DietWatch: Towards Low-effort Fine-grained Dietary Monitoring via Smartwatch in Open-World Scenarios

Zhen Hou
Indiana University-Purdue University
Indianapolis, Indiana, USA
zlhou@purdue.edu

Yucheng Xie*
Yeshiva University
New York, USA
yucheng.xie@yu.edu

Feng Li
Purdue University
Indiana, USA
fengli@purdue.edu

Abstract

Dietary behaviors play a pivotal role in preventing chronic diseases and promoting overall health. The widespread adoption of smartwatches offers a promising platform for continuous dietary monitoring in real-world scenarios. However, existing smartwatch-based approaches struggle with challenges in open-world scenarios, including dynamic interference, unrestricted eating gestures, and previously unseen daily behaviors. In this work, we present *DietWatch*, an open-world dietary monitoring system that utilizes a commercial smartwatch to capture fine-grained dietary behaviors. First, *DietWatch* mitigates acoustic and inertial interferences via Conv-TasNet and Bi-GRU. Then, *DietWatch* leverages a contrastive learning-based method to analyze human gestures, ensuring accurate eating gesture identification even in the presence of previously unseen daily gestures. Furthermore, a clustering algorithm is designed to estimate dietary time, while an attention-based multimodal fusion method is employed to classify food categories. Experimental evaluations in open-world settings demonstrate that *DietWatch* achieves 79.75% tIoU for eating time detection, 86.0% accuracy in food classification.

Keywords

Dietary Monitoring, Contrastive Learning, Multimodal Fusion, Open-World Recognition

ACM Reference Format:
Zhen Hou, Yucheng Xie, and Feng Li. 2025. DietWatch: Towards Low-effort Fine-grained Dietary Monitoring via Smartwatch in Open-World Scenarios. In *ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE '25), June 24–26, 2025, New York, NY, USA*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3721201.3721403

1 Introduction

Dietary behavior has a profound impact on human health. According to reports from the World Health Organization, poor dietary habits are key risk factors for chronic conditions, including obesity, diabetes, and cardiovascular diseases [23]. Given the significant health risks associated with poor diet, it becomes essential to monitor and understand individual dietary behaviors—including dietary time and food category selections. These detailed insights are key to uncovering underlying dietary issues, providing a basis for developing effective health intervention strategies, and delivering personalized nutritional advice. For instance, identifying the timing of dietary intake could reveal snacking habits, which in turn can uncover hidden contributors to weight gain [7]. In addition, understanding food categories consumed is crucial for nutritionists to create balanced and personalized dietary plans tailored to individual needs [5].

Traditional dietary behavior analysis approaches rely on self-reporting tools, including food diaries and 24-hour recall questionnaires [17]. To overcome these limitations, researchers propose employing wearable devices for dietary monitoring [3, 8, 13, 14, 19]. These devices are worn on the body. They facilitate continuous dietary-related data collection and enable automated dietary monitoring. However, these wearable devices face challenges for dietary monitoring in daily scenarios, as they are often cumbersome and can make users feel self-conscious in social settings, particularly in public dining environments. Furthermore, the high costs associated with specialized components, along with the complexity of use for general users, make them impractical for widespread use [14, 19]. Due to the widespread adoption and relatively low cost, smartwatches have emerged as an ideal platform for daily dietary monitoring [4]. Moreover, smartwatches integrate multiple sensors, such as inertial measurement units (IMUs) and microphones, allowing for comprehensive dietary feature extraction from inertial and acoustic domains. Existing smartwatch-based eating monitoring methods have shown initial success in tracking various aspects of dietary behaviors [9, 12, 16, 18, 21, 22]. However, these approaches remain limited in scope, providing only partial tracking of dietary behavior. A more pressing challenge is that most existing smartwatch-based eating monitoring methods only perform optimally in controlled conditions.

To bridge the gap between controlled settings and the complexities of real-world settings, the concept of *open-world* scenarios has been proposed. The open-world scenarios bring about the presence of dynamic interference, unrestricted eating gestures, and unseen daily activities. To achieve fine-grained dietary monitoring in open-world scenarios, several key challenges should be addressed: (1) Dynamic interferences, including background noises (e.g., music in public environments), as well as motion-induced noise (e.g., head scratching), are unpredictable and can significantly affect data collected from the smartwatch, leading to errors in dietary behavior recognition. Mitigating the impact of dynamic interference is essential for enhancing the robustness of dietary monitoring in

*Yucheng Xie is the corresponding author.

CHASE '25, June 24–26, 2025, New York, NY, USA

ACM ISBN 979-8-4007-1539-6/2025/06
https://doi.org/10.1145/3721201.3721403

open-world conditions. (2) Consistent features of intake gestures must be extracted, as motion range and speed can vary even when the same gesture is performed across different scenarios or emotional states. In addition, it is impractical to expect users to provide samples for every potential non-eating gesture during the system training phase. Open-world dietary monitoring system needs to accurately identify eating time even when unseen daily gestures are encountered during the testing phase. (3) A more fine-grained and comprehensive dietary monitoring system is required, encompassing dietary time recognition and food category classification, to assist accurate dietary assessments and support personalized nutritional recommendations for users in open-world settings.

To address these challenges, we propose a low-effort open-world dietary monitoring system *DietWatch*, which achieves fine-grained dietary behavior monitoring using a commercial smartwatch. (1) We develop an interference mitigation method for efficiently reducing open-world interference in both acoustic and inertial domains. To address dynamic background noise, we develop a time-domain convolutional neural network (Conv-TasNet) [10] to build adaptive masks for the accurate extraction of dietary-related acoustic signals. We also develop a lightweight Bidirectional Gated Recurrent Unit (Bi-GRU) [6] and Least Mean Square (LMS) filter [20] to effectively extract motion signals caused by eating behaviors. (2) To extract consistent features of eating behaviors and recognize unseen daily gestures, we develop a contrastive learning-based eating gesture identification method. By strategically increasing the distance between eating and non-eating gestures in the feature space and decreasing it for eating ones, we significantly enhance *DietWatch*'s ability to identify eating gestures in open-world scenarios. (3) To achieve fine-grained dietary monitoring, we design a clustering algorithm to estimate dietary time based on eating gesture identification. We further develop an attention-based multimodal fusion strategy that combines inertial and acoustic features, allowing for food category classification. Our main contributions are summarized as follows:

- We propose *DietWatch*, a low-effort dietary monitoring system that provides fine-grained dietary information in open-world scenarios using a commercial smartwatch.
- We develop an interference mitigation approach to mitigate both acoustic and inertial interferences, and develop a contrastive learning-based method for recognizing dietary gestures without imposing restrictions on users' daily and eating activities. In addition, we design a multimodal feature fusion strategy with a multi-head attention mechanism to effectively combine dietary features from different domains for food category classification.
- We evaluate *DietWatch* with 23 activity types and 40 food types in open-world scenarios. Experimental results demonstrate that *DietWatch* achieves robust performance with 79.75% temporal Intersection over Union (tIoU) for eating time detection and 86.0% accuracy for food type classification.

2 DietWatch Design

DietWatch is a fine-grained dietary monitoring system design for open-world scenarios. The core of *DietWatch* consists of 4 components, including dynamic interference mitigation, eating gesture identification, multimodal feature fusion, and fine-grained dietary behavior derivation. As shown in Figure 1, *DietWatch* takes time-series inertial and acoustic readings from smartwatches as input. We develop interference mitigation methods to efficiently reduce noise in both domains. In the acoustic domain, we develop a Conv-TasNet to filter out background noise. In the inertial domain, we develop a Bi-GRU network and a LMS filter to mitigate motion-induced noise. To accurately recognize eating gestures among diverse daily activities, we develop a contrastive learning-based approach to maximize the distances between eating and non-eating gestures in the feature space, enabling our system to robustly differentiate eating gestures from unseen daily activities. To leverage the complementary strengths of multiple data sources (e.g., motion and acoustic features) and achieve robust dietary behavior recognition, we design a multimodal feature fusion framework based on attention mechanism. Based on the above modules, Our system enables simultaneous monitoring of dietary time detection and food category classification.

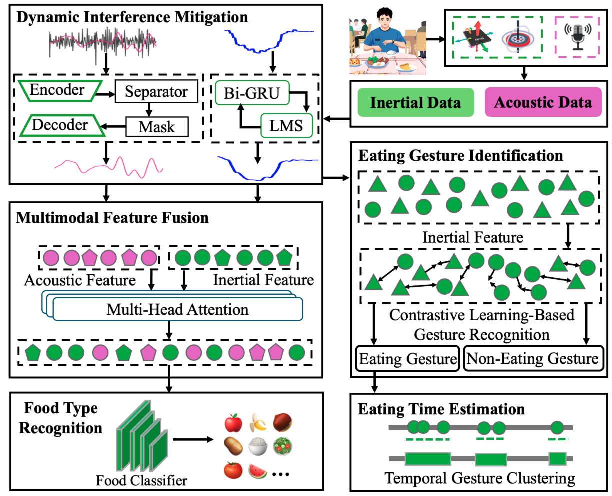


Figure 1: DietWatch system design.

2.1 Dynamic Interference Mitigation

Inertial Signal Processing. In open-world scenarios, inertial signals collected from smartwatches are inevitably affected by motion-induced noise, including unstructured hand motions (e.g., head scratching or repositioning utensils) or environment-induced vibrations (e.g., those generated by walking or traveling in a moving vehicle). To address this challenge, we develop a denoising approach based on Bi-GRU and LMS. We choose Bi-GRU because eating gestures exhibit distinct temporal patterns, such as periodic wrist motions, which Bi-GRU can effectively capture by leveraging its ability to process sequential data in both forward and backward directions. In contrast, unstructured hand motions, such as head scratching, lack temporal coherence, often appearing as abrupt, irregular signal deviations. Bi-GRU learns to differentiate these patterns during training, focusing on the structured periodicity of eating gestures while ignoring unstructured noise signals.

To further reduce the impact of environment-induced vibrations, we develop a LMS adaptive filter, which is effective for dynamic noise due to its ability to iteratively adapt to signal variations in

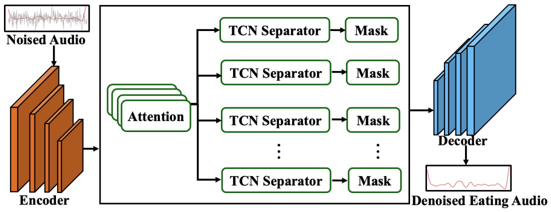


Figure 2: The multi-branch Conv-TasNet architecture for acoustic signal denoising.

real time. The LMS algorithm operates by iteratively adjusting its filter weights to minimize the mean squared error (MSE) between the noisy signals and the reference signals. The reference signals are collected in advances from stable settings (e.g., inertial signals of eating gesture without any environment introduced vibrations). The adaptive filter leverages cross-correlation between the noisy signals (e.g., inertial signals of eating gesture in a moving vehicle) and the reference signals to iteratively refine its weights, thereby mitigating noise while preserving the essential temporal and spatial characteristics of eating gestures.

Acoustic Signal Processing. Analyzing eating-related sounds is crucial for eating behavior monitoring, as these signals provide key information about food texture. However, acquiring high-quality eating sounds in open-world scenarios is challenging due to environmental noise, particularly in public spaces where noise levels vary unpredictably in intensity (50–70 dB) and frequency (100 Hz-10 kHz). Traditional frequency-domain denoising methods, such as Wiener filtering and spectral subtraction, are inadequate as they often distort the signal and fail to preserve subtle acoustic features critical for food type recognition.

To overcome these limitations, we design a time-domain denoising approach based on a multi-branch Conv-TasNet architecture [10]. As shown in Figure 2, we develop an encoder that transforms input waveforms into high-dimensional feature representations using convolutional filter banks. These features are dynamically weighted by an attention mechanism, ensuring that texture-specific patterns are emphasized. The attention-weighted features are routed to different Temporal Convolutional Network (TCN) branches, each specialized in handling a particular texture category. The key to background noise mitigation lies in the mask generated by each TCN branch. The mask isolates the texture-specific acoustic features by selectively enhancing components corresponding to the desired eating sounds while attenuating noise components that do not match the target texture.

2.2 Contrastive Learning-based Eating Gesture Recognition

Eating gestures refer to the distinct motion patterns involved in food consumption, including the manipulation of eating utensils, food preparation movements (e.g., stirring, cutting), and hand-to-mouth trajectories [15]. Recognizing eating gestures in open-world scenarios presents unique challenges due to unrestricted eating gestures and unseen daily behaviors. Traditional supervised learning methods often fail to generalize in open-world scenarios due to the labor-intensive task of collecting exhaustive samples for every possible non-eating activity. To address this issue, we propose a contrastive learning framework to extract consistent features that encapsulate the shared characteristics of diverse eating gestures while effectively discriminating them from visually and kinematically similar non-eating activities [16].

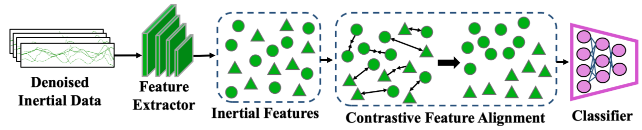


Figure 3: The contrastive learning-based framework for eating gesture recognition.

Our training framework consists of two phases to ensure robust and generalizable gesture recognition. In the first phase, training pairs are constructed such that variations of the same eating gesture are treated as positive pairs, while pairs comprising eating gestures and similar non-eating activities are designated as negative pairs. In the second phase, a GRU-based temporal feature extractor is employed to model the sequential dependencies in the denoised inertial signal and capture motion dynamics essential for gesture identification. The extracted temporal features are then processed by a contrastive feature alignment process, which minimizes the cosine distance for positive pairs to promote intra-class consistency while enforcing a margin for negative pairs to maximize inter-class separability, addressing the overlap between eating and non-eating behaviors. As shown in Figure 3, the feature extraction module processes the inertial data to obtain representative embeddings, which are then refined through contrastive feature alignment to enhance the discriminative power of the learned features by maximizing the separation between eating and non-eating gestures. Finally, the aligned features are fed into a classifier for gesture recognition, enabling robust discrimination between eating and non-eating activities.

2.3 Multimodal Feature Fusion

Multimodal data integration, combining inertial and acoustic signals, enhances fine-grained food category classification in open-world scenarios. Inertial data captures hand motion dynamics, while acoustic signals reflect auditory patterns of food texture. By leveraging these complementary strengths, multimodal approaches address the limitations of single-modal systems, such as difficulties in distinguishing foods with overlapping features [1, 2, 11]. For example, inertial data can differentiate noodle dishes with similar sounds by capturing utensil use and associated hand gestures.

To achieve effective feature fusion, we propose an attention-based framework that integrates acoustic and inertial features. Initially, denoised acoustic signals are converted to time-frequency representations using Short-Time Fourier Transform. Temporal dependencies are captured through GRU-based feature extractors tailored for each modality. The resulting outputs, obtained by concatenating the inertial and acoustic features, are projected into query (Q), key (K), and value (V) matrices through learned linear transformations. The cross-modal attention mechanism computes attention weights as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{1}$$

Table 1: Activity categories in dataset

Category Groups	Activity Types
Eating Activities (7)	Using Spoon, Using Fork, Using Chopsticks, Drinking from Cup, Drinking from Bottle, Cutting Food, Eating by Hand
Non-eating Activities (16)	Phone Call, Standing, Shaving Face, Cleaning Ears, Conversation, Whistling, Nail Biting, Teeth Brushing, Sitting, Head Scratching, Teeth Picking, Hair Combing, Walking, Keyboard Typing, Book Reading, Phone Browsing

where d_k is the dimensionality of the key matrix. These attention weights ensure that the model focuses on the most relevant features for food classification. For example, the framework can prioritize acoustic features when food textures vary significantly or inertial features when gestures provide better differentiation.

2.4 Fine-grained Dietary Behavior Derivation

Eating Time Derivation. Eating time refers to the intervals during which eating behaviors occur, and it is derived by analyzing the temporal patterns of continuous eating gestures identified by the proposed gesture recognition module. We propose a clustering algorithm for eating time derivation. Specifically, for continuous eating behavior recognition, the collected data is segmented into 3-second windows, each classified independently. Given a sequence of identified eating gestures $G = \{g_1, g_2, \ldots, g_n\}$, where each gesture g_i is a single eating action (e.g., eating with a fork) with a corresponding timestamp t_i, an eating period C_k is defined as a continuous time interval containing multiple temporally adjacent eating gestures. Formally:

$$C_k = \{g_i, g_{i+1}, \ldots, g_j \mid t_{m+1} - t_m < \theta, \forall m \in [i, j-1]\}, \quad (2)$$

where θ is the maximum allowed time gap between consecutive gestures within the same eating period. Based on typical meal patterns, we empirically set θ to 25 seconds [18]. Consecutive gestures separated by time gaps exceeding θ are assigned to different eating periods. To ensures reliable eating moment detection, we require each eating moment to contain at least 4 gestures per minute and last for a minimum duration of 3 minutes.

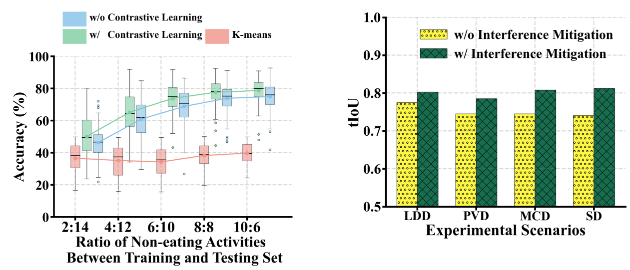
Food Type Classification. Food type classification is analyzed within the identified eating periods. Our food type classifier takes the latent feature representations derived from the feature fusion module as input and outputs food categories. The latent feature representations are extracted from the multimodal fusion features module. The classification network comprises two fully connected layers with ReLU activation, followed by a softmax-based output layer that predicts probabilities for C food categories. The model parameters are trained using a cross-entropy loss function. To address ambiguous cases, a confidence threshold (i.e., 0.85) is used. In addition, a majority voting strategy is applied over a 3-second sliding window, reducing sporadic misclassifications and ensuring stable predictions.

3 Performance

Experimental Setup and Evaluation. We develop a *DietWatch* prototype on two commercial smartwatches (Samsung Galaxy Watch 5 and Google Pixel Watch 2), both equipped with 6-axis IMU sensors and microphones (41 kHz sampling rate). We evaluate across 4 real-life scenarios: *1) Low Disturbance Dining (LDD)*: Individual eating in a private room with small background noise (<30 dB noise) and

Table 2: Food category groups in dataset

Category Groups	Food Types
Staple Foods (8)	Rice, Corn, Bread, Crackers, Fried buns, Boiled potatoes, Potato, Cereal
Hard/Crispy Foods (8)	Chips, Cookies, Fries, Peanut, Pecans, Gum, Chocolate, M&M's
Soft Foods (8)	Yogurt, Pudding, Cake, Egg, Ice cream, Mousse, Marshmallow, Meat
Fruits/Vegetables (8)	Apple, Pear, Orange, Tangerine, Grape, Carrot, Tomato, Cucumber
Beverages (8)	Water, Tea, Coffee, Milk, Juice, Cola, Wine, Parsley



(a) Performance of eating gesture recognition under different ratios.

(b) Performance of eating time derivation under different scenarios.

Figure 4: Performance evaluation of eating gesture recognition and eating time derivation.

without other daily activities; *2) Public Venue Dining (PVD)*: Eating in coffee shops during business hours (50-70 dB noise); *3) Mobile Context Dining (MCD)*: Eating portable foods while walking (4-5 km/h); *4) Social Dining (SD)*: Dining with 1-3 others. This scenario introduces acoustic interference from conversations (60–65 dB) and frequent transitions between eating and non-eating activities.

We conducted experiments with 18 participants (11 males, 7 females, aged 20-30) wearing smartwatches on their dominant wrists. Our dataset spans four months, comprising over 400 eating sessions (10 minutes each), 7 eating activities and 16 non-eating activities (Table 1), and 40 food types across 5 categorie groups (Table 2). Ground truth was captured via a GoPro Hero 12. For evaluation, we use *Accuracy* and *Temporal Intersection over Union (tIoU)*. *tIoU* is defined as: $tIoU = \frac{|D \cap G|}{|D \cup G|}$, where D representing detected eating periods and G denoting ground truth periods.

Performance of Eating Time Derivation. To evaluate the robustness of our system against unseen daily behaviors, we varied non-eating activity categories between the training and testing datasets. Figure 4(a) shows our method outperforms two baseline methods across all configurations. Even under the most challenging configuration, where the training set includes only 2 non-eating activity categories and the testing set contains 14, our approach achieves the best recognition *accuracy* of 82.40% compared with baselines. Figure 4(b) demonstrates eating time derivation across 4 real-life scenarios. With the assist of our interference mitigation method, *tIoU* improves by over 5.25% across all scenarios, reaching 79.75% in average, confirming effectiveness of our method.

Performance of Food Category Classification. To evaluate the food classification capabilities, we compared our multimodal fusion approach against single-modality baselines. Our framework achieved an *accuracy* of 86.0%, significantly outperforms inertial-only (35.6%) and acoustic-only (52.4%) models. Among food categorie groups, Hard/crispy foods achieved highest *accuracy* at 91.3%, followed by staple foods (86.5%), fruits/vegetables (84.9%), soft foods

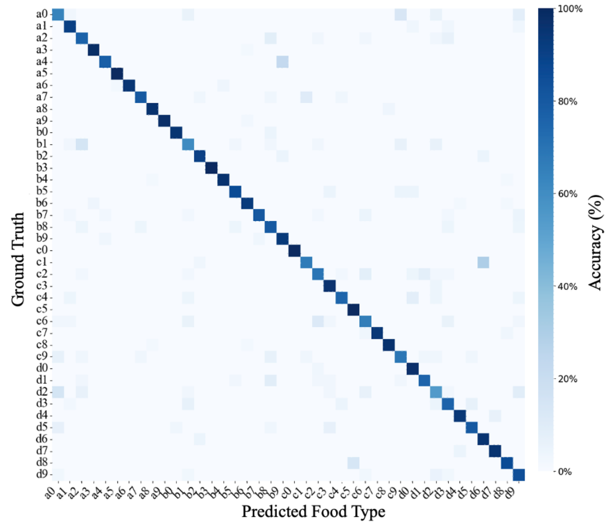


Figure 5: Performance of food classification (40 food types across 5 groups).

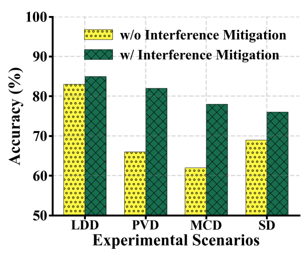


Figure 6: Performance of food classification under different real-life scenarios.

(84.0%), and beverages (83.5%). The confusion matrix in Figure 5 reveals robust category separation, with minimal cross-category confusion rate of just 0.4%. To evaluate the robustness of food category classification in open-world scenarios, we assess the food classification performance with and without the interference mitigation. As shown in Figure 6, with the interference mitigation enabled, the system demonstrates consistent performance improvements, with *accuracy* increasing by more than 10% in most scenarios.

4 Conclusion

In this paper, we propose *DietWatch*, a fine-grained dietary monitoring system designed for open-world scenarios using a commercial smartwatch. By addressing challenges such as dynamic interference, unrestricted eating gestures, and previously unseen daily behaviors, DietWatch offers a scalable and practical solution for dietary monitoring, paving the way for advancements in personalized nutrition and public health interventions.

References

[1] 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.

[2] Shengjie Bi and David Kotz. 2022. Eating detection with a head-mounted video camera. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*. IEEE, 60–66.

[3] 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.

[4] Demand Sage. 2024. Smartwatch Statistics 2024: Worldwide Market Data. https://www.demandsage.com/smartwatch-statistics Accessed: 15-January-2024.

[5] Michael J Gibney and Marianne C Walsh. 2013. The future direction of personalised nutrition: my diet, my phenotype, my genes. *Proceedings of the Nutrition Society* 72, 2 (2013), 219–225.

[6] Shipeng Han, Zhen Meng, Xingcheng Zhang, and Yuepeng Yan. 2021. Hybrid deep recurrent neural networks for noise reduction of MEMS-IMU with static and dynamic conditions. *Micromachines* 12, 2 (2021), 214.

[7] Rebecca M Leech, Anna Timperio, Katherine M Livingstone, Anthony Worsley, and Sarah A McNaughton. 2017. Temporal eating patterns: associations with nutrient intakes, diet quality, and measures of adiposity. *The American journal of clinical nutrition* 106, 4 (2017), 1121–1130.

[8] Jindong Liu, Edward Johns, Louis Atallah, Claire Pettitt, Benny Lo, Gary Frost, and Guang-Zhong Yang. 2012. An intelligent food-intake monitoring system using wearable sensors. In *2012 ninth international conference on wearable and implantable body sensor networks*. IEEE, 154–160.

[9] Yadnyesh Y Luktuke and Adam Hoover. 2020. Segmentation and recognition of eating gestures from wrist motion using deep learning. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 1368–1373.

[10] Yi Luo and Nima Mesgarani. 2019. Conv-tasnet: Surpassing ideal time–frequency magnitude masking for speech separation. *IEEE/ACM transactions on audio, speech, and language processing* 27, 8 (2019), 1256–1266.

[11] Mark Mirtchouk, Drew Lustig, Alexandra Smith, Ivan Ching, Min Zheng, and Samantha Kleinberg. 2017. Recognizing eating from body-worn sensors: Combining free-living and laboratory data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–20.

[12] Mehrab Bin Morshed, Samruddhi Shreeram Kulkarni, Richard Li, Koustuv Saha, Leah Galante Roper, Lama Nachman, Hong Lu, Lucia Mirabella, Sanjeev Srivastava, Munmun De Choudhury, et al. 2020. A real-time eating detection system for capturing eating moments and triggering ecological momentary assessments to obtain further context: System development and validation study. *JMIR mHealth and uHealth* 8, 12 (2020), e20625.

[13] Vasileios Papapanagiotou, Christos Diou, Janet van den Boer, Monica Mars, and Anastasios Delopoulos. 2021. Recognition of food-texture attributes using an in-ear microphone. In *International Conference on Pattern Recognition*. Springer, 558–570.

[14] Jianing Qiu, Frank P-W Lo, and Benny Lo. 2019. Assessing individual dietary intake in food sharing scenarios with a 360 camera and deep learning. In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 1–4.

[15] Jenna L Scisco, Eric R Muth, and Adam W Hoover. 2014. Examining the utility of a bite-count–based measure of eating activity in free-living human beings. *Journal of the Academy of Nutrition and Dietetics* 114, 3 (2014), 464–469.

[16] 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 Transactions on Computing for Healthcare* 1, 4 (2020), 1–19.

[17] Jee-Seon Shim, Kyungwon Oh, and Hyeon Chang Kim. 2014. Dietary assessment methods in epidemiologic studies. *Epidemiology and health* 36 (2014).

[18] Simon Stankoski, Marko Jordan, Hristijan Gjoreski, and Mitja Luštrek. 2021. Smartwatch-based eating detection: Data selection for machine learning from imbalanced data with imperfect labels. *Sensors* 21, 5 (2021), 1902.

[19] Mingui Sun, Lora E Burke, Zhi-Hong Mao, Yiran Chen, Hsin-Chen Chen, Yicheng Bai, Yuecheng Li, Chengliu Li, and Wenyan Jia. 2014. eButton: a wearable computer for health monitoring and personal assistance. In *Proceedings of the 51st annual design automation conference*. 1–6.

[20] Maher Tarek and Shady Zahran. 2021. Enhanced IMU sensors fusion based on adaptive least square windowing technique. In *2021 International Telecommunications Conference (ITC-Egypt)*. IEEE, 1–4.

[21] Chunzhuo Wang, T Sunil Kumar, Walter De Raedt, Guido Camps, Hans Hallez, and Bart Vanrumste. 2024. Eating speed measurement using wrist-worn imu sensors towards free-living environments. *IEEE journal of biomedical and health informatics* (2024).

[22] Boyang Wei, Shibo Zhang, Xingjian Diao, Qiuyang Xu, Yang Gao, and Nabil Alshurafa. 2023. An end-to-end energy-efficient approach for intake detection with low inference time using wrist-worn sensor. *IEEE Journal of Biomedical and Health Informatics* 27, 8 (2023), 3878–3888.

[23] World Health Organization. 2020. Healthy diet. https://www.who.int/news-room/fact-sheets/detail/healthy-diet Accessed: 2023-09-30.