A Multimodal AI-Enabled Framework for Characterizing Overeating Behaviors and Consumption Patterns

Farzad Shahabi[1,2], Jessica Li[1], Chris Romano[2], Rowan McCloskey[1], Glenn Fernandes[1,2], Mahdi Pedram[3], Jacob Schauer[2], Tammy Stump[4], Nabil Alshurafa[1,2]
[1]Department of Computer Science and [2]Department of Preventive Medicine, Northwestern University, Evanston, IL, USA
[3]Department of Computer Science and Engineering, University of North Texas, Denton, TX, USA
[4]Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA

***Abstract*—Overeating is a key contributor to obesity, yet identifying and characterizing its underlying causes remains challenging. While prior research has leveraged Ecological Momentary Assessment (EMA) to capture psychological and contextual factors in real-time, few studies have integrated EMA with passive sensing to uncover fine-grained, individualized consumption behaviors. In this work, we present a multimodal framework combining psychological and contextual data from a custom-built EMA app with validated camera-derived meal microstructure features from a neck-worn activity-oriented wearable camera. Across 41 participants, the camera captured 6,343 hours of footage over 312 days, yielding annotated bites, chews, meal start/end times, and dietitian-confirmed caloric intake. Using supervised contrastive learning, we generated meal-level representations, projected them using UMAP, and applied k-means clustering to identify behavioral phenotypes. We then conducted a z-score analysis to highlight features most distinctive to each cluster. Among the eight discovered groups, three consistently showed high purity for overeating meals (average purity = 0.99), revealing nuanced, data-driven overeating phenotypes that may inform targeted intervention strategies.**

***Index Terms*—Overeating phenotypes, multimodal sensing, supervised contrastive learning, wearable camera**

I. Introduction

As of 2023, 40.3% of US adults have obesity [1], a disease caused by energy intake in excess of expenditure [2]. Individual factors shape the formation of overeating habits that sustain energy imbalance over time [3]. Consequently, understanding the patterns underlying individual dietary habits is crucial for designing effective long-term weight loss interventions [4]. Recent advances in sensor analytics enable automated capture of such patterns. This approach supports identifying both known (e.g., emotional eating, impulsive eating in response to cues, hedonic eating) and novel eating phenotypes, enabling personalized, meal-level interventions.

Significant efforts have been made to understand eating behaviors through objective and subjective monitoring methods such as passive sensing and Ecological Momentary Assessment (EMA) [5]. Previous passive sensing approaches include wrist-worn accelerometers to detect eating, inertial sensors to monitor chewing, and multi-sensor systems combining proximity, IMU, and ambient light sensors to track eating gestures via jaw and hand movements [5]. Compared to traditional self-reports, EMA provides more accurate and nuanced data by capturing eating contexts in real time, which aids in identifying co-occurring psychological or situational exposures over time [6]. Combining image-based EMA with 24-hour dietary recalls can further enhance accuracy by reducing errors in dietary reporting [7]. The wealth of data these tools provide invites us to explore new analytic approaches to overeating.

Prior studies examine overeating through individual factors such as stress, nighttime eating, and loss of control (LOC). Emotional states (e.g., anxiety, boredom, depression) are closely linked to emotional eating [8], while behaviors can increase intake (e.g., rapid eating) by overriding satiety cues [9]. Environmental influences, (e.g., access to high-calorie foods, social settings) also contribute to overeating [2]. Most existing methods examining these determinants have primarily relied on linear statistical approaches, which utilize a limited set of predictors to tackle the complex and multifaceted issue of overeating and thus fall short in capturing the interplay of physiological, psychological, behavioral, and environmental factors at work in overeating.

To address limitations in prior work, we introduce a multimodal framework that integrates passive sensing and EMA to explore overeating. Leveraging a dataset of 700 dietitian-verified meals from 41 adults with obesity, we extracted psychological, contextual, and camera-derived features across 14 days. Our contributions include: (1) collecting a high-resolution, multimodal dataset linking passive sensing and EMA in real-world eating contexts; (2) developing a supervised contrastive learning model to derive meaningful meal-level embeddings; and (3) identifying interpretable overeating phenotypes through clustering, revealing novel sensor-EMA behavioral signatures that may guide targeted intervention strategies for nutrition and obesity care.

II. Methods

A. Study Design and Data Collection

The 14-day observational study involved 65 adults with obesity (BMI $\geq 30\ \mathrm{kg/m^2}$) in the Chicago area, designed to examine real-world eating behaviors using a combination of

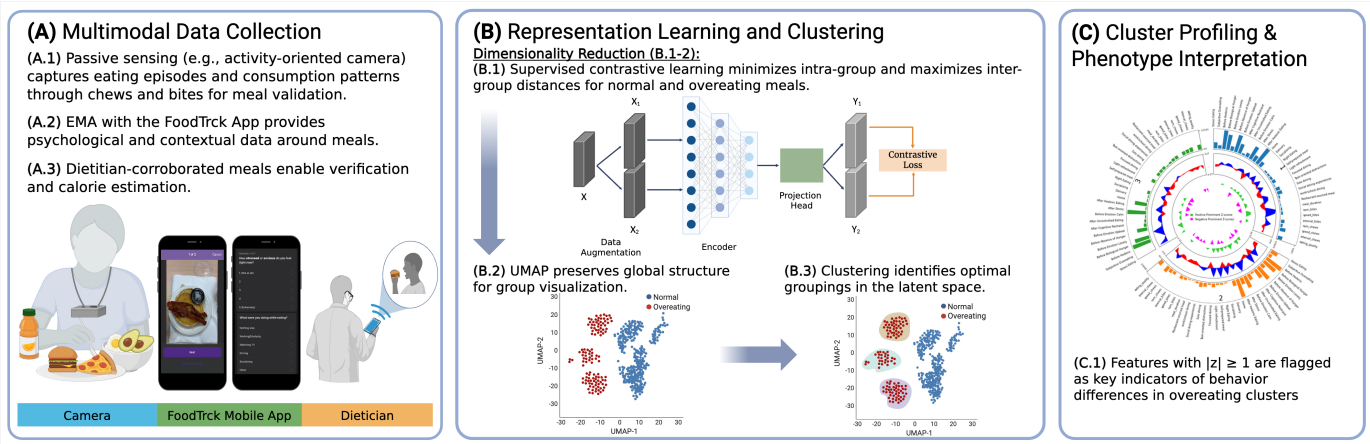

Fig. 1. Multimodal pipeline for overeating phenotype discovery: (A) Passive sensing, EMA, and dietitian-verified meals enable rich data collection. (B) Supervised contrastive learning and UMAP uncover latent clusters. (C) Key features ($|z\text{-score}| \geq 1$) profile behavioral differences.

passive sensing and EMA. Participants were equipped with a wearable sensing suite that included a neck-mounted, activity-oriented wearable camera capable of capturing thermal and RGB video data to monitor eating activity in naturalistic settings. A custom smartphone app prompted participants to complete event-based EMA surveys before and after meals. Dietary intake was further assessed through 24-hour recalls conducted by trained dietitians using the multiple-pass method, supported by participant-captured food images. Of the 65 initial participants, 41 participants were included in the final analysis after excluding individuals who dropped out (n=7), lacked dietitian-administered recalls (n=5), had fewer than 10 valid meals (n=5), or had insufficient wearable camera data (n=7). The resulting dataset comprised 700 meals with complete passive sensing and EMA data. Figure 1 displays an overview of the analysis pipeline. More details about the study design and data collection can be found in the SenseWhy design paper [10].

B. Feature Extraction

EMA-derived features captured psychological and contextual factors surrounding each meal. Prior to and following each eating episode, participants completed surveys via a mobile app assessing hunger, affect, stress, and emotional states using Likert-scale items. Contextual features included meal source (e.g., self-prepared, restaurant), setting (e.g., solo or social dining), and concurrent activities (e.g., task-oriented distraction or focused eating). Evening meals were defined as those consumed between 5 PM and 6 AM. Following a standardized protocol validated by trained raters, we conducted fine-grained annotations of eating activity using wearable camera footage. For each eating episode, we manually labeled the start and end times. Within each episode, we annotated individual bites, defined as large jaw open/close movements when food enters the mouth, and chews, defined as smaller, consecutive jaw movements during mastication. From these annotations, we derived passive sensing features including total number of chews/bites, chew-to-bite ratio, average time interval between chews and bites, meal duration from start/end of eating episodes, eating pace, and percent of meal time spent chewing. The outcome of interest, overeating, was defined as a binary variable indicating whether a meal's energy intake exceeded 1 z-score above the individual's 14-day average intake, based on validated data from 24-hour dietary recalls [10].

C. Eating Pace

In obesity research, eating speed influences satiety, intake, and weight outcomes, making it a key behavioral marker. Traditional assessments of eating speed, however, often rely on self-report or coarse measurements of meal duration, which may be prone to recall bias or distortion by long inactive periods during eating. To address these limitations, we propose a refined, behaviorally grounded method of defining eating based on chews, allowing for precise quantification of eating dynamics. We applied DBSCAN, a density-based clustering algorithm, to identify natural groupings of chews and remove temporally isolated outliers. DBSCAN parameters (ϵ and minimum samples) were optimized for each participant using elbow-method clustering to accommodate individual differences in chewing frequency. Chews occurring within ϵ seconds of one another were grouped into chewing sequences, while those separated by longer intervals were considered isolated and excluded. This segmentation filters out spurious chewing activity—such as a single chew minutes apart from others—that could otherwise inflate eating time or distort derived metrics. An eating episode was defined as chewing sequences containing at least 15 consecutive chews, in line with prior work, which established this threshold as a minimal criterion for sustained eating behavior [11]. Building on this segmentation, we introduced a novel metric for eating pace, defined as the number of chews per meal divided by the total chewing duration—calculated from the time between the first and last chew in the identified sequence. Unlike conventional definitions based on bite timing or meal start/end timestamps, our chew-based approach is robust to pauses in eating and better reflects actual masticatory engagement.

TABLE I
PHENOTYPE PROFILES OF OVEREATING CLUSTERS

Phenotype Name	Psychological Features	Contextual Features	Camera Features
"Fast-Fork Social Dining"	Loss of control$^{(p)}$ Pre-meal biological hunger$^{(p)}$ Desire for food$^{(p)}$	Social dining experience$^{(p)}$ Restaurant-sourced meal (dine-in)$^{(p)}$ In-restaurant dining$^{(p)}$ Socializing$^{(p)}$ In-home dining$^{(n)}$	Eating pace$^{(p)}$
"Serene Screen Suppers"	Pre-meal biological hunger $^{(p)}$ Pre-meal calm$^{(p)}$ Hedonic eating$^{(p)}$ Desire for food$^{(p)}$ Perceived overeating$^{(n)}$ Pre-meal upbeat$^{(n)}$ Post-meal stress$^{(n)}$	Evening eating$^{(p)}$ Task-oriented distraction (watching TV)$^{(p)}$ Self-prepared meal$^{(p)}$ Solo dining$^{(p)}$	Chews number$^{(p)}$ Bites number$^{(p)}$ Meal duration$^{(p)}$ Chew-bite ratio$^{(n)}$ Eating pace$^{(n)}$ Chew time-meal duration ratio$^{(n)}$
"Stressed Solo Bites"	Pre-meal stress$^{(p)}$ Post-meal stress$^{(p)}$ Pre-meal lonely$^{(p)}$ Perceived overeating$^{(p)}$ Cognitive restraint$^{(p)}$ Hedonic eating$^{(n)}$ Pre-meal upbeat$^{(n)}$ Pre-meal calm$^{(n)}$	Restaurant-sourced meal (takeout)$^{(p)}$ In-home dining$^{(p)}$	–

The table shows the principal features and characteristics associated with each derived overeating phenotype cluster. *Note.* Superscripts $^{(p)}$ and $^{(n)}$ indicate positive and negative z-scores, respectively.

To our knowledge, this is the first study to examine the relationship between overeating and detailed passive sensing features—such as eating pace, chew-to-bite ratio, percent of time spent chewing, and average chew/bite intervals—using validated, fine-grained annotations from wearable sensor data.

D. Supervised Contrastive Learning Framework

We developed a supervised contrastive learning framework, SupConNet, inspired by the architecture and training strategy proposed in [12]. SupConNet consists of an MLP-based encoder followed by a projection head, trained using a supervised contrastive loss to learn discriminative embeddings of meal-level behavioral features. During training, two augmented views of each sample are generated through stochastic transformations, and the model is optimized to pull together embeddings from the same class (e.g., overeating vs. non-overeating) while pushing apart those from different classes. Formally, let $f_\theta : \mathbb{R}^d \to \mathbb{R}^p$ denote the full network parameterized by θ, where $\mathbf{x} \in \mathbb{R}^d$ is a high-dimensional input feature vector and $\mathbf{z} = f_\theta(\mathbf{x}) \in \mathbb{R}^p$ is the learned embedding in a latent space. After training, we discard the projection head and extract embeddings from the encoder for downstream tasks. To visualize the learned representations, we apply Uniform Manifold Approximation and Projection (UMAP) to project $\mathbf{z} \in \mathbb{R}^p$ into a two-dimensional space $\mathbb{R}^2$, facilitating qualitative exploration of latent structure and clustering in eating behaviors.

E. Latent Space Clustering and Z-Score Analysis

We applied k-means clustering to the two-dimensional UMAP projections of the SupConNet embeddings to identify distinct subgroups of participants based on their learned behavioral representations. To determine the optimal number of clusters k, we computed the average silhouette score across a range of candidate values $k \in [2, 10]$, selecting values near local maxima for further evaluation.

We evaluated the resulting clustering using standard metrics including homogeneity, entropy, Davies-Bouldin Index (DBI), purity, and silhouette score. Homogeneity measures whether clusters contain only members of a single class (range: 0–1, higher is better); DBI evaluates cluster compactness and separation (range: 0 to ∞, lower is better); entropy quantifies class distribution uncertainty within clusters (range: 0–1, lower is better); and the silhouette score captures cohesion vs. separation (range: -1 to 1, higher is better). Purity was computed as:

$$\text{Purity} = \frac{1}{N} \sum_k \max_j |C_k \cap L_j| \,,$$

where C_k is the set of samples in cluster k, L_j is the set of samples in class j, and N is the total number of samples.

To assess the distinctiveness of clusters, we conducted a z-score analysis across all features included in the study. For each cluster c and feature j, we calculated:

$$z_{c,j} = \frac{\mu_{c,j} - \mu_j}{\sigma_j},$$

where $\mu_{c,j}$ is the mean of feature j within cluster c, μ_j is the overall mean across all clusters, and σ_j is the corresponding standard deviation. This analysis enabled standardized comparisons across clusters to identify behavioral patterns and characterize each subgroup.

III. RESULTS AND DISCUSSION

A. SupConNet Specifications

We trained SupConNet using a fully connected encoder with three hidden layers (128–128–64 units, ReLU activations) followed by a two-layer projection head with a final projection dimension of 32. Input features were augmented by adding Gaussian noise and applying random masking before contrastive training. The model was optimized using a supervised contrastive loss with a temperature parameter of 0.07. Training was conducted for 500 epochs using the Adam optimizer with a learning rate of 10^{-3} and a batch size of 4. After training, embeddings from the encoder were extracted and used for UMAP projection and then downstream clustering and phenotype profiling.

B. Overeating Subgroup Discovery via Clustering in Latent Space

After performing k-means clustering with $k = 8$, selected based on the maximum silhouette score, on the two-dimensional UMAP projections of SupConNet embeddings, we obtained clusters that exhibited strong separation and internal consistency. The resulting clustering achieved an average silhouette score of 0.55, homogeneity = 0.98, DBI = 0.51, average purity = 0.99, and average entropy = 0.05 across all clusters. Among these, three clusters were highly pure with respect to the overeating class and collectively accounted for all of the overeating meals. These clusters were selected for further profiling as candidate overeating phenotypes.

C. Phenotype Profiling

Table I summarizes the three overeating phenotypes identified via clustering, highlighting features with $|z\text{-score}| \geq 1$ as the most distinctive. For each cluster, we report cluster size (n), purity score (p), mean energy intake with standard deviation (kCal), entropy (e), and silhouette score (s). Additionally, we assign a concise name to each phenotype, highlight novel behavioral patterns, and contextualize them with existing literature.

"Fast-Fork Social Dining" ($n = 66$, $p = 100\%$, 1190.3 ± 355.2 kCal, $e = 0.00$, $s = 0.83$): reflects elevated loss of control and biological hunger, with a tendency to eat socially or at restaurants and at a faster pace. Our findings suggest that loss of control, as experienced in the context of social dining, may be coupled with an elevated eating pace—providing a behavioral lens through which sensor-derived features reflect internal states. This extends prior associations between eating speed and reduced satiety awareness [13].

"Serene Screen Suppers" ($n = 27$, $p = 92.6\%$, 1206.2 ± 536.2 kCal, $e = 0.381$, $s = 0.93$): are marked by biological hunger, low stress, and calm affect. Individuals in this group often eat self-prepared, evening meals while watching TV alone. Despite frequent chewing and biting, their low chew-to-bite ratio and slow eating pace suggest distracted, extended meals. Although prior work has linked screen-related distractions to overeating [14], we found that such distracted, mindless eating may also be reflected in altered chewing behavior (e.g., a lower chew-to-bite ratio and slower eating pace), highlighting new sensor-derived indicators of disengaged eating.

"Stressed Solo Bites" ($n = 18$, $p = 100\%$, 1143.4 ± 259.5 kCal, $e = 0.00$, $s = 0.94$): characterized by stress-related solitary takeout eating, elevated pre- and post-meal stress, loneliness, and perceived overeating, alongside low positive affect and reduced hedonic motivation. This pattern reinforces and extends prior research linking emotional distress and social isolation to maladaptive eating behaviors and increased risk of overeating [8].

IV. Conclusion and Future Work

These findings highlight distinct phenotypes of overeating meals, integrating psychological, contextual, and sensor-derived features. The outlined phenotypes provide a better understanding of prominent overeating behaviors at the meal level, which can then be utilized for tailored interventions deployed in real-time when passive sensing detects eating patterns consistent with specific phenotypes. Future work will focus on validating these phenotypes in larger, diverse populations and exploring their utility for targeted intervention strategies.

Acknowledgment

This work was supported by the National Institutes of Health (NIH) under award numbers K25DK113242, R03DK127128, R21EB030305, and R01DK129843. The figure was created using BioRender software (https://biorender.com/1j1qu42).

References

[1] S. D. Emmerich, C. D. Fryar, B. Stierman, and C. L. Ogden, "Obesity and severe obesity prevalence in adults: United states, august 2021–august 2023," 2024.

[2] I. Romieu, L. Dossus, S. Barquera, H. M. Blottière, P. W. Franks, M. Gunter, N. Hwalla, S. D. Hursting, M. Leitzmann, B. Margetts, *et al.*, "Energy balance and obesity: what are the main drivers?," *Cancer causes & control*, vol. 28, pp. 247–258, 2017.

[3] G. A. Bray, W. E. Heisel, A. Afshin, M. D. Jensen, W. H. Dietz, M. Long, R. F. Kushner, S. R. Daniels, T. A. Wadden, A. G. Tsai, *et al.*, "The science of obesity management: an endocrine society scientific statement," *Endocrine reviews*, vol. 39, no. 2, pp. 79–132, 2018.

[4] R. R. Wing and S. Phelan, "Long-term weight loss maintenance," *The American journal of clinical nutrition*, vol. 82, no. 1, pp. 222S–225S, 2005.

[5] B. M. Bell, R. Alam, N. Alshurafa, E. Thomaz, A. S. Mondol, K. de la Haye, J. A. Stankovic, J. Lach, and D. Spruijt-Metz, "Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review," *NPJ digital medicine*, vol. 3, no. 1, p. 38, 2020.

[6] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological momentary assessment," *Annu. Rev. Clin. Psychol.*, vol. 4, no. 1, pp. 1–32, 2008.

[7] C. Boushey, M. Spoden, F. Zhu, E. Delp, and D. Kerr, "New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods," *Proceedings of the Nutrition Society*, vol. 76, no. 3, pp. 283–294, 2017.

[8] O.-R. Ha and S.-L. Lim, "The role of emotion in eating behavior and decisions," *Frontiers in Psychology*, vol. 14, p. 1265074, 2023.

[9] G. Privitera, K. Cooper, and A. Cosco, "The influence of eating rate on satiety and intake among participants exhibiting high dietary restraint," *Food & nutrition research*, vol. 56, no. 1, p. 10202, 2012.

[10] N. I. Alshurafa, T. K. Stump, C. S. Romano, A. F. Pfammatter, A. W. Lin, J. Hester, D. Hedeker, E. Forman, and B. Spring, "Rationale and design of the sensewhy project: A passive sensing and ecological momentary assessment study on characteristics of overeating episodes," *Digital health*, vol. 9, p. 20552076231158314, 2023.

[11] I. Ioakimidis, M. Zandian, L. Eriksson-Marklund, C. Bergh, A. Grigoriadis, and P. Södersten, "Description of chewing and food intake over the course of a meal," *Physiology & behavior*, vol. 104, no. 5, pp. 761–769, 2011.

[12] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," *Advances in neural information processing systems*, vol. 33, pp. 18661–18673, 2020.

[13] S.-q. Yuan, Y.-m. Liu, W. Liang, F.-f. Li, Y. Zeng, Y.-y. Liu, S.-z. Huang, Q.-y. He, B. Quach, J. Jiao, *et al.*, "Association between eating speed and metabolic syndrome: a systematic review and meta-analysis," *Frontiers in nutrition*, vol. 8, p. 700936, 2021.

[14] D. Garg, E. Smith, and T. Attuquayefio, "Watching television while eating increases food intake: A systematic review and meta-analysis of experimental studies," *Nutrients*, vol. 17, no. 1, p. 166, 2025.