NutriEar: Robust Nutrition-Aware Food Classification from In-Ear Acoustic Signals

ZOEY XIAOCHEN TAN, University of Cambridge, United Kingdom
YANG LIU, Florida State University, United States
KAYLA-JADE BUTKOW, University of Cambridge, United Kingdom
CECILIA MASCOLO, University of Cambridge, United Kingdom

Convenient tracking of food intake is essential for linking diet to health, enabling personalised nutrition guidance, early metabolic risk detection, and effective prevention of chronic disease. Recent wearable sensing advances have begun to automate eating monitoring. However, these systems largely focus on detecting *when* users eat and only weakly address *what* they eat. In particular, state-of-the-art solutions typically cover a narrow set of foods or textures, rely on strong assumptions about individual eating behaviour, and often produce labels that are poorly aligned with the nutritional information most useful to end users. As a result, their outputs remain difficult to translate into meaningful everyday dietary guidance.

In this paper, we present *NutriEar*, an audio sensing system that performs nutrition-aware classification of food intake from in-ear chewing sounds. Rather than recognising arbitrary food identities, *NutriEar* maps in-ear acoustics to an eight-class nutrition-texture taxonomy grounded in food science, jointly capturing dominant macronutrient role and mechanical texture to provide an interpretable target space that links chewing acoustics to nutritionally meaningful categories.

To accommodate natural eating behaviour, *NutriEar* detects bite events directly from the audio stream using a lightweight segmentation method that does not assume fixed bite–chew–swallow cycles. For each detected bite, *NutriEar* extracts short post-bite chewing segments, focusing on the early chewing phase where texture-dependent acoustic cues are most prominent. From each segment, *NutriEar* derives a hybrid representation that combines engineered features and learned embeddings. The engineered component captures interpretable cues related to food texture and chewing dynamics, including energy variation, spectral structure, envelope patterns, and pseudo-motion characteristics inferred from the audio. In parallel, the system uses a pre-trained CLAP [1] audio encoder with supervised contrastive fine-tuning to obtain embeddings that better separate nutrition-texture classes across users and recording conditions. These complementary features are fused and passed to a lightweight classifier, and segment-level predictions are further aggregated into bite-level outputs through a temporally weighted voting scheme that emphasises the most informative windows immediately after bite onset.

An evaluation on a dataset we collected with over 30 food types from 15 users under varied eating conditions demonstrates that *NutriEar* achieves 80.18% average leave-one-subject-out (LOSO) accuracy and outperforms state-of-the-art baselines. These findings highlight the potential of combining earable acoustic sensing, structured nutrition-aware labels, and hybrid representation learning to move dietary monitoring beyond coarse eating-event detection toward more meaningful nutritional inference.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing design and evaluation methods**.

Additional Key Words and Phrases: Earable Sensing, Dietary Monitoring, In-Ear Acoustic Sensing

References

[1] Yusong Wu*, Ke Chen*, Tianyu Zhang*, Yuchen Hui*, Taylor Berg-Kirkpatrick, and Shlomo Dubnov. 2023. Large-scale Contrastive Language-Audio Pretraining with Feature Fusion and Keyword-to-Caption Augmentation. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*.

Authors' Contact Information: Zoey Xiaochen Tan, University of Cambridge, Cambridge, United Kingdom, xt229@cam.ac.uk; Yang Liu, Florida State University, Tallahassee, United States, yl25r@fsu.edu; Kayla-Jade Butkow, University of Cambridge, Cambridge, United Kingdom, kjb85@cam.ac.uk; Cecilia Mascolo, University of Cambridge, Cambridge, United Kingdom, cm542@cam.ac.uk.