Published: 16 March 2026

RESEARCH-ARTICLE

Detecting Eating Events with Inertial Sensing in a Ring Wearable

DAWEI LIANG, The University of Texas at Austin, Austin, TX, United States

BOBBY CHIU, The University of Texas at Austin, Austin, TX, United States

CODY ARVONEN, The University of Texas at Austin, Austin, TX, United States

AUSTIN JAMES, The University of Texas at Austin, Austin, TX, United States

YEJI KIM, Cockrell School of Engineering, Austin, TX, United States

EDISON THOMAZ, The University of Texas at Austin, Austin, TX, United States

Open Access Support provided by:

The University of Texas at Austin

Cockrell School of Engineering

Detecting Eating Events with Inertial Sensing in a Ring Wearable

DAWEI LIANG[*†], The University of Texas at Austin, USA
BOBBY CHIU[*], The University of Texas at Austin, USA
CODY ARVONEN, The University of Texas at Austin, USA
AUSTIN JAMES, The University of Texas at Austin, USA
YEJI KIM, The University of Texas at Austin, USA
EDISON THOMAZ, The University of Texas at Austin, USA

Automated dietary monitoring (ADM) is increasingly recognized as a critical component to advance health tracking and inform behavioral interventions. Although prior research has investigated ADM through various forms of wearable devices and sensing modalities, e.g. wrist-worn motion sensors, acoustic sensing, and vision-based approaches, the potential of smart rings, an emerging and increasingly popular wearable form factor, remains largely underexplored. In this work, we focus on detecting eating events, a key component of ADM, leveraging inertial data acquired from a custom-designed wireless ring. We developed a prototype ring device accompanied by a smartphone application to facilitate data collection. To evaluate the feasibility and effectiveness of ring-based ADM, we conducted studies with 42 participants in three settings: a controlled laboratory environment, a semi-controlled environment, and a deployment in the wild. Our analysis addresses key research questions regarding the accuracy of eating event detection in various contexts, the influence of ring placement on performance, the impact of personalization, and the potential benefits of integrating inertial data from both the ring and a commodity smartwatch. Our ring wearable achieved F1-scores of 82.6% in the controlled laboratory setting and 65.9% in the semi-controlled condition. We showed that while a ring wearable alone does not perform well in naturalistic settings, it can complement a smartwatch sensor at this task, and performance can be further improved through personalization. To foster continued advancement in ring-based ADM research, we will publicly release a paired ring and smartwatch dataset comprising 80 hours of annotated eating and non-eating activities. To our knowledge, this will constitute the largest publicly available dataset focused on eating event detection using ring-based inertial sensing.

CCS Concepts: • **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; **Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: Eating Detection, Wearable Sensing, Activity Recognition

ACM Reference Format:
Dawei Liang, Bobby Chiu, Cody Arvonen, Austin James, Yeji Kim, and Edison Thomaz. 2026. Detecting Eating Events with Inertial Sensing in a Ring Wearable. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 1, Article 9 (March 2026), 29 pages. https://doi.org/10.1145/3789678

[*]Both authors contributed equally to this research.
[†]Corresponding author.

Authors' Contact Information: Dawei Liang (corresponding author), The University of Texas at Austin, Austin, Texas, USA, dawei.liang@utexas.edu; Bobby Chiu, The University of Texas at Austin, Austin, Texas, USA, bychiu@utexas.edu; Cody Arvonen, The University of Texas at Austin, Austin, Texas, USA, codyarvonen@utexas.edu; Austin James, The University of Texas at Austin, Austin, Texas, USA, ajamesut@utexas.edu; Yeji Kim, The University of Texas at Austin, Austin, Texas, USA, yejikim@utexas.edu; Edison Thomaz, The University of Texas at Austin, USA, ethomaz@utexas.edu.

ACM 2474-9567/2026/3-ART9
https://doi.org/10.1145/3789678

1 INTRODUCTION

Understanding eating behaviors and dietary intake is essential for promoting long-term health and supporting goals such as addressing nutritional deficiencies, managing weight, and encouraging healthier food choices. Traditional dietary monitoring methods often rely on self-reporting, which can be prone to bias and inaccuracies due to forgetfulness or incomplete recall [14, 15]. To address these limitations, recent research has focused on automated dietary monitoring (ADM) using on-body wearable sensors. Advances in wearable technologies such as by using inertial measurement units (IMUs), acoustic sensors, and visual cameras have enabled lightweight and unobtrusive solutions for capturing dietary behaviors in real-world settings.

Numerous studies have explored ADM through wearable technologies designed to track food intake. Among these, IMU-based approaches have been especially popular due to their low computational overhead, minimal privacy concerns, and lower manufacturing costs compared to acoustic or vision-based sensing [26]. IMU sensors have been integrated into various wearable form factors such as smartwatches [30, 31, 35], eyeglasses [20, 29, 33], and earbuds [8]. Smart rings, in contrast, are a relatively recent addition to the wearable sensing landscape. While the adoption of smart rings is growing, thanks to their unobtrusive form factor and health tracking capabilities, their potential remains largely underexplored in ADM. Preliminary studies have begun to investigate their use in this context, but several key research questions remain unanswered. For instance, most prior work has been limited to controlled lab environments [13, 18], with little evaluation of ADM performance using ring-based sensing in more realistic, everyday settings. Additionally, direct comparisons between ring-based ADM and other form factors like smartwatches are lacking. Design considerations such as optimal ring placement on the hand have also received limited attention. Compounding these challenges is the fact that most commercial smart rings do not provide open-source APIs for accessing raw sensor data, further limiting research progress. In this work, we address these limitations with the following contributions:

- The design and implementation of a wireless ring wearable with inertial sensing capabilities and a companion smartphone application for data collection. We also introduce a technical framework for the detection of eating events, including an architecture for processing multi-device data.
- A comprehensive evaluation across diverse settings. We conduct experiments involving 42 participants of a diverse background across three scenarios, including controlled lab, semi-naturalistic, and unconstrained in-the-wild environments. Using only IMU data from a single ring, we achieved class-level F1-scores of 83% and 66% for detecting eating events in the lab and semi-naturalistic settings, respectively.
- A detailed investigation of challenges and opportunities affecting eating episode detection. We provide an in-depth analysis of eating detection performance across different settings, examine the effects of ring placement on various fingers, assess the benefits of augmenting ring data with smartwatch data, and discuss practical considerations such as finger placement variability and the role of model personalization for improving the detection.
- The release of ADM datasets[1] containing 80 hours of eating and non-eating activities from the 42 participants captured with our ring prototype and smartwatch. To our knowledge, this is the largest publicly available dataset for ADM based on motion sensing from a single ring device.

2 RELATED WORK

Over the last decade, researchers have explored the automation of various dimensions of dietary behaviors such as detecting when people are eating, predicting when they are likely to eat, and recognizing what and how much is consumed [1, 4, 9, 35, 41]. In this section, we first provide an overview of prior work in ADM leveraging multimodal sensing. Many recent systems in this category have been designed in the form of eyeglasses. Next, we review how inertial sensing has been specifically used for detecting intake gestures and eating episodes,

[1] https://doi.org/10.18738/T8/IFTEDK

largely with head-mounted sensing. We conclude with a discussion of related research using wrist-mounted and ring-based sensing, which are particularly relevant in light of the contributions of this work.

2.1 Eyeglasses and Multimodal Sensing

Eating activities often involve multiple steps such as fetching and mixing foods, chewing, swallowing and digesting. Using multiple sensors to capture these complex motions reveals relevant information regarding dietary activities. One of the earlier attempts to utilize multimodal sensing for ADM was made by Liu *et al.*, who presented a food-logging application based on audio and images [19]. The system identified chewing events from acoustic data, then a wearable camera was used to capture a video of the eating activity [19]. In 2016, Rahman *et al.* presented a wearable sensing framework consisting of an array of sensors [27] and in the same year, Merck *et al.* used acoustic data from the ear, motion data from the head, and wrist sensors to capture eating-related movements [22]. In 2020, Bedri *et al.* developed a pair of eyeglasses called FitByte that was outfitted with inertial sensors, optical sensors, and a camera [3]. It captured jaw movements (IMU), swallowing sounds (microphone), hand-to-mount gestures (proximity), and visuals of the food (camera). FitByte recognized eating episodes with 94.1% accuracy, detected eating activity within 7 seconds, and estimated the duration of the eating episode with 96.3% accuracy. More recently, Shin *et al.* [33] demonstrated chewing detection using eyeglass frames, achieving an F1-score of 0.92 in real-world settings. Follow-up work by Bedri *et al.* [5] and Saphala *et al.* [29] also used eyeglass-based sensing.

2.2 Head-mounted and Inertial Sensing

The availability of inertial sensors in many consumer electronic products has encouraged the use of accelerometers and gyroscopes for detecting eating activities. In 2017, Bedri *et al.* presented EarBit, which detected chewing instances using a behind-the-ear inertial sensor in two studies conducted in a semi-controlled laboratory environment and an outside-the-lab environment. In the semi-controlled environment, EarBit detected all eating episodes and achieved an F1 score of 90.9% in detecting chewing instances. In the outside-the-lab evaluation, EarBit achieved an F1 score of 80.1% in detecting chewing instances [4]. In another study, Wang *et al.* utilized a headband fashioned with a 3-axis accelerometer located at the right temple. This sensing technique used the accelerometer to detect eating episodes and count chews, and it resulted in 94.4% accuracy and an F-score of 87.2% [37]. Morshed *et al.* [8] used motion signals from earbuds to detect snack consumption with personalized models.

In 2020, Chun *et al.* proposed a small-scale, Bluetooth-enabled device equipped with an inertial sensor mounted underneath the chin to monitor chewing movement by the jaw. A lab study was conducted and achieved 91.7% precision and 91.3% recall, and a naturalistic study achieved 92.3% precision and 89.0% recall. These studies resulted in a >10% improvement in F1 score as compared to prior works using necklace-based ADM devices. While discrete, the placement of the device is unlikely to be favored for everyday non-clinical applications [10].

2.3 Smartwatches and Wrist-based Sensing

Several studies have explored wrist-worn devices for tracking eating-related activities. More recently, this has been the case due in large part to the popularity and social acceptability of smartwatches. However, the first explorations of tracking food intake gestures with wrist-mounted devices made use of a smartphone strapped to the wrist. In 2013, Dong *et al.* compiled 449 hours of inertial sensor data from 43 participants [11]. The average detection accuracy for eating activities was reported as 81% [11]. As a follow up to this work, Thomaz *et al.* collected hand gesture data from 20 participants in a semi-controlled laboratory setting, but using a smartwatch [35]. With models trained on the semi-controlled laboratory environment, Thomaz *et al.* applied the models to

Table 1. Comparison of Ring-Based Sensing for ADM and Eating Gesture Detection. DL: Deep learning approaches; FS: Sensor fusion.

Reference	Models	Setting	Platform
[13]	Statistical	1 subject, 7 activities	Shimmer
[18]	Heuristic	2 subjects, 10 meals	Shimmer
[2]	Heuristic, DL	16 subjects, home and cafeteria	Wave Ring [38]
Ours	Statistical, DL, FS	42 subjects, Campus, Free-living	Low-cost prototype

two independently collected, free-living data sets. The detection performance was reported as 76.1% and 71.3% from the free-living study [35].

In following years, this research was further validated. Weiss *et al.* [39] demonstrated that smartwatches outperform smartphones in detecting eating behaviors by analyzing hand motions. Similarly, Sen *et al.* [30, 31] investigated eating gesture and episode recognition using wrist-worn smartwatches and successfully distinguished eating from non-eating activities with both personalized and generalized models. Mekruksavanich *et al.* [21] applied a hybrid deep learning model to recognize a range of activities including eating by using the public activity recognition dataset proposed by Weiss *et al.* [40], achieving high accuracy in detecting eating events.

To further support research in smartwatch-based dietary monitoring, several smartwatch-based eating-related datasets have been made publicly available. These include the OREBA [28] dataset, collected under controlled laboratory conditions, and free-living datasets such as Clemson All-Day [32] and FreeFIC [17], which capture eating behaviors in natural, real-world environments.

2.4 Rings and Finger-based Sensing

As an emerging technology, smart rings are attracting increased interest thanks to their small size and capabilities as a health tracker. Their popularity is also growing as a research platform, and prior studies have explored the potential of leveraging ring-like form factors for ADM. Fan *et al.* [13] demonstrated an approach to infer eating gestures using finger motion data collected from a commercially available device, the Shimmer sensor kit. Their analysis, based on a controlled study with 375 gestures across seven activities, achieved high binary classification accuracy using statistical algorithms. Similarly, Lin and Hoover [18] proposed a bite-counting algorithm based on finger motion data. These studies were constrained by small datasets and controlled test environments, and did not explore modern deep learning approaches that have since demonstrated strong performance in ADM tasks.

The most recent and closely related work to ours is by Armin *et al.* [2], where the authors proposed a mobile application to unlock commercial smartwatches and smart rings for intake gesture data collection. Using deep learning models, they achieved an F1-score of 0.74 for eating action detection based on data collected from a commercial smart ring device (the Wave ring [38]) with 16 participants.

While inspired by these prior efforts, as shown in Table 1, our work differs from the previous ring-based ADM studies in several important ways. Firstly, we propose a low-cost, high-fidelity technical framework for building a single ring-based sensing prototype for ADM instead of using an existing smart ring. Secondly, we conducted an in-depth investigation of finger-based eating event detection, exploring the effects of sensor fusion (i.e., ring and smartwatch combined), the impact of finger placement, and the sensitivity of model performance to window size configurations and also personalization. Lastly, we conducted human subject studies in controlled laboratory, semi-naturalistic, and naturalistic settings with 42 participants of diverse backgrounds. To the best of our knowledge, the inertial sensing dataset we compiled using a ring and smartwatch is the largest to date for research on ADM, and detection of eating events specifically.

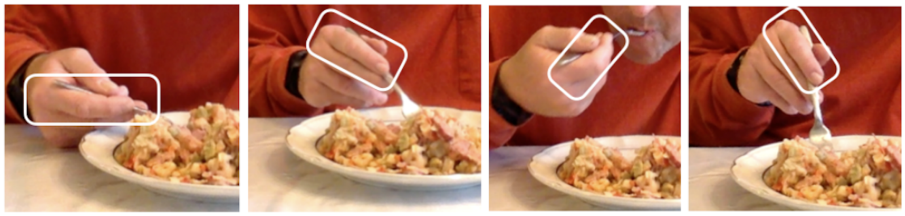

Fig. 1. The range of motion of the index finger while eating is considerably larger than wrist movements and rotation.

3 DIETARY TRACKING WITH A RING

As previously outlined, the goal of this work is to advance an approach to detect eating events with an instrumented ring. Our motivation for investigating the potential of a ring for dietary activity tracking is based on a hypothesis formulated while observing participant videos from our previous dietary monitoring studies. These videos showed that the range of motion of the index finger while eating is significantly more pronounced than the motion and rotation of the wrist, as shown in Figure 1. Due to the popularity of smartwatches with increasing complexity, both in terms of processing and sensing capabilities, researchers have investigated wrist motion as a proxy in a wide variety of human activity recognition applications, including dietary monitoring. Finger movement, on the other hand, has received much less attention in this context, and thus provides a promising framing for the research questions in this work:

(1) How reliably can ADM, particularly eating event detection, be performed using a single ring-based sensing prototype, and how does its performance compare to motion sensing from adjacent form factors, such as commercially available smartwatches?
(2) How do various factors such as the placement of the rings on different fingers, the choice of inference models, and window size configurations affect the performance of eating detection using a ring-based sensing approach?
(3) To what extent can ADM models developed with ring sensing data generalize across different settings, including controlled, semi-naturalistic, and fully unconstrained eating behaviors, i.e., in-the-wild, for the detection tasks?

To address these research questions, we first introduce the ring wearable device and its companion smartphone application. Section 4 details our data collection efforts using the ring prototype in a controlled lab setting, where 20 participants performed both eating and non-eating activities according to a scripted protocol. Section 5 presents our framework for data collection, modeling and analysis, which are aimed at addressing Research Questions (1) and (2). Sections 6 and 7 explore the generalization capabilities of our models through semi-naturalistic and fully in-the-wild studies, addressing Research Question (3). We conclude with a discussion of our findings and the limitations of the work.

3.1 Ring Wearable Hardware Design

Our ring prototype strives to preserve the sleek and compact form factor of a conventional ring to ensure comfortable everyday wear while achieving performance comparable to that of a smartwatch. Unlike many existing smart rings that distribute their internal components uniformly around the finger, our design adopts a form factor similar to that of a signet or championship ring. This configuration allows the bulk of the electronic components to be positioned at the top and bottom of the ring by minimizing the volume along the sides. The ring body is made with a customized 3D-printed polylactic acid (PLA) biopolymer, with internal cutouts to house wiring for the battery and IMU board. The battery is a 3.7V 140mAh Lithium Polymer battery (GEB401015) from

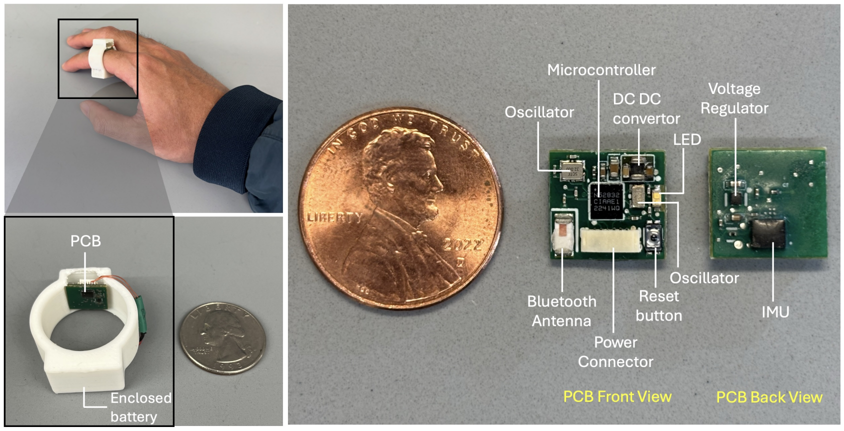

Fig. 2. Overview of the custom-built IMU ring prototype and its components. *Left*: The ring worn on the finger (top) and a close-up view (bottom) showing the 3D-printed enclosure housing the custom PCB and battery, with a U.S. quarter for scale. *Right*: Front and back views of the custom PCB placed beside a U.S. penny. Only one board is utilized in the ring; two are shown for illustrative purposes.

General Electronics Technology Co., Ltd. We designed a custom functional printed circuit board (PCB) that is inserted into the 3D-printed housing. This enabled us to optimize the design of the system to promote low-power operation, which is necessary for deploying it in-the-wild. The custom designed IMU board consists of a Nordic nRF52832 microcontroller and a Bosch BMI323 6-axis IMU. The nRF52832 was chosen because of the low-power capabilities and Bluetooth Low Energy (BLE) features. The nRF52832 was chosen over others in the nRF lineup (including the nRF52811 and nRF52840) because it provides more memory and processing power while still maintaining low-power capabilities. Other components on the custom IMU board include a Bluetooth antenna to increase BLE strength, an indicator LED, and an external oscillator. The PCB prototype is a fully functional implementation. The finger-facing side of the board was coated with a conformal waterproof layer to protect against moisture exposure during everyday wear.

Figure 2 presents snapshots of our ring prototypes used for collecting data from 6-axis accelerometer and gyroscope sensors. The ring is designed to closely resemble the size and form factor of modern commercial smart rings. Specifically, the sensor board measures 10.75mm x 10.25mm x 3.6mm in dimension and weighs approximately 0.36 g. The battery weighs 1.14 g. As can be seen in Figure 2, the overall size of this custom designed sensor board and the battery are comparable to the size of a U.S. quarter. To accommodate different finger sizes, we developed multiple ring sizes. The typical diameter of the ring prototypes ranges from 14.1 cm (U.S. ring size 3) to 22.2 cm (U.S. ring size 13), with a total weight varying between 2.22 g and 2.6 g.

3.2 Custom Smartphone App for Data Collection

Since the ring is too small to accommodate a user interface, we developed a custom iOS app using Swift to connect the ring to a user's smartphone. This app allowed us to control the start and stop of data recording and to retrieve sensor data after each session. It includes the main controls for initiating and ending data collection. The app managed the recording sessions for each participant, including the ability to pause data collection mid-session. Upon completion, the app provided options to export the recorded data to a computer for further processing.

To support a comparative study, we also collected motion data using a commercial smartwatch in addition to the ring. The iOS app was therefore designed to interface with both the ring and the smartwatch using the same set of controls for recording and data management. For deployment, the app was installed on an iPhone 13 mini and used in conjunction with an Apple Watch Series 8. The custom ring was designed to temporarily store sensor data in a buffer before transmitting it to the iPhone. Both the ring and the watch streamed data live to the phone, with the watch sending packets of 100 samples and the ring sending packets of 20 samples. All data was sampled at 100 Hz. To ensure proper synchronization, each data packet from both the ring and the watch included a local timestamp with sample-level precision. These timestamps enabled accurate alignment of multi-modal sensor data during analysis.

4 CONTROLLED LAB STUDY

To evaluate the performance of ADM using our ring prototype, we first conducted a lab-based study in which participants were asked to perform a series of eating and non-eating activities according to a predefined study script. While the study followed a structured protocol, it retained naturalistic elements. In particular, participants were encouraged to follow their own behavioral styles with minimal restrictions. They were also allowed to speak freely and complete the tasks in a public kitchen space. Food and utensils were prepared in advance to support the eating activities. The study was specifically designed to address Research Questions (1) and (2): to gain an initial understanding of the performance of ring-based eating detection, to compare it with that of a commercially available smartwatch, and to investigate how factors such as modeling strategies and ring placement on different fingers affect performance.

4.1 Participants

In total, we recruited 20 participants from a university campus for this phase of the study. Participant ages ranged from 21 to 33, with a mean age of 24. With the exception of one individual, all participants were right-handed. The participant group included 4 females and 16 males, with self-reported ethnicity as follows: 11 Asian, 7 White, 1 Latino, and 1 Mixed. Data from one participant (ring on the middle finger) was excluded from further analysis due to data loss and the participant's inability to complete the eating portion of the study. As a result, data from 19 participants was used for model development and subsequent analysis.

4.2 Experimental Procedures

Each participant completed a single session that involved eating a meal in a lounge area and performing a set of non-eating activities outlined in a study script. These non-eating activities were selected to closely resemble typical hand movements involved in eating, thereby creating scenarios likely to cause confusion in the classification of dietary gestures, and thus providing a more realistic test of model robustness.

Before starting data collection, each participant signed an IRB-approved consent form and was informed about study procedures and potential risks. Participants were asked to wear both the custom ring and an Apple Watch Series 8 on their dominant hand. A researcher assisted in fitting the ring to the participant's finger by selecting the appropriate ring size. The researcher also helped the participant with the smartwatch if needed.

To capture data from all five fingers, we divided the 20 participants into five evenly distributed groups, with each group assigned to wear the ring on a different finger of the dominant hand. Each participant wore only one ring on one finger. This design allowed us to collect data with balanced finger placement across participants.

Once participants were ready to initiate the sequence of activities, a researcher initiated data capture from both the ring and watch via the smartphone app. During the session, a video camera recorded the entire session to provide ground truth labels. The study activities are described in Appendix A.1.

Fig 3 shows example images captured during the lab study, illustrating participants performing both eating and non-eating activities. Although the activities were pre-scripted, we emphasized a naturalistic data collection environment. Participants were allowed to speak freely with the researchers or other individuals present in the

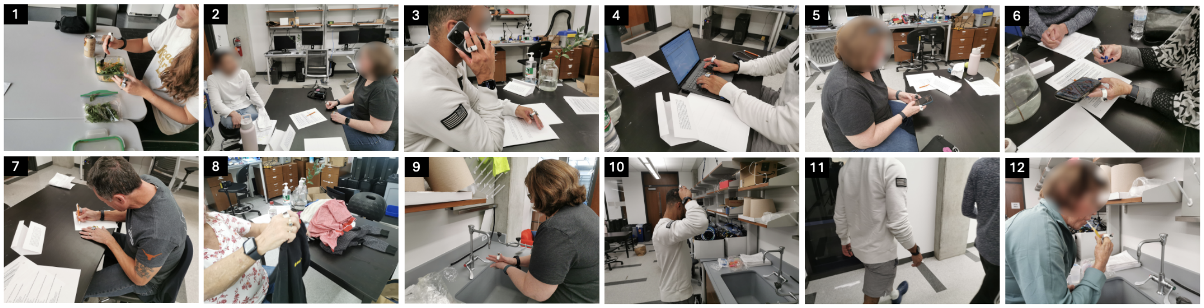

Fig. 3. Sample photos of participants performing the target activities wearing our ring prototypes in the lab study. Activity indices: 1 - eating; 2 - conversation; 3 - giving a telephone call; 4 - typing on a laptop; 5 - typing on a mobile phone; 6 - using a mobile phone; 7 - writing on paper; 8 - folding laundry; 9 - simulating washing hands; 10 - combing hair; 11 - walking around; 12 - brushing teeth

building throughout the session. After each session, all devices were sanitized. Hygiene-related items such as combs, toothbrushes, and toothpaste were either given to participants or discarded if left behind.

After completing the study session, each participant was asked to fill out a demographic questionnaire. As compensation for their time, participants received a $10 USD gift card. The researcher saved all collected data locally on the smartphone and backed it up to a secure server for later processing and analysis.

5 DATA COLLECTION, MODELING AND ANALYSIS

In total, we collected approximately 13.5 hours of sensor data from the ring prototype and a comparable amount from the smartwatch, averaging around 42 minutes per participant. Due to occasional disconnections between the devices and the control smartphone, a small amount of data loss occurred in several sessions. After filtering out unusable segments, we retained around 13.1 hours of synchronized IMU data from the ring and watch for actual analysis. Common causes of data loss included power disconnections during vigorous hand movements and temporary device removal, such as when participants visited the restroom. The IMU data captured from both the ring and the watch consisted of six-dimensional motion data, including 3-axis acceleration and 3-axis gyroscope signals. This data was retrieved from the devices at the end of each session using our custom iOS app and securely stored on a server for further processing. For specific details about ground truth annotation, refer to Appendix A.2.

5.1 Pre-processing

With the annotated data, we conducted a series of pre-processing steps to prepare for ADM model training and evaluation. This involved data windowing and standardization for machine learning models. We applied empirical window sizes ranging from 2 to 30 seconds to capture different granularity of eating events, following practices commonly used in prior ADM research [7, 29, 35]. Windows that did not contain enough sample points at the end of an event were discarded. A 50% overlap was applied between adjacent windows to ensure smoother temporal coverage. For each participant's data, we performed global standardization by subtracting the mean and dividing by the standard deviation at each sample point, improving the suitability of the data for computational modeling.

5.2 Classification

To address Research Questions (1) and (2), we first focused on the task of eating event detection at the window level, as widely explored in prior work [7, 29, 35]. We experimented with several classification models, including

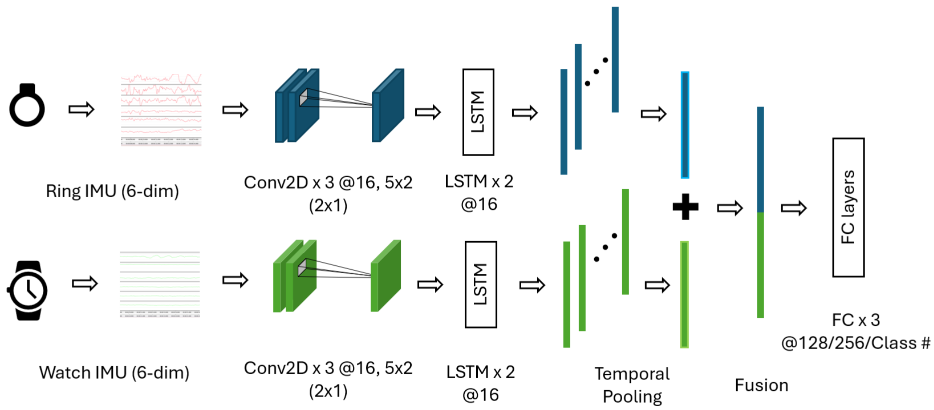

Fig. 4. The DeepConvLSTM-split neural network architecture we developed for sensor fusion with the ring and the watch IMU data. We developed separate input branches to process the ring and the watch data individually.

Random Forest, Support Vector Classifier (SVC), and Deep Learning models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Transformers [36]. Specific configuration details of these models are in Appendix A.3.

5.3 Ring and Smartwatch Sensor Fusion

In addition to capturing finger movements from our custom ring prototype and wrist movements from the smartwatch independently, another important aspect we explored is the fusion of both sensor streams for joint modeling of eating behaviors, thereby leveraging the unique motion patterns captured by each device. To enable this, we modified our CNN and DeepConvLSTM architectures to accept inputs from both sensors simultaneously. We explored two strategies for sensor fusion: early-fusion and split-branch fusion.

- **Early-Fusion**. We applied early fusion by stacking the sensor data from the ring and watch to form a combined input. Specifically, we concatenated the accelerometer and gyroscope signals from both devices into a 12-dimensional input (X, Y, Z for acceleration and gyroscope from each device). This approach enables information from both sensors to be jointly optimized within the convolutional layers without significantly increasing model complexity. Early fusion has been widely adopted for multi-sensor modeling tasks, where data is combined before being fed into a model [16].
- **Split-Branch**. We explored a split-branch fusion architecture in which separate input branches were created for the ring and the watch sensor streams: one block of convolutional layers was used to process the ring's data, and an identical but independent block processed the watch's data. The outputs from these two branches were then fused at an intermediate layer, either by direct concatenation or pooling. As illustrated in Fig. 4, for the CNN model, fusion was performed after the convolutional blocks, whereas for the DeepConvLSTM model, fusion was conducted after the outputs of each branch's LSTM layers.

In the following sections, we refer to these models as *CNN-early*, *DeepConvLSTM-early* for the early fusion strategy, and *CNN-split*, *DeepConvLSTM-split* for the split-branch fusion strategy, respectively.

Table 2. Overall performance of window-level eating detection using various modeling approaches, comparing our ring prototype with a commercial smartwatch as the sensing device, based on data from 13 participants in the lab study. Sensor fusion combining IMU data from both the ring and the watch consistently yields the best detection performance under the same model configurations.

	Model	**Class F1**	**Precision**	**Recall**
Smartwatch	Random Forest	0.874	**0.897**	0.860
	SVC	0.837	0.803	0.890
	CNN	0.868	0.826	0.933
	DeepConvLSTM	0.889	0.848	**0.947**
	Transformer	**0.895**	0.877	0.926
Ring	Random Forest	0.808	**0.862**	0.768
	SVC	0.761	0.732	0.805
	CNN	0.773	0.705	0.866
	DeepConvLSTM	0.823	0.769	**0.914**
	Transformer	**0.826**	0.793	0.872
Fusion (ring + watch)	CNN-early	0.903	0.874	0.941
	CNN-split	0.902	0.860	0.958
	DeepConvLSTM-early	**0.913**	**0.905**	0.929
	DeepConvLSTM-split	0.904	0.864	**0.963**

5.4 Results for Window-level Eating Detection

For evaluation, we adopted a leave-one-participant-out (LOPO) cross-validation scheme. Model performance was assessed using class-level F1 score, precision, and recall, which are commonly used metrics in dietary monitoring tasks [6, 20, 33, 35]. Table 2 presents the overall LOPO results obtained from the lab study. This analysis focuses on Research Question (1): whether IMU data collected from a ring device can support accurate eating event detection. For each combination of sensor input type and model, we report results using the optimal window size (within the tested range) that yielded the highest F1 score. For the Transformer models, the input window size was limited to a maximum of 15 seconds due to computational constraints.

As shown in Table 2, the best results for ring-based detection were primarily achieved using advanced deep learning models such as DeepConvLSTM and Transformers, highlighting their superior capacity to model complex temporal patterns compared to pure CNNs and the tested statistical methods. To contextualize these results, we also evaluated model performance using IMU data collected from the Apple Watch. Similarly, the Transformer model achieved the best performance, closely followed by DeepConvLSTM; both consistently outperformed the other modeling approaches.

Overall, detection performance using ring data was lower than that achieved with watch-based sensing across all model types, which is a trend also observed in prior work [2]. This may be attributed to the smartwatch's ability to capture broader, less subtle wrist movements, which may be less affected by confounding non-eating gestures. Further exploration into the differential sensitivity of wrist- versus finger-based sensing could provide valuable insight into optimizing ring-based ADM systems.

Table 2 also compares ADM performance between single-modality input and sensor fusion approaches. We observe that models using sensor fusion consistently outperform those relying on a single sensor type, highlighting the benefit of combining coarse-grained wrist movements with fine-grained finger movements for

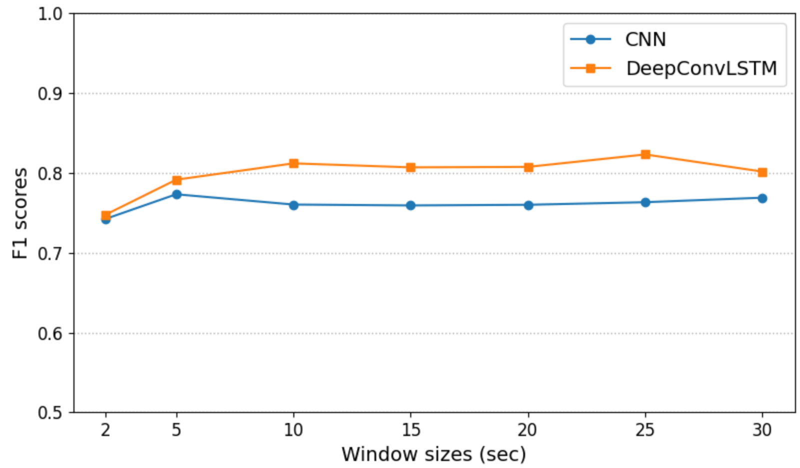

Fig. 5. Effects of window size on eating detection performance using ring data in our lab study. Overall, model performance remains relatively consistent for window sizes between 5 and 30 seconds, suggesting robustness to window duration within this range.

dietary monitoring. Interestingly, the more complex fusion methods with multi-input branches did not always yield better detection performance than fusion methods with a simpler model size, such as early fusion. A possible explanation is that simpler fusion models may generalize better to our dataset, given the balance between model complexity and the available data size.

Figure 5 illustrates the effect of varying window sizes on eating detection performance using ring data, comparing results from the CNN and DeepConvLSTM models. The results show that optimal window sizes tend to be larger, at typically above 5 seconds, rather than shorter durations such as 2 seconds. A likely explanation is that longer windows allow the models to capture more global patterns in hand movement and are less susceptible to transient finger motion artifacts. This effect is especially relevant given that our annotations are made at the event level. Notably, we observed that the performance of neural network models remains relatively stable once the window size exceeds a certain threshold (e.g., 5 seconds), indicating that the models are able to effectively capture the necessary hand motion patterns within that temporal range.

5.5 Finger Placement Performance Analysis

Beyond evaluating the overall eating detection performance using the ring, we also examined how ring placement on different fingers may influence the detection performance, given that individuals commonly wear rings on various fingers in daily life. As described in Section 4, participants were divided into five groups based on the specific finger (thumb, index, middle, ring, or pinky) of their dominant hand on which the ring was worn.

To assess the effect of finger placement, we conducted two types of analyses as illustrated in Fig 6: *Finger-Agnostic LOPO Evaluation* and *Finger-Specific LOPO Evaluation*.

- **Finger-Agnostic LOPO Evaluation**. This evaluation first trains across all participants except one in a LOPO style evaluation before grouping participants according to the finger on which the ring was worn and averaging the results in the group. It compares model performance across different fingers to determine whether some finger placements leads to better eating detection accuracy. This helps assess the model's robustness when the finger placement is not predetermined.

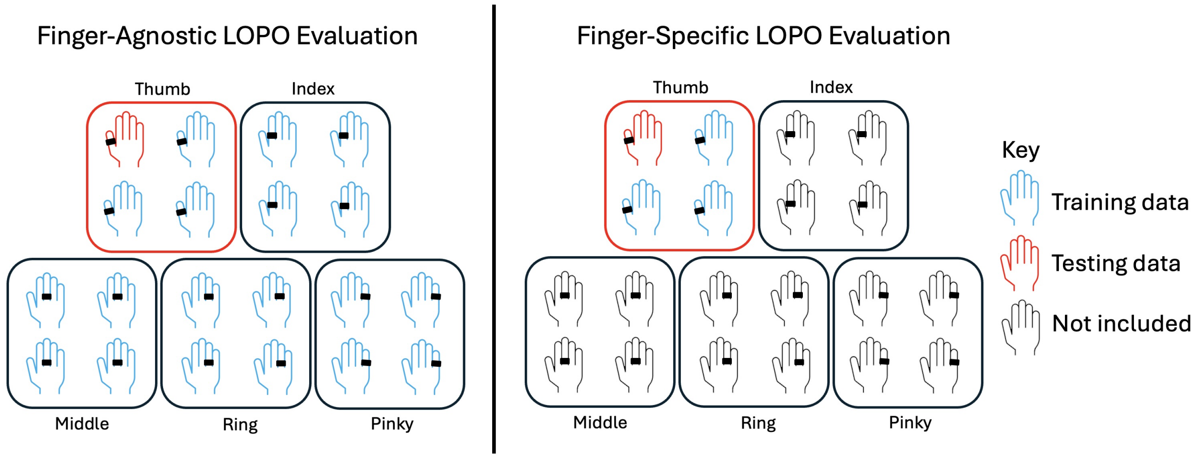

Fig. 6. Diagram illustrating the two evaluation settings used for analyzing the effect of finger placement on eating detection performance. Left: *Finger-Agnostic LOPO Evaluation*, where results for participants wearing the ring on the same finger are grouped from the global LOPO evaluation. Right: *Finger-Specific LOPO Evaluation*, where a separate LOPO evaluation is performed within each group of participants wearing the ring on the same finger.

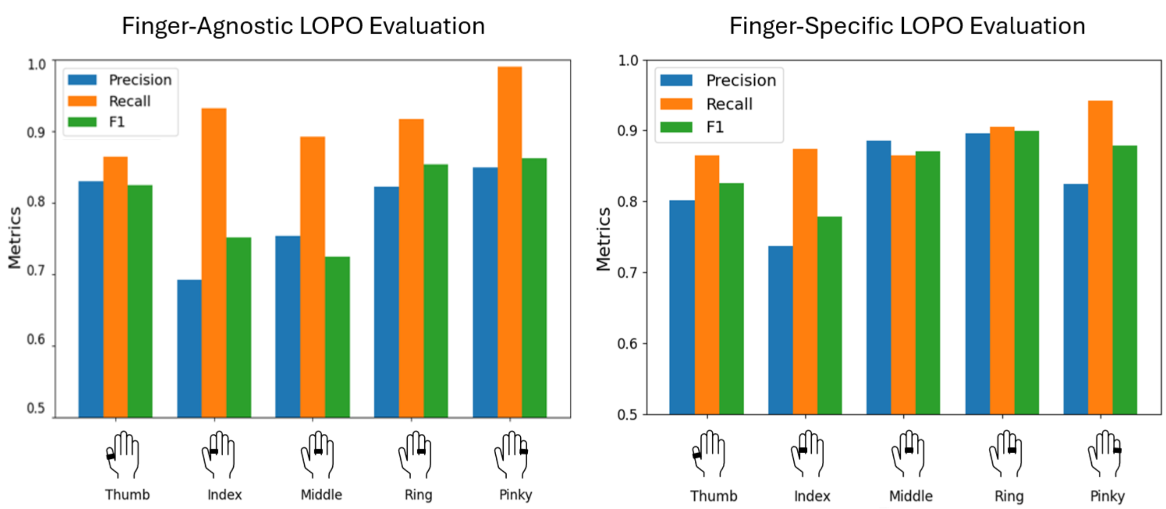

Fig. 7. Model performance for eating detection in the lab study under Finger-Agnostic LOPO evaluation (left) and Finger-Specific LOPO evaluation (right). Among the five fingers, models trained and tested on data from the thumb, ring, and pinky fingers demonstrated the most robust and consistent performance.

- **Finger-Specific LOPO Evaluation**. This evaluation trains models using data from participants who all wore the ring on the same finger, holding out one participant at a time (LOPO within each finger group).

Average results for each finger group are reported. This provides finger-specific performance metrics and helps identify which finger yields the most reliable detection, guiding optimal ring placement.

Fig 7 presents eating detection performance under both evaluation settings, using a 20-second window size optimized for the DeepConvLSTM model. Across both settings, models trained and tested on ring data from the thumb, ring, and pinky fingers demonstrated the most robust and consistent window-level eating detection performance.

In contrast, models tested on data from the index and middle fingers showed noticeably lower performance. This likely reflects greater variability in motion patterns associated with these fingers across participants, which can hinder model generalization. A plausible explanation is that the index and middle fingers are frequently involved in a wide range of individualized, non-eating activities such as typing or writing, which differ significantly between users and introduce confounding motion signals for ADM models. In comparison, the thumb, ring, and pinky fingers tend to exhibit more stable and consistent movement patterns during eating, making them more reliable locations for ring-based sensing in ADM tasks.

6 SEMI-NATURALISTIC STUDY

To evaluate how ring-based eating detection performs in more naturalistic conditions, particularly in relation to Research Question (3), we conducted a semi-naturalistic study following a similar setup to the lab study but with a more relaxed protocol. This study was designed to be more naturalistic in several key ways: participants were not restricted to specific food types and were allowed to bring their own food and utensils for the eating session. Additionally, we recruited participants from the broader local community, rather than exclusively from the university campus, resulting in a more diverse sample in terms of age, ethnicity, and eating behaviors. Twenty participants were recruited for this semi-naturalistic study and their demographic information can be found in Appendix B.

6.1 Experimental Procedures

For this study, we followed a study script that included the same list of eating and non-eating activities as in the lab study. The use of a predefined script allowed us to introduce specific non-eating activities with hand movement patterns similar to eating, enabling a more controlled evaluation of ADM performance using ring-based sensing. However, unlike the lab study, no food or utensils were provided; participants brought their own meals and necessary utensils and ate at their natural pace, as they normally would.

During the eating sessions, participants either sat at a table or on a sofa with a nearby couch table. For non-eating activities, participants mostly sat at a table, although some preferred to stand. To introduce more natural variability, we varied the social interaction setting: for half of the participants, a researcher engaged them in casual conversation during eating activities to mimic daily scenarios where people eat while chatting; for the other half, participants ate alone while the researcher remained outside the lounge area.

6.2 Data Annotations and Pre-processing

The entire semi-naturalistic study was conducted over approximately one week, with two to three participants scheduled each day. However, we experienced significant data losses during this phase due to technical and participant-related issues. As a result, usable data were obtained from 13 out of the 20 participants originally recruited. Examples of data loss included hardware failures such as power disconnections between the battery and the ring sensor board caused by vigorous hand movements and communication issues resulting from the physical isolation between participants and researchers during sessions. In total, we collected approximately 11.4 hours (41,056 seconds) of usable ring sensor data across participants, averaging around 0.9 hours per participant.

Table 3. Overall performance for window-level eating detection using different modeling approaches, comparing our ring prototype against a commercial smartwatch as the sensing device in the semi-naturalistic study. Results are based on data from 13 participants. Sensor fusion by combining IMU data from both the ring and the watch yielded the highest inference F1 performance under the same model configurations.

IMU Source	Model	Class F1	Precision	Recall
Smartwatch	Random Forest	0.670	**0.753**	0.621
	SVC	0.613	0.660	0.585
	CNN	0.682	0.597	**0.907**
	DeepConvLSTM	0.660	0.573	0.836
	Transformer	**0.691**	0.633	0.824
Ring	Random Forest	0.592	**0.676**	0.536
	SVC	0.559	0.588	0.547
	CNN	0.643	0.523	0.877
	DeepConvLSTM	**0.659**	0.600	0.827
	Transformer	0.598	0.451	**0.928**
Fusion (ring + watch)	CNN-early	0.659	0.596	0.759
	CNN-split	0.686	0.596	0.841
	DeepConvLSTM-early	0.677	0.594	0.800
	DeepConvLSTM-split	**0.706**	**0.625**	**0.864**

6.3 Results

We applied the LOPO evaluation scheme to assess intake gesture detection performance using the ring data. Table 3 summarizes the overall detection results across different modeling approaches, i.e., Random Forest and SVM, CNN, DeepConvLSTM, and Transformers, using the same training setup as in the lab study. Similar to the setup in the lab study, the temporal window size was selected to optimize performance for each sensor/model pair, with window lengths ranging from 2 to 30 seconds and a 50% window overlap. For Transformers, the tested range was up to 15 seconds.

For comparison, the table also includes results obtained using IMU data from the smartwatch and the sensor fusion of ring and watch signals. As expected, the eating detection performance metrics for all devices declined compared to the lab study results, due to the greater variability of intake gestures when participants ate meals of their own choice. Nonetheless, the overall trends remained consistent with the lab study findings: smartwatch-based models generally outperformed ring-based models, and deep learning approaches (CNN, DeepConvLSTM, and Transformers) consistently outperformed statistical models, demonstrating the superior capability of complex models for ADM tasks. Interestingly, in this phase of the study, the Transformer model did not outperform CNN and DeepConvLSTM for eating detection using ring data. A possible explanation is that the limited dataset size and increased variability in the dataset may have hindered the Transformer's ability to learn effectively from scratch. In contrast, CNN and DeepConvLSTM demonstrated greater stability and robustness for the detection task under these conditions.

Fig 8 further illustrates the models' sensitivity to training set size. To examine this effect, we compared the detection performance of the Transformer and DeepConvLSTM models while varying the amount of training data from the full dataset down to only 0.5% of the total training samples. A fixed 15-second window with 50% overlap was applied for segmentation. As shown in the figure, both models exhibited improved performance as

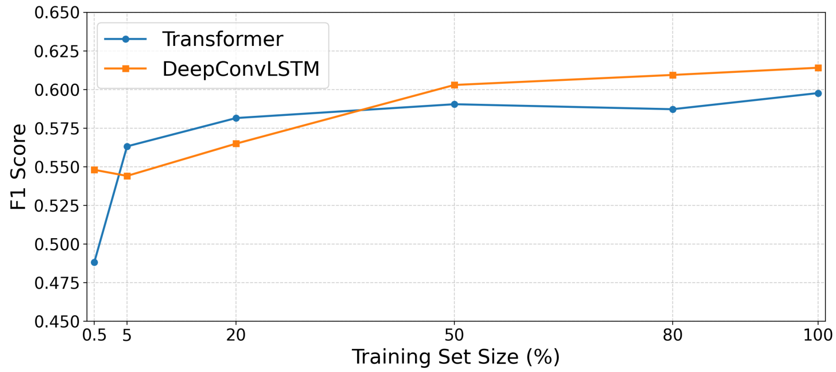

Fig. 8. Effects of training set size on model performance in our semi-naturalistic study. While the Transformer exhibited greater sensitivity under extremely low-data conditions, both models showed consistent performance gains across the tested range. This trend suggests that substantially larger datasets and/or the consideration of additional sensing modalities may be required to achieve further performance improvements.

more training data was utilize but the rate of improvement slowed once the training set reached approximately 20 − 50% of its full size, suggesting diminishing sensitivity to additional data beyond this range.

Under the same model configurations, sensor fusion achieved the best overall detection performance, with higher F1-scores compared to using either the ring or the watch alone. This highlights the effectiveness of combining motion patterns from different parts of the hand (fingers and wrist) for more accurate eating behavior detection. It suggests that finger and wrist movements offer complementary information during eating activities. Interestingly, while fusion improved F1 performance, the smartwatch alone still achieved comparable or even higher precision and recall values in some cases. This may be due to our model optimization being targeted specifically toward maximizing the F1-score. Furthermore, architectures with split input branches for ring and watch data consistently outperformed those that fused sensor streams at the raw input level, indicating that separately processing the sensor inputs before fusion captures their unique motion characteristics more effectively for our dataset collected in this study.

We also conducted a subgroup analysis based on participant age to examine potential performance differences. As shown in Fig 9, participants were divided into two groups: those younger than 60 years ($n = 6$) and those aged 60 years or older ($n = 8$). Independent samples t-tests revealed no statistically significant differences between the two age groups in Precision, Recall, or Macro-F1 with all p-values greater than or equal to 0.208, which was much higher than the 0.05 threshold needed for significance. The absence of statistically significant differences between younger and older participants suggests that age did not meaningfully affect performance across precision, recall, or macro-F1 metrics. This finding indicates that the system performed consistently across age groups, supporting its robustness and potential age fairness. However, given the small sample sizes ($n = 6$ and $n = 8$), these results should be interpreted cautiously, as the study may have been underpowered to detect subtle effects. Future work with larger and more diverse samples is needed to confirm the system's age-invariant performance.

7 IN-THE-WILD STUDY

To further evaluate the performance and generalization of ADM models utilizing ring-based sensing across diverse real-world scenarios, we conducted a fully naturalistic study with minimal constraints on participant behavior. Unlike the previous lab and semi-naturalistic settings, this experiment allowed participants to engage

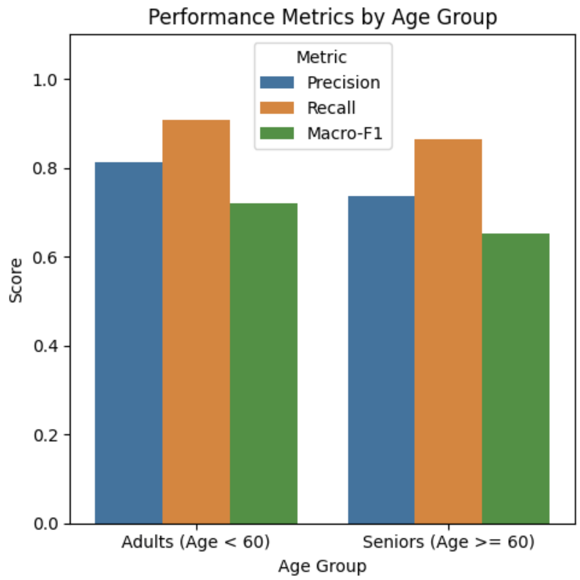

Fig. 9. Comparison of model performance between adults and seniors, showing precision, recall, and macro-F1 scores. The system performed consistently across age groups, supporting its robustness and potential age fairness.

Fig. 10. Participant self-captured photos taken before and after eating. A variety of food types and associated intake gestures are represented, reflecting the diversity of real-world eating behaviors.

in both eating and non-eating activities freely throughout the day, closely reflecting their everyday routines. While participants were instructed to take a photo before and after each eating episode for basic documentation, they were otherwise free to choose when and what to eat, as well as which non-eating activities to perform.

7.1 Participants

This in-the-wild study involved ten participants (5 male, 5 female) recruited from a university campus, with demographics and characteristics shown in Appendix C.1. All participants provided informed consent in accordance with IRB guidelines. They were equipped with sensing devices and instructed to carry out their regular daily activities over a 6-hour period.

To capture ground truth labels, participants were asked to take photos immediately before and after eating or drinking events. These images, along with their timestamps, were used to align and annotate corresponding IMU sensor data based on the system time metadata. Most sessions began between 8:30 AM and 1:00 PM and concluded between 4:30 PM and 7:00 PM to include possible meals within the day. Participants received a $10 gift card as compensation.

7.2 Experimental Procedures

For the unconstrained study, participants first signed an IRB consent form and were briefed on the study procedures. They were then outfitted with an Apple Watch and a sensing ring on a finger of their choice. An iPhone was also provided to collect Bluetooth data from both wearable devices. The iPhone and Apple Watch remained in a locked display mode, allowing participants to confirm Bluetooth connectivity without interacting with the screens. This design choice ensured data integrity by preventing unintended device interactions such as tapping or swiping. Participants used their own phones to photograph their meals and snacks before and after consumption. At the end of the session, devices were returned to a researcher, and participants submitted the food-related photos they had taken.

7.3 Data Annotations and Pre-processing

Since video recording was not feasible in this phase of the study, ground truth eating intervals were annotated by aligning the IMU sensor data with timestamps from the pre- and post-meal / snack photos taken by participants. Although this method introduces some potential temporal inaccuracies, it allowed us to minimize participant burden and preserve the naturalistic nature of the study. The collected raw IMU data was then standardized following the same procedure used in the previous study phases. The standardized data was subsequently fed into the models for evaluation. Fig 10 shows sample photos taken by participants during the in-the-wild study to serve as ground truth references. As illustrated, the food types varied widely, introducing significant diversity in intake gestures encountered during this phase of the study. In total, among the 10 participants enrolled, we collected approximately 55 hours (197,902 seconds) of eating and non-eating recordings for analysis. Within this dataset, we identified around 3 hours (10,879 seconds) corresponding to eating moments. This imbalance between eating and non-eating durations highlights the naturally skewed distribution of eating behaviors in daily life, reflecting the reality that eating occupies only a small fraction of a person's daily activities.

7.4 Evaluation for In-the-Wild Study

7.4.1 Detection of Eating Windows. For the in-the-wild study, we focused our analysis on one of the most reliable modeling setups identified in our lab and semi-naturalistic studies: the DeepConvLSTM model and its associated training configuration. As in previous phases, we compared model performance using IMU data collected from both the ring and the smartwatch, with a focus on detecting windows corresponding to eating events:

- **Lab-Only LOPO**. Models were trained using only the in-lab dataset and then directly applied to in-the-wild user data without any further fine-tuning. This setting aimed to assess how well models trained in controlled environments generalize to unconstrained, real-world scenarios.
- **Lab-Wild Fine-Tuned**. Models were initialized with weights trained on in-lab data and fine-tuned using a small portion of the in-the-wild participant's data. We used a random, stratified 20% of the participant's data for training and the remaining 80% for evaluation. We used a 20-second window size with a 50% hop and report global average metrics across all participants. This approach evaluates the benefit of model customization using minimal real-world data.
- **Lab-Wild Fine-Tuned LOPO**. Models were initialized with weights trained on in-lab data and then fine-tuned using in-the-wild participant's data in LOPO iterations. This configuration assesses how using a baseline of weights obtained from the lab data can improve the LOPO evaluation of in-the-wild data.
- **Wild-Only**. Models were trained with in-the-wild user data only. We used a random, stratified 20% of the participant's data for training and the remaining 80% for evaluation. We used a 20-second window size with a 50% hop and report global average metrics across all participants. This evaluation is useful to verify how the model performs when trained and tested with data drawn from the same in-the-wild distribution.

Table 4. Window-level eating detection performance for the in-the-wild study under different training setups. Initializing deep learning models with lab data and fine-tuning them on a small portion (20%) of in-the-wild participant data yields the best performance. Across all settings, models generally exhibit much higher recall than precision.

	Model	Class F1	Precision	Recall
Watch	Lab-Only	0.145	0.088	0.629
	Lab-Wild Fine-Tuned	**0.370**	**0.244**	**0.811**
	Lab-Wild Fine-Tuned LOPO	0.249	0.196	0.626
	Wild-Only	0.148	0.100	0.354
	Wild-Only LOPO	0.196	0.130	0.591
Ring	Lab-Only	0.142	0.092	0.432
	Lab-Wild Fine-Tuned	**0.436**	**0.333**	**0.741**
	Lab-Wild Fine-Tuned LOPO	0.214	0.140	0.566
	Wild-Only	0.125	0.103	0.208
	Wild-Only LOPO	0.112	0.071	0.362

- **Wild-Only LOPO**. Models were trained and evaluated with the leave-one-participant-out (LOPO) scheme using only the in-the-wild dataset. This approach is particularly useful to determine how well the model generalizes to unseen users.

Table 4 presents a comparison of model performance for window-level eating detection in the in-the-wild study. From the table, we can see a few observations. First, fine-tuning the models with user-specific data significantly improved eating detection performance, yielding a class-level F1-score gain of approximately 0.1 to 0.3. This improvement was consistent across both ring and watch data, as well as in both the random train–test split and the LOPO scenarios, highlighting the benefit of personalization for adapting models to individual behavior patterns. The performance gain also reflects the substantial distribution shift between controlled laboratory conditions, where participants consumed the same type of food at a consistent pace, and the more varied, self-paced eating behaviors observed in real-world settings.

Second, models trained solely on lab data and directly applied to in-the-wild participants achieved a comparable F1-score to models trained from scratch on a small portion (20%) of in-the-wild data or following the LOPO setting on the in-the-wild dataset. Notably, the lab-trained models achieved higher recall, suggesting that lab data can still provide a useful foundation for building generalizable ADM models, even under domain shifts. Interestingly, the LOPO setting, despite providing more overall training data, did not clearly outperform the *wild-only* setup. This indicates that the inter-participant domain shift is substantial, and that simply increasing training set size without incorporating subject-specific information may offer limited benefits. Finally, consistent with trends observed in earlier phases of our study, the models tended to achieve much higher recall than precision, indicating a high false positive rate in classifying eating events. This pattern suggests that while models are effective at capturing most true eating instances, they may also over-predict, mistaking non-eating activities for eating events.

To further investigate the impact of fine-tuning with user-specific data, we compared model performance using ring data across varying levels of user personalization (10%, 20%, and 30% of each participant's data). The results are shown in Table 5. As expected, increasing the amount of personalized data used for fine-tuning led to improved detection performance, demonstrating the importance of model customization for eating event detection in real-world, unconstrained environments.

Table 5. Performance of eating window detection in the in-the-wild study using different percentages of each participant's ring data for model personalization. Increasing the amount of personalized data improves detection performance.

% of Personalization	Class F1	Precision	Recall
10%	0.383	0.290	0.620
20%	0.436	0.333	0.741
30%	0.480	0.369	0.754

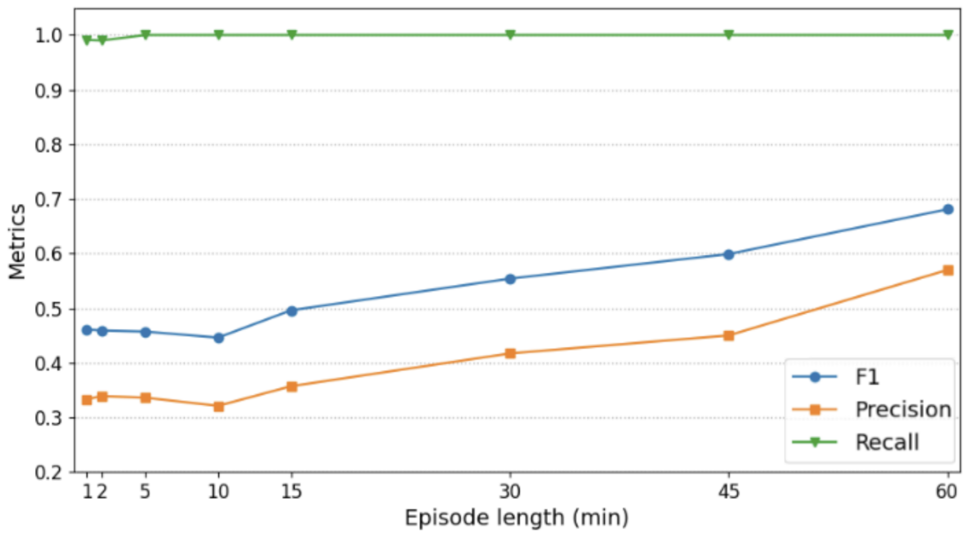

Fig. 11. Performance of dietary episode detection using participants' ring data. The results show that increasing the duration of the detection episodes improves overall episode detection performance, albeit at the cost of reduced temporal resolution.

7.4.2 Effects of Data Imbalance. One notable challenge in the in-the-wild study was the severe imbalance between eating and non-eating samples, both across participants and over time. In our dataset, the vast majority (>90%) of samples corresponded to non-eating moments. This imbalance substantially affected model training and inference performance, as deep neural networks tend to minimize global loss dominated by the majority (non-eating) class.

We addressed the imbalance through two approaches. First, we applied class-specific sample weights inversely proportional to class frequency. For example, samples from the positive (eating) class were assigned larger weights so they contributed more to the loss function, encouraging the model to better attend to minority-class examples. Conversely, non-eating samples were given smaller weights. The results reported in Table 4 are based on this weighted-loss strategy. Without weighting, both the ring and watch models trained on *wild-only* data produced near-zero precision.

Second, we explored down-sampling the negative class to match the size of the positive class. For the *wild-only* data, the resulting F1 / precision / recall scores were 0.167 / 0.101 / 0.736 for the watch and 0.106 / 0.060 / 0.600 for the ring, respectively. However, this approach did not yield clear improvements, likely because the down-sampling process substantially reduced the overall training size, limiting the variability and richness of the training data.

7.4.3 Detection of Eating Episodes. In addition to detecting dietary events at the window level (e.g., 20-second windows), we also evaluated detection at the eating *episode* level. In this work, we refer to an eating episode as a larger temporal segment encompassing multiple eating windows [35]. To model eating episodes, we clustered predicted eating moments into temporal segments based on their density over time. Specifically, adjacent eating moments that were sufficiently close to one another within a specified time window and exceeded a preset density threshold were grouped into an eating episode. We used the DBSCAN clustering algorithm [12] for this purpose. DBSCAN is well-suited for episode detection because it does not require predefining the number of clusters and can naturally form temporal clusters based on sample point density—aligning closely with the irregular, burst-like nature of eating behavior. For evaluating episode-level ADM performance, we divided each participant's full evaluation time series into fixed-duration segments representing eating episodes. An inferred episode was counted as a true positive if it overlapped with any ground truth eating episode. In our implementation of DBSCAN, we set the maximum temporal gap between points in a cluster to 10 samples (i.e., 200 seconds, given 20-second windows), and the minimum number of samples to form a cluster to 7. These values were empirically chosen based on preliminary testing with the in-the-wild data to optimize clustering outcomes. The clustering process was implemented using the Scikit-learn Python library [25].

Figure 11 illustrates our model's ADM performance at the episode level. We selected a DeepConvLSTM checkpoint that achieved above-average performance in the lab study and fine-tuned it using 20% of each participant's in-the-wild data. Performance metrics (F1 score, precision, and recall) were calculated for each user and then averaged to obtain global results. As expected, the eating detection performance improved with longer episode durations. This is because larger episode windows increase the likelihood that predicted segments will overlap with true eating episodes. However, this improvement comes at the cost of reduced temporal resolution in inference. Notably, the model achieved near-perfect recall (close to 100%) for detecting eating episodes as short as one minute. This suggests strong sensitivity in identifying true eating events. Nevertheless, consistent with our window-level detection results, precision remained relatively low, leading to an overall F1 score below 0.5 for shorter episode durations. As the episode length increased, precision improved and the F1-score rose accordingly, reaching approximately 0.7 at a 60-minute episode duration.

8 DISCUSSION

The results from the lab, semi-naturalistic, and in-the-wild studies presented in this paper highlight the potential of using IMU data from a ring wearable for eating event detection, especially when combined with smartwatch data. Our findings offer several insights into the performance of different sensing configurations, model types and parameters, and strategies for improving the accuracy of eating event detection in real-world scenarios. We discuss some of these insights in this section, and include additional ones in the Appendix (i.e., power consumption characteristics of the ring wearable (Appendix D.1) and results from a wearability survey (Appendix D.2).

8.1 Sensor Selection: Ring vs. Smartwatch

Our experiments show that eating event detection performance with ring-based IMU data is generally weaker than with smartwatch data across lab and semi-naturalistic studies, largely due to sensor placement. While the ring captures fine finger movements, the smartwatch detects broader wrist movements that are more consistent during eating and less affected by non-eating activities. This aligns with prior findings that wrist movements provide more reliable ADM signals. Although ring data alone is less robust, it remains valuable when combined with smartwatch data, as demonstrated by our sensor fusion models. A key takeaway from our findings is that ring-based eating event detection systems exhibit a higher rate of false positives (i.e., lower precision), likely due to the noisier nature of finger movements, which occur as frequently and often as variably as hand movements.

8.2 Model Performance and Sensor Fusion

The comparison of model performance across different sensor types revealed that the more sophisticated DeepConvLSTM and Transformer models generally outperformed other models for detection, including pure CNN and statistical models, even with the limited training set size. This supports the effectiveness of these models in capturing the complex, time-series nature of eating behavior data. Furthermore, our experiments also show that sensor fusion by combining data from both the ring and the smartwatch consistently improved the eating detection performance compared to using a single sensor under the same model configurations. This highlights the complementary nature of the different hand movements captured by each sensor: the more coarse wrist movements from the smartwatch and the fine-grained finger movements from the ring. The improved performance from fusion models is consistent with previous research, which has demonstrated the benefits of combining multiple sensing modalities for behavioral detection tasks.

However, it is noteworthy that more complex fusion models did not necessarily outperform simpler models, suggesting that the added complexity of multi-input branch models may not always lead to significant gains, particularly when the dataset is limited. This finding emphasizes the importance of model complexity relative to the size and diversity of the dataset, as overly complex models can lead to overfitting or diminished returns on performance.

8.3 Ring Finger Placement

One key aspect of our research was the investigation of how the placement of the ring on different fingers affects the eating detection performance. Our finger-agnostic and finger-specific analyses showed that certain fingers, such as the ring and pinky fingers, provided more consistent performance compared to others, such as the index finger. This finding suggests that finger movements are not uniform across participants and that some fingers may be more prone to variations due to their involvement in non-eating tasks. This variance is particularly pronounced in the index finger, which is often used for tasks such as typing or handling objects. This highlights the need for further exploration into how finger-specific movements contribute to ADM, and how models can be tailored to account for such variations.

8.4 Personalization for In-the-Wild Evaluation

The in-the-wild study results underscore the importance of fine-tuning models with user-specific data to account for the distribution shifts between controlled lab settings and real-world environments. Fine-tuning the models with even a small proportion of personalized data significantly improved the ADM performance. This demonstrates that while models trained on lab data can provide a useful starting point, they are not sufficient for handling the diverse and variable eating behaviors observed in the wild. Personalization is crucial for improving the performance of ADM systems in real-world applications, particularly when participants engage in a wide range of eating behaviors at their own pace.

Our results also highlight that, despite the improvements gained through fine-tuning, models trained directly on a small proportion of personalized in-the-wild data (from scratch) still lag behind in terms of performance. This highlights the ongoing challenges of developing reliable models with limited behavioral data in unconstrained and highly personalized environments, and underscores the need for effective strategies to augment or adapt models under such conditions.

8.5 Episode-level Eating Detection

The episode-level eating event detection analysis further emphasizes the advantages of using longer temporal durations to detect eating behaviors. Our findings show that, as the episode duration increases, the model detection performance significantly improves, particularly in terms of F1-score and precision. This is because a longer inference duration allows the capture of more temporal context, increasing the likelihood that inferred eating moments overlap with true eating events. However, this comes at the cost of reduced temporal resolution for the detection, which may not be suitable for applications that require high temporal precision. The balance between episode duration and resolution should be carefully considered depending on the specific use case of the eating event detection system.

9 LIMITATIONS

While our ring implementation, human subject studies, and data analyses led to new contributions to ADM, several limitations remain. First, the ring prototype was less robust than anticipated and will require further refinement. In particular, cable connections between the power source, circuit board, and Bluetooth module were prone to wear under physical stress, leading to data loss in some sessions; future designs could mitigate this by better embedding these connections within the ring enclosure to improve durability. We also observed minor data loss due to participants moving phones out of range or interference from iPhone features (e.g., NameDrop), which occasionally disrupted Bluetooth data access.

Although the study included both controlled and naturalistic settings, the dataset does not fully capture the diversity of dietary behaviors across contexts such as meal types, eating styles, age groups, and socioeconomic backgrounds. Expanding data collection to a larger and more heterogeneous participant pool will be essential for improving generalizability and enabling more powerful machine learning models. With improvements to the ring's reliability and power consumption, we plan to collect multi-day dietary data to increase ecological validity.

Lastly, in the naturalistic study, participants were asked to take photos immediately before and after eating or drinking events. These images, along with their timestamps, were used as ground truth for eating episodes. We decided on this approach to minimize burden on participants while still obtaining verifiable evidence of eating activities. However, this method has limitations; it is possible that participants ate foods and either forgot or decided not to take photos of what they consumed. It is common for wearable cameras to be used in ADM studies, passively recording all dietary and non-dietary activities. In this study, we decided not to follow this practice as it also introduces issues in naturalistic settings that are undesirable such as privacy concerns and social discomfort.

10 CONCLUSION

Rings have become increasingly popular due to their compact and unobtrusive form factor, which makes them comfortable for everyday wear while supporting a variety of sensing and interaction capabilities. This paper highlighted the challenges and opportunities of leveraging a ring wearable with inertial sensing capabilities to passively detect eating moments across various environments. Through a series of studies, we showed that while a ring wearable alone does not perform well in naturalistic settings, it can complement a smartwatch sensor at this task; the combination of both sensors leveraging fusion models improves performance. We also demonstrated that differences in finger movement patterns significantly impact eating event detection model performance based on ring motion data, with certain fingers yielding more consistent results for eating detection. Additionally, we showed that fine-tuning models with personalized data is essential for achieving robust performance in real-world settings. Despite the challenges posed by distribution shifts, our results suggest that IMU-based wearable devices, especially when used with advanced deep learning models, hold great promise for in-the-wild eating event detection.

Acknowledgments

This research was supported by the NSF CAREER award number 2239943 and NIH/NIDDK grant R01DK134546.

References

[1] Oliver Amft. 2010. A wearable earpad sensor for chewing monitoring. *Sensors, 2010 IEEE* (2010), 222–227.

[2] Faria Armin, James Jolly, James Nguyen, Lakshmi Rangaraju, Daniel Regan, Zoe Brown, Sharon O'Toole, Hollie Raynor, Leslie Brick, Elissa Jelalian, et al. 2024. MealWatcher: A New Toolset to Record Meals for the Development of Automated Energy Intake Estimation. In *2024 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 1–8.

[3] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel. 2020. FitByte: Automatic Diet Monitoring in Unconstrained Situations Using Multimodal Sensing on Eyeglasses. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[4] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 3 (2017), 1–20.

[5] Abdelkareem Bedri, Yuchen Liang, Sudershan Boovaraghavan, Geoff Kaufman, and Mayank Goel. 2022. Fitnibble: A field study to evaluate the utility and usability of automatic diet monitoring in food journaling using an eyeglasses-based wearable. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. 79–92.

[6] Brooke M Bell, Ridwan Alam, Nabil Alshurafa, Edison Thomaz, Abu S Mondol, Kayla de la Haye, John A Stankovic, John Lach, and Donna Spruijt-Metz. 2020. Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review. *NPJ digital medicine* 3, 1 (2020), 38.

[7] Shengjie Bi, Tao Wang, Nicole Tobias, Josephine Nordrum, Shang Wang, George Halvorsen, Sougata Sen, Ronald Peterson, Kofi Odame, Kelly Caine, et al. 2018. Auracle: Detecting eating episodes with an ear-mounted sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–27.

[8] Mehrab Bin Morshed, Harish Kashyap Haresamudram, Dheeraj Bandaru, Gregory D Abowd, and Thomas Ploetz. 2022. A personalized approach for developing a snacking detection system using earbuds in a semi-naturalistic setting. In *Proceedings of the 2022 ACM international symposium on wearable computers*. 11–16.

[9] Keum San Chun, Sarnab Bhattacharya, and Edison Thomaz. 2018. Detecting Eating Episodes by Tracking Jawbone Movements with a Non-Contact Wearable Sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 4.

[10] Keum San Chun, Hyoyoung Jeong, Rebecca Adaimi, and Edison Thomaz. 2020. Eating Episode Detection with Jawbone-Mounted Inertial Sensing. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 4361–4364.

[11] Yujie Dong, Jenna Scisco, Mike Wilson, E Muth, and A Hoover. 2013. Detecting periods of eating during free living by tracking wrist motion. *IEEE Journal of Biomedical Health Informatics* (Sept. 2013).

[12] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, Vol. 96. 226–231.

[13] Dawei Fan, Jiaqi Gong, and John Lach. 2016. Eating gestures detection by tracking finger motion. In *2016 IEEE Wireless Health (WH)*. IEEE, 1–6.

[14] John PA Ioannidis. 2018. The challenge of reforming nutritional epidemiologic research. *Jama* 320, 10 (2018), 969–970.

[15] David R Jacobs Jr. 2023. Challenges in research in nutritional epidemiology. In *Nutritional health: strategies for disease prevention*. Springer, 21–31.

[16] Bahador Khaleghi, Alaa Khamis, Fakhreddine O Karray, and Saiedeh N Razavi. 2013. Multisensor data fusion: A review of the state-of-the-art. *Information fusion* 14, 1 (2013), 28–44.

[17] Konstantinos Kyritsis, Christos Diou, and Anastasios Delopoulos. 2021. A Data Driven End-to-End Approach for In-the-Wild Monitoring of Eating Behavior Using Smartwatches. *IEEE Journal of Biomedical and Health Informatics* 25, 1 (2021), 22–34. https://doi.org/10.1109/JBHI.2020.2984907

[18] Basil Lin and Adam Hoover. 2019. A Comparison of Finger and Wrist Motion Tracking to Detect Bites during Food Consumption. In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 1–4. https://doi.org/10.1109/BSN.2019.8771066

[19] Jindong Liu, E Johns, L Atallah, C Pettitt, B Lo, G Frost, and Guang-Zhong Yang. 2012. An Intelligent Food-Intake Monitoring System Using Wearable Sensors. In *Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference on*. IEEE Computer Society, 154–160.

[20] Saif Mahmud, Devansh Agarwal, Ashwin Ajit, Qikang Liang, Thalia Viranda, Francois Guimbretiere, and Cheng Zhang. 2024. MunchSonic: Tracking Fine-grained Dietary Actions through Active Acoustic Sensing on Eyeglasses. In *Proceedings of the 2024 ACM International Symposium on Wearable Computers*. 96–103.

[21] Sakorn Mekruksavanich and Anuchit Jitpattanakul. 2020. Smartwatch-based human activity recognition using hybrid lstm network. In *2020 ieee sensors*. IEEE, 1–4.

[22] Christopher Merck, Christina Maher, Mark Mirtchouk, Min Zheng, Yuxiao Huang, and Samantha Kleinberg. 2016. Multimodality Sensing for Eating Recognition. *Proceedings of PervasiveHealth* (March 2016), 1–8.

[23] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16, 1 (2016), 115.
[24] Adam Paszke et al. [n. d.]. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS 2019*.
[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[26] Temiloluwa Prioleau, Elliot Moore, and Maysam Ghovanloo. 2017. Unobtrusive and wearable systems for automatic dietary monitoring. *IEEE Transactions on Biomedical Engineering* 64, 9 (2017), 2075–2089.
[27] Tauhidur Rahman, Mary Czerwinski, Ran Gilad-Bachrach, and Paul Johns. 2016. Predicting "About-to-Eat" Moments for Just-in-Time Eating Intervention. In *DH '16: Proceedings of the 6th International Conference on Digital Health Conference*. Cornell University, ACM.
[28] Philipp V. Rouast, Hamid Heydarian, Marc T. P. Adam, and Megan E. Rollo. 2020. OREBA: A Dataset for Objectively Recognizing Eating Behavior and Associated Intake. *IEEE Access* 8 (2020), 181955–181963. https://doi.org/10.1109/ACCESS.2020.3026965
[29] Addythia Saphala, Rui Zhang, and Oliver Amft. 2022. Proximity-based eating event detection in smart eyeglasses with expert and data models. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers*. 59–63.
[30] Sougata Sen, Vigneshwaran Subbaraju, Archan Misra, Rajesh Balan, and Youngki Lee. 2020. Annapurna: An automated smartwatch-based eating detection and food journaling system. *Pervasive and Mobile Computing* 68 (2020), 101259.
[31] Sougata Sen, Vigneshwaran Subbaraju, Archan Misra, Rajesh Krishna Balan, and Youngki Lee. 2015. The case for smartwatch-based diet monitoring. In *2015 IEEE international conference on pervasive computing and communication workshops (PerCom workshops)*. IEEE, 585–590.
[32] Surya Sharma, Phillip Jasper, Eric Muth, and Adam Hoover. 2020. The Impact of Walking and Resting on Wrist Motion for Automated Detection of Meals. *ACM Trans. Comput. Healthcare* 1, 4, Article 24 (Sept. 2020), 19 pages. https://doi.org/10.1145/3407623
[33] Jaemin Shin, Seungjoo Lee, Taesik Gong, Hyungjun Yoon, Hyunchul Roh, Andrea Bianchi, and Sung-Ju Lee. 2022. MyDJ: Sensing food intakes with an attachable on your eyeglass frame. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.
[34] Han Sloetjes and Peter Wittenburg. 2008. Annotation by category-ELAN and ISO DCR. In *6th international Conference on Language Resources and Evaluation (LREC 2008)*.
[35] Edison Thomaz, Irfan Essa, and Gregory D Abowd. 2015. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 1029–1040.
[36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
[37] Shuangquan Wang, Gang Zhou, Yongsen Ma, Lisha Hu, Zhenyu Chen, Yiqiang Chen, Hongyang Zhao, and Woosub Jung. 2018. Eating detection and chews counting through sensing mastication muscle contraction. *Smart Health* 9 (2018), 179–191.
[38] Wave. [n. d.]. The future of music is now within reach. Retrieved April 24, 2025 from https://genkiinstruments.com/products/wave?srsltid=AfmBOopWkETDBw4VHC9S7sJn58kADXF1N5PTM1pO-CoG7zQplOHuzqKP
[39] Gary M Weiss, Jessica L Timko, Catherine M Gallagher, Kenichi Yoneda, and Andrew J Schreiber. 2016. Smartwatch-based activity recognition: A machine learning approach. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 426–429.
[40] Gary M Weiss, Kenichi Yoneda, and Thaier Hayajneh. 2019. Smartphone and smartwatch-based biometrics using activities of daily living. *Ieee Access* 7 (2019), 133190–133202.
[41] Rui Zhang and Oliver Amft. 2020. Retrieval and Timing Performance of Chewing-Based Eating Event Detection in Wearable Sensors. *Sensors* 20, 2 (2020), 557.

A CONTROLLED LAB STUDY

A.1 Study Activities

(1) **Eat a meal (not limited in time)**: Participants were asked to eat the meal that we prepared beforehand for the study. Participants ate at their regular eating pace. The food we prepared was standardized purchased from markets, with a pack of instant rice containing rice and meat (255 grams), a bag of popcorn (0.5 ounces), a cup ice cream (3 fluid ounces). In special cases, such as for participants who are vegetarian or allergic to soy source, we prepared instant rices of a similar size but without such ingredients.

(2) **Conversation with the researcher or participant (2 minutes)**: Participants conversed indoors with a researcher about a topic they were interested in or had knowledge about.

(3) **Telephone call with a researcher or assistant (1 minute)**: The participant placed a telephone call to a remote researcher. The script did not specify any requirements regarding the location or content of the conversation, allowing the call to be performed in a completely natural way.

(4) **Type a paragraph on a laptop (2 minutes)**: The participant was asked to type a 125 word paragraph in a document editor like Microsoft Word.

(5) **Type a paragraph on a mobile phone (2 minutes)**: The participant was asked to type the same paragraph as they did on the laptop but on their phone, in their iPhone Notes or equivalent Android app.

(6) **Use a mobile phone normally (2 minutes)**: Participants were asked to use their phone as they normally would (e.g., check social media, play games, surf the web, etc.).

(7) **Write on paper (2 minutes)**: The participant was asked to write the same paragraph as they did on the laptop and phone, but on a blank sheet of paper.

(8) **Fold laundry (2 minutes)**: Participants were asked to fold and stack clean laundry, and encouraged to fold and stack as they normally would.

(9) **Simulate washing hands (30 seconds)**: Participants stood at a sink and performed the motions associated with hand washing, such as turning on water, pumping soap, and scrubbing their hands.

(10) **Comb hair (30 seconds)**: Participants combed their hair using a provided individual use comb / brush while standing up.

(11) **Walk around (2 minutes)**: Participants walked indoors with a researcher through the building hallways in a tour around the building.

(12) **Brush teeth (not limited in time)**: Participants brushed their teeth as they normally would at a sink, using provided individual use brushes and toothpaste.

A.2 Ground Truth and Annotation

To obtain ground truth for eating and non-eating activities, all participants were recorded on video during their session. For activities such as *walk around*, the researcher carried the camera and followed the participant to ensure that all activities were captured. Participants were instructed to firmly clap three times at the beginning of each target activity to create clear reference points for our data synchronization and segmentation afterwards. These clapping events were clearly visible as spikes in the IMU sensor signals and were used to align data streams from the ring and the smartwatch.

Activity segmentation and annotation were performed using the ELAN tool [34]. To extract the timestamps of eating and non-eating activities from the collected sensor data, a researcher manually reviewed the full-length video recordings for each participant. Activities were annotated at the activity level by marking the start and end times of each distinct behavior based on the video recordings. Following the predefined activity list, we identified 12 distinct activity classes along with an additional "transition" class to represent the periods between adjacent activities. Based on the annotated timestamps from the video ground truth, we segmented the sensor data streams from the ring and the watch separately, assigning a single activity label to each sensor segment falling within the corresponding time window.

Since our primary focus was detecting eating behaviors, all non-eating activity segments were grouped into a single negative class for binary classification, including the transition class. Across participants, eating activities accounted for approximately 47% of the total recorded sensor duration. Due to slight mismatches in start times between the ring and the watch, as well as occasional data loss from connection failures, the segmented sensor streams were not always of identical duration. To ensure proper alignment for computational model inputs, we compared the clapping signal spikes across devices and made manual adjustments as needed. Specifically, sensor data streams for each participant were trimmed or zero-padded at the beginning of the signals to synchronize the timelines. For most participants, this adjustment of the segments was largely negligible, averaging around 3 seconds out of the averaged entire duration of 40 minutes per participant.

A.3 Classifiers

For the Random Forest classifier, we set the number of estimators to 150 and limited the maximum number of features to 3. For the SVC, we used a radial basis function (RBF) kernel, which yielded the best performance under our LOPO evaluation based on preliminary testing. For both models, we first extracted statistical features from the raw sensor signals, resulting in a 30-dimensional feature vector per input sample. These features included the mean, standard deviation, skewness, kurtosis, and sum of energy across each X, Y, and Z axis for both the accelerometer and gyroscope signals within a given time series window. For this analysis, we used the Python scikit-learn library [25].

The CNN architecture used in this analysis consisted of three convolutional layers followed by three fully connected layers. The model inputs were six-dimensional acceleration and gyroscope signals from either the ring or the watch, stacked along the feature dimension to form a 2D input of shape – *batch size, 6, temporal dimension*. The size of the temporal dimension was determined by the input window size selected for each experiment, ranging from 2 to 30 seconds. To control the model size across different input window lengths, we applied temporal pooling after the final convolutional layer to squeeze the temporal dimension to 1, regardless of the input window size. The resulting features were then flattened and fed into the fully connected layers to produce probability-level class outputs for model inference. Each convolutional layer in the CNN was followed by a ReLU activation and batch normalization. Similarly, each fully connected layer was activated by ReLU and regularized with a dropout rate of 0.8. The kernel size for each convolutional layer was set to (5, 2), with a stride of (2, 1) and no additional padding. The output dimensions of the fully connected layers were 128, 256, and 2, corresponding to the binary classification task of eating event detection.

In addition to CNNs, we also evaluated ADM performance using the DeepConvLSTM model [23], which incorporates additional LSTM layers to better capture the temporal dynamics of the input signals. Similarly to the CNN configuration, the DeepConvLSTM models processed six-dimensional stacked acceleration and gyroscope signals as input. We applied two LSTM layers, each with a hidden size of 16 units, followed by the same fully connected layers used in the CNN architecture.

Finally, our Transformer model was implemented with 8 layers of Transformer encoders followed by a sequence of three fully connected layers. For the multi-head attention mechanism, we used 8 attention heads and set the feedforward network dimension to 256. To prepare the inputs, we initialized a learnable positional embedding of size 184, which was directly added to the input sequence. The full sequence output from the encoder was used for classification. The dimensions of the fully connected layers were kept consistent with those used in our CNN and DeepConvLSTM models, with 128, 256, and 2 neurons respectively, each activated by ReLU. Before passing the encoder outputs into the fully connected layers, we applied average pooling along the temporal dimension to enable the model to handle input sequences of varying lengths while maintaining similar model complexity.

For all neural network models, we used the Adam optimizer with a learning rate of 10^{-4}, betas of (0.9, 0.999), an epsilon value of 10^{-8}, and no weight decay. The loss function was set to cross-entropy loss. Following our LOPO evaluation scheme, model training incorporated early stopping with a patience of 5 epochs to save a model checkpoint if no improvement was observed on the validation set, and within a maximum of 100 training epochs. These hyper-parameters were selected based on our preliminary experiments showing that a patience of 5 epochs was sufficient for convergence across all participants. The batch size was set to 128 for training. To ensure reproducibility and comparability across models, we fixed random seeds during training. All the neural network models were implemented using the PyTorch [24] framework.

B SEMI-NATURALISTIC STUDY

B.1 Participants

Table 6. Overview of participants in the semi-naturalistic study. Participants were recruited to ensure diversity across age, ethnicity, and occupation. Each participant received USD $100 compensation.

Participant	Gender	Age	Ethnicity	Occupation	Handiness
1	M	29	Caucasian	Realtor	R
2	F	61	Black	Operations Manager	R
3	M	33	Caucasian	Sales Development Representative	L
4	M	68	Black	Retired Administrator	R
5	F	61	Caucasian	Project Manager	L
6	F	39	Caucasian	Stay-At-Home Parent	R
7	M	63	Hispanic	Retired Research Coordinator	R
8	F	68	Caucasian	Recovery Coach	R
9	M	26	Black	Unemployed	R
10	F	65	Caucasian	Retired	R
11	M	66	Caucasian	Retired	R
12	F	46	Other	Inventory Control	R
13	M	60	Caucasian	Brand Ambassador	L
14	M	60	Hispanic	Project Management Specialist	L
15	M	40	Other	Chief Marketing Officer	R
16	M	69	Asian	Customer Service	R
17	F	58	Caucasian	Stay-At-Home Parent	R
18	M	56	Caucasian	Photographer	R
19	F	38	Caucasian	Small Business Owner	L
20	F	54	Caucasian	Project Manager	R

C IN-THE-WILD STUDY

C.1 Participants

Table 7. Overview of the participants in our in-the-wild study.

Participant	Gender	Age	Ethnicity	Handiness	Chosen Finger
1	F	21	Persian	R	Ring
2	F	20	Asian/White	L	Ring
3	M	31	Asian	R	Middle
4	F	21	Asian	L	Ring
5	M	25	Asian	R	Index
6	M	27	Asian	R	Thumb
7	M	21	Asian	R	Ring
8	F	21	Asian/White	R	Ring
9	F	21	Asian	R	Index
10	M	20	White	R	Pinky

D RING WEARABLE

D.1 Power Consumption

Battery endurance tests comparing the prototype ring, a smartwatch, and a smartphone revealed meaningful differences in power efficiency. The ring operated continuously for 23 hours 50 minutes, the smartwatch for 19 hours 30 minutes, and the smartphone required a full recharge during the same period. During testing, the smartphone display remained on at low brightness to maintain uninterrupted Bluetooth data streaming, and the smartwatch screen was periodically activated to verify connectivity. These results indicate that the ring supports full-day operation and is not the limiting device in multi-device data collection scenarios. Although the system was capable of running for longer durations, we selected 6-hour study sessions to align with participant availability, which typically spanned morning through late afternoon.

D.2 Wearability and Comfort

As seen in Figure 12, in a voluntary post-study survey of 16 participants, the majority reported that the ring was comfortable to wear and did not interfere with their eating or non-eating activities. Specifically, all but one participant agreed or strongly agreed that the ring did not interfere with eating, and most participants indicated similar non-interference with general daily tasks. Comfort ratings were largely positive, with over two-thirds of participants agreeing or strongly agreeing that the ring was comfortable during the study. However, responses were more mixed regarding long-term wearability: only about half of participants stated they would be happy wearing the ring every day, with several expressing neutral or disagreeing responses. Feelings of self-consciousness were generally low, though a few participants noted mild awareness of the ring in public settings. These findings suggest that while the prototype was well-tolerated and functionally unobtrusive, future iterations should prioritize improved ergonomics and aesthetics to enhance comfort and social acceptability for everyday use. We expect that further iterations will bring the long-term wearability of the prototype on par with existing health tracking rings in comfort, aesthetic and design.

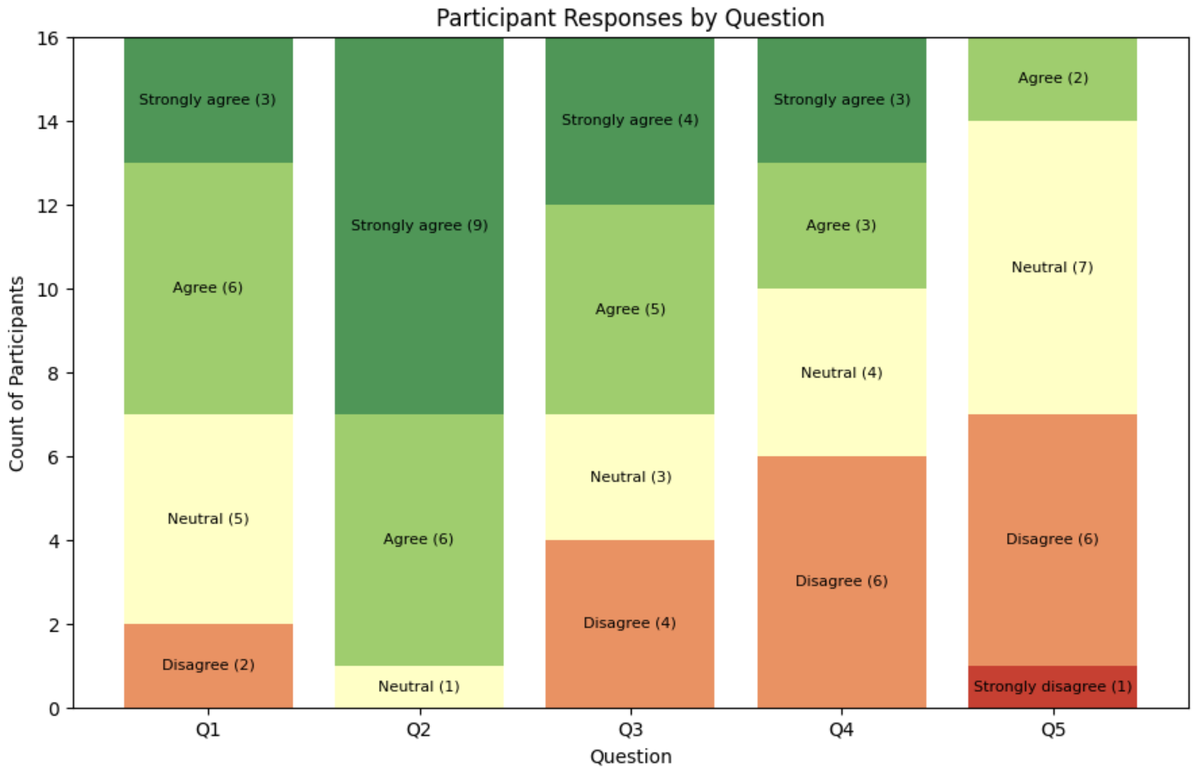

Fig. 12. Post-study participant responses by question. **Q1:** Wearing the ring was comfortable. **Q2:** Ring did not interfere with eating activities. **Q3:** Ring did not interfere with non-eating activities. **Q4:** I would be happy wearing the ring every day. **Q5:** I felt self-conscious wearing the ring in public.