Foodback

Sensing How Consumers Experience Food

Tommaso Califano, Giovanni Ligato, Nicola Ramacciotti, Gabriele Suma
t.califano@studenti.unipi.it, g.ligato@studenti.unipi.it, n.ramacciotti2@studenti.unipi.it, g.suma@studenti.unipi.it

ABSTRACT

Online food reviews play a crucial role in shaping consumer choices and improving food services, yet a vast majority of dining experiences remain undocumented due to users' lack of time or motivation to write reviews. To address this gap, we propose a novel prototype smartphone application that automatically generates numerical food ratings by analyzing physiological and neural signals acquired via wearable devices—specifically electroencephalography headsets and smartwatches. Our system captures synchronized brain activity, heart rate, and electrodermal responses during brief tasting sessions, processing these multimodal data through a compact deep learning architecture. We collected data from 10 volunteers across 45 sessions, training an end-to-end model to predict subjective hedonic ratings on a 5-point scale. Experimental results demonstrate the feasibility of decoding food enjoyment from wearable biosignals, highlighting the critical importance of evaluation methodology: a session-level data split avoiding information leakage yields a realistic test accuracy of ~33%, whereas a random epoch-level split inflates performance to ~73% due to shared session-specific noise. This work lays foundational steps toward scalable, objective, and unobtrusive food experience assessment via wearable neurotechnology.

1 Introduction

Millions of food reviews are published online every day, yet most dining experiences go undocumented. Despite the central role that user's feedback plays in guiding consumer choices and improving food services, many people are unwilling to write reviews due to lack of time, motivation, or interest.

To address this issue, the long-term goal of our work is to develop a smartphone application that automatically generates food reviews by analyzing physiological and emotional data collected via electroencephalography (EEG) headsets and smartwatches. The envisioned system would monitor the user's neural and biometric responses—such as brain activity, heart rate, and stress levels—during and after a meal. These signals would then be processed using machine learning algorithms to produce a numerical score or descriptive review, without requiring any active user input.

However, due to time constraints, only a prototype of the system has been developed. The current version focuses on evaluating a single food item using a numerical rating from 1 to 5, based on biometric data captured over a brief 2-second window, as shown in Figure 5. While limited in scope, this prototype serves as a proof of concept for a broader vision of automated, data-driven food evaluation, and provides initial insights into the feasibility of such an approach.

2 Related Works

Recent research in emotion and sensory perception recognition has laid a promising foundation for systems that aim to capture human affective and sensory responses through wearable and neurophysiological technologies. The current work, which focuses on generating automated reviews of food experiences using EEG and smartwatch data, builds upon this literature while addressing key limitations in generalization, ecological validity, and integration across modalities.

The study *EmotionSense* by Wang et al. [1] presents a robust emotion recognition system using wearable smartwatches, effectively tackling the confounding impact of physical activity on physiological signals. A notable strength of this work is its context-aware emotion recognition framework, which uses a two-stage pipeline: first, identifying physical activity through accelerometer data with high accuracy (98.27%), and then applying *adaptive* emotion recognition using Support Vector Machines. The dual modelling approach—combining discrete emotional states with dimensional representations (valence-arousal)—adds flexibility to the system, enabling real-time, minimally invasive emotion tracking.

However, while *EmotionSense* achieves a commendable emotion recognition accuracy (74.3%), its focus remains limited to generic emotional states rather than context-specific experiences like taste or food enjoyment. Moreover, the experimental design introduces potential biases and temporal inconsistencies. Specifically, the emotion-stimulus clips used in the study were self-selected by participants, which may have introduced subjective bias and reduced the generalizability of the system's evaluation. Additionally, the protocol involved participants first performing a physical action and then being exposed to emotional stimuli via video, with physiological data collected only during the post-action phase, when no movement occurred. This design decouples emotion elicitation from concurrent physical activity, which may not accurately reflect the dynamic and overlapping nature of emotion and behaviour in real-life scenarios. Finally, although the system demonstrates progress toward real-world applicability, its reliance solely on smartwatch-derived physiological signals limits the granularity and depth of affective information compared to neural data such as EEG. Thus, while *EmotionSense* contributes significantly to activity-aware emotion classification, it lacks the specificity and multimodal integration required to capture the rich, subjective nuances of food-related experiences.

In contrast, the work by Zhang et al., titled *Decoding Human Taste Perception by Reconstructing and Mining Temporal-Spatial Feature of Taste-Related EEGs* [2], offers a novel and highly granular approach to decoding subjective taste experiences using EEG data. The study's deep learning architecture, enhanced by a Multi-View Channel Attention (MVCA) module and a Temporal-Spatial Reconstruction Data Augmentation (TSRDA) technique, achieves near-perfect classification performance (accuracy: 99.56%). This represents a substantial advancement over prior taste recognition methods, both human and machine based.

The strength of this work lies in its deep modelling of neurophysiological responses to basic taste stimuli and its methodological innovations in data augmentation, which address the common challenge of *limited* EEG datasets. Moreover, the controlled experimental protocol and *precise* delivery of taste stimuli ensure high-quality data collection.

Nevertheless, the system's experimental design is highly constrained, relying on artificial taste delivery under laboratory conditions and focusing only on four primary taste categories. This limits its direct applicability in naturalistic food consumption settings. Additionally, while the architecture is sophisticated, it remains unvalidated in real-world or multisensory contexts, such as those involving texture, aroma, or emotional associations with food.

However, an important limitation that potentially affects the reliability of the reported results lies in the ambiguity regarding data partitioning. The paper refers to the use of training and test sets but lacks explicit mention of a distinct validation set. This raises concerns about data split ambiguity, as it remains unclear whether a proper three-way split (training, validation, test) was employed.

Moreover, hyperparameter tuning uncertainty persists due to the absence of detailed information on whether tuning was performed on a dedicated validation subset or inadvertently on the test data. This methodological omission can compromise the objectivity of performance metrics.

Without a dedicated validation set, it becomes difficult to ascertain whether the model's hyperparameters were optimally calibrated or whether the classification accuracy reflects genuine generalization capability rather than overfitting. This uncertainty may undermine the broader generalizability and replicability of the results in real-world scenarios.

Our proposed system addresses the complementary limitations of both these studies by integrating real-world wearable sensing (from smartwatch) with EEG-based neural decoding in an *everyday* setting—thus capturing not only the physiological and emotional aspects of food experiences, but also the underlying perceptual and cognitive components. In doing so, we aim to provide a holistic and scalable solution for automatically generating food experience reviews, extending the state of the art in both emotion recognition and sensory decoding research.

3 Application Components

The prototype application consists of three main components: an authentication module powered by a Flask [3] based server, a data collection subsystem, and a data processing engine embedded within a Kotlin [4] based smartphone app. These systems work together seamlessly to produce quantitative assessments of the user's sensory experience.

3.1 User Authentication

The authentication system is implemented using Flask and follows a layered Model–Controller–Service–Repository (MCSR) architecture.

At the foundation of the system are manually defined data models that reflect the structure of the underlying MySQL [5] database. The repository layer handles all direct interactions with the database, providing a consistent Python [6] interface for CRUD operations.

The service layer manages the core authentication processes, including input validation and secure password hashing, ensuring that sensitive information like credentials is never exposed in plaintext.

The controller layer exposes a RESTful API to the mobile application. It handles user registration, login, and session management through signed JSON Web Tokens [7], using both access and refresh tokens to support secure and efficient authentication flows.

The system distinguishes between three user roles, each with clearly defined access privileges:

- **Anonymous User**: This is the default state for unauthenticated users. Anonymous users are only permitted to perform registration and login actions. They have no access to application functionalities beyond the authentication interface.
- **Regular User**: Upon successful authentication, a user becomes a Regular one. This role grants access to the rating inference functionality. Regular users can initiate sessions during meals, from which overall ratings are inferred and associated with specific dining experiences.
- **Administrator**: Admin users possess elevated privileges, although their scope is intentionally constrained to data collection activities. Unlike regular users, administrators are automatically redirected to a dedicated interface for managing and acquiring new data samples, which are later used to train the machine learning model. Notably, administrators do not engage in inference operations within the mobile app.

3.2 Data Collection

Data collection is orchestrated by two independent sources: the EEG headset and the smartwatch, both interfacing with a central smartphone application.
Before initiating data acquisition, the application verifies the active connection of both the smartwatch and the EEG device, ensuring system *readiness*. Upon launching a tasting session via the app's

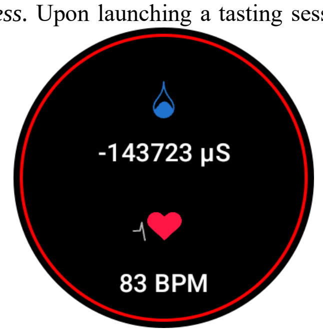

Figure 1: Interface of the smartwatch app during data acquisition.

dedicated button, a command is dispatched through Google's Wearable Data Layer API [8]—chosen for its low latency and reliability across the Android ecosystem. This triggers the companion app on the smartwatch, which runs a service continuously sampling physiological, specifically heart rate (HR, ~1 Hz) and electrodermal activity (EDA, ~10 Hz), to display them on the smartwatch screen as depicted in *Figure 1*.

Figure 2: Interface of the smartphone app during data acquisition.

Upon receiving the command, the companion app begins collecting these signals and sends them back to the smartphone. The signals are transmitted in batches, either in ten-second windows for offline model training or in two-second windows for real-time inference.

In parallel, the EEG headset streams high-resolution neural signals (~500 Hz) continuously and in real time directly to the mobile application, which manages buffering and storage according to the same dual-mode architecture. This process, which shows the EEG's data flow during collection, is illustrated in Figure 2.
As the two wearable devices function independently and do not communicate directly with each other, their respective data streams remain segregated until they converge at the application layer. At this point, synchronized processing and interpretation of multimodal data take place.

Figure 3: Interface for managing CSV files with the option to enable preview mode.

In data collection mode, once the smartwatch data has been fully received and temporally aligned with the EEG stream, the application serializes the data into three distinct CSV (comma-separated values) files—one per sensor. These files, also shown in Figure 3, preserve the *original* sampling rate of each sensor, ensuring the fidelity of the raw signals during post-processing.
Notably, these CSV files grow cumulatively with each session, meaning the system does not create separate files per experiment; instead, it maintains three evolving datasets, one for each signal type. This strategy promotes efficient long-term data aggregation while enabling precise multimodal alignment for subsequent offline analysis and machine learning.

3.3 Data Processing

To enable joint learning across heterogeneous data modalities—specifically, high-resolution EEG signals and low-frequency physiological signals from a smartwatch—it is essential to project all signals onto a *common temporal framework*. This ensures temporal alignment between modalities, which is critical for both the training and inference phases of the model.
To achieve this, separate preprocessing pipelines were developed for each data source. The EEG signals, acquired at 500 Hz, are subjected to a tailored sequence of filtering and downsampling

operations to match the temporal resolution of the wearable data. Conversely, smartwatch-derived signals—including electrodermal activity and heart rate—are linearly upsampled to the target sampling rate of 125 Hz. Once temporally aligned, the outputs of these two pipelines are fused at the normalization layer, which compensates for differences in scale and statistical distribution across modalities.

Except for manual upsampling applied to the smartwatch signals—necessary due to the low temporal resolution of heart rate measurements—preprocessing is fully integrated into the TensorFlow [9] computational graph, thereby minimizing the need for external preprocessing or on-device signal processing. This upsampling step is performed using Python during training and replicated in Kotlin for real-time inference. The result is a fully integrated model capable of ingesting raw 2-second data windows during both training and inference with minimal overhead. Specifically, during training, each 10-second trial is segmented into five non-overlapping 2-second *epochs* prior to preprocessing, allowing the model to learn from finer temporal granularity. In contrast, inference data arrives natively in 2-second windows and is processed directly, enabling real-time application without additional segmentation steps.

The EEG in-model preprocessing pipeline comprises the following stages:

- **High-pass filtering at 0.5 Hz** to remove slow drifts and baseline wander (detrending).
- **Band-pass filtering between 0.5–50 Hz** to isolate the canonical EEG frequency bands relevant for cognitive decoding.
- **Notch filtering at 49–51 Hz** to suppress power-line interference.
- **Anti-aliasing low-pass filtering at 62.5 Hz**, applied prior to decimation.
- **Decimation from 500 Hz to 125 Hz**, reducing temporal redundancy while preserving spectral fidelity.

All filtering stages are implemented using fixed-coefficient finite impulse response (FIR) filter, ensuring that the signal conditioning behaves identically across training and deployment. Importantly, the decimation step is performed at the final stage of the EEG pipeline, just before data enters the model, and remains within the TensorFlow graph.

The wearable signals—heart rate (originally sampled at 1 Hz) and electrodermal activity (approximately 10 Hz)—are linearly interpolated outside the model to reach the common sampling rate of 125 Hz. This preprocessing step enables direct temporal alignment with the EEG data once the signals enter the TensorFlow computation graph. Finally, the two modality-specific pipelines converge at a *shared* normalization layer, which harmonizes the preprocessed signals prior to model ingestion. Given the significant differences in dynamic ranges—EEG signals measured in microvolts can span millions in both positive and negative values, EDA signals in microsiemens can reach hundreds of thousands with predominantly negative values, and heart rate, measured in beats per minute, typically ranges only in the low hundreds with positive values—normalization is a critical step to ensure comparability across modalities. The layer performs per-channel standardization according to:

$$x \mapsto \frac{x - \mu_c}{\sigma_c}$$

where μ_c and σ_c represent the mean and standard deviation of each channel, computed over the training dataset. This normalization ensures balanced feature representation, preventing modalities with larger numerical magnitudes from disproportionately influencing the learning process.

Following preprocessing and normalization, each 2-second multimodal window, consisting of synchronized EEG and smartwatch signals, serve as input to the neural network. The model employed is EEGNet [10], a compact convolutional neural network originally designed for EEG signal classification. It leverages depthwise and separable convolutions to efficiently extract spatial–spectral patterns while maintaining a small footprint—approximately 1,000 trainable parameters. This makes it particularly well suited for edge scenarios involving limited data availability and constrained computational budgets.

After training, the —complete with embedded signal conditioning, resampling, and normalization operations—is exported to the TensorFlow Lite (TFLite) format for deployment on mobile and embedded platforms. This lightweight runtime enables on-device execution of the trained model, offering key advantages such as low-latency inference, reduced binary size, efficient memory usage, and seamless interoperability with Kotlin-based Android environments. An example of the live inference process running directly on the device is shown in Figure 5.

4 Experimental Results

This section presents the experimental protocol used for data collection and the methodology for evaluating the performance of the proposed multimodal classification model. It first details the procedure followed during each tasting session, including device setup, timing cues, and participant instructions. Subsequently, it outlines the training strategy, the evaluation metrics, and the resulting classification performance.

The experiment consisted of 10 volunteers aged between 16 and 56 years, with most participants in their mid-twenties. A total of 45 tasting sessions were conducted, each producing a labelled EEG segment synchronized with smartwatch sensor recordings and self-reported hedonic ratings. These multimodal samples formed the dataset used for training and validating the proposed classification model.
To acquire the physiological signals, the following devices were employed:

- **Mindrove Arc** EEG headset [11] for capturing neural activity across six channels at 500 Hz,
- **Google Pixel Watch 2** [12] for recording heart rate and electrodermal activity,
- **LG G8 ThinQ** smartphone [13] for orchestrating session timing, collecting responses, and coordinating device synchronization.

4.1 Experimental Setup

To ensure consistency and reduce procedural variability, each experimental session began with an optional *Preview Mode*, activated via a dedicated button (see Figure 3). During this phase, participants could rehearse the complete tasting experience as many times as desired. Although no data was recorded in this mode, all stimuli and interactions — including acoustic cues, real-time visual feedback, and post-experience rating prompts — were identical to those used in the actual trials. This *familiarization period* was designed to minimize anxiety and motion-related EEG artifacts, promoting higher-quality data acquisition during the official protocol.
The formal data collection procedure followed a structured timeline composed of the following steps:

1. **Cue 1 – Initiation Beep**: The participant is prompted to bring the food sample to their mouth.
2. **Stabilization Phase**: For the next five seconds, the participant holds the sample without chewing, remains completely still with relaxed facial muscles and a steady posture, breathes naturally, and keeps their eyes closed to reduce muscular and ocular interference in the EEG recording, as shown in Figure 4.
3. **Cue 2 – Recording Start Beep**: for the next ten-second window the participant must remain motionless and maintain the same relaxed state as in the stabilization phase to ensure clean data acquisition.
4. **Cue 3 – Recording End Beep**: Indicates the conclusion of the recording phase, signalling that is possible to move or swallow the sample.

Immediately after each trial, the participant is asked to provide a hedonic evaluation of the sample using a five-point scale (via the user interface shown in Figure 5), where a score of one reflects *not appreciated at all* and a score of five indicates *highly appreciated.* These subjective ratings serve as the reference ground truth for training and evaluation phase of the recognition model. To avoid sensory adaptation and carry-over effects, participants were instructed to rinse their mouth with water between tastings.

Figure 4: A volunteer keeping the required posture during data recording phase.

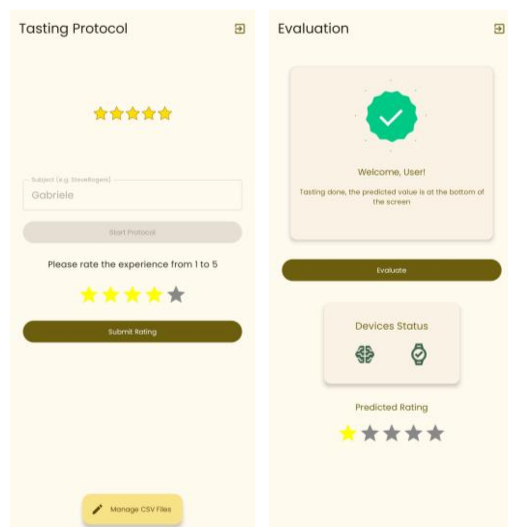

Figure 5: On the left there is the user interface for manual score input while on the right the interface displays the automatically inferred rating by the system.

4.2 Performance Evaluation

Two distinct evaluation strategies were employed to assess model performance: one using a session-level split, and another using a random epoch-level split. The model was trained over 1000 epochs with a batch size of 128, using synchronized inputs from EEG, HR, and EDA signals. Additionally, 10% of the training data was set aside as a validation subset to monitor performance during training.
In the first approach, the dataset was partitioned using a strict 80/20 train–test split applied at the experiment level, ensuring that all 2-second epochs from a given tasting session were assigned exclusively to either the training or the test set. This prevented overlap of sessions between partitions and avoided contamination

of the test set with session-specific signals. While subjects could appear in both partitions, this method preserved generalization by mitigating the pitfalls of overly conservative subject-wise splits, which often degrade EEG classification performance due to high inter-subject variability. Using this setup, the model achieved a final test accuracy of approximately **33.33%**.

The second evaluation randomly split individual 2-second epochs into training and test sets at an 80/20 ratio without respecting session boundaries. This approach yielded a substantially higher test accuracy of approximately **73.33%**. However, it suffers from information leakage since adjacent epochs within the same session share subject- and setup-specific noise. As a result, the model effectively "peeks" at the test data during training, artificially inflating performance. Despite the appealing accuracy numbers, this evaluation overestimates the model's true generalization ability and should be interpreted with caution.

These results highlight the critical importance of careful data partitioning in EEG-based multimodal classification. While the epoch-level split may suggest impressive performance, only the session-level split provides a reliable estimate of the model's true generalization capabilities, free from information leakage.

5 Conclusion

This work presents a proof-of-concept system for automatic, biometric-based food evaluation that integrates EEG, heart rate, and electrodermal activity signals collected via wearable devices. Although limited in scope—restricted to short, controlled tasting sessions and a simple five-point rating scale—the prototype successfully demonstrates the technical feasibility of real-time multimodal data acquisition, preprocessing, and on-device inference using a lightweight neural network architecture.

Importantly, the current system represents only a partial implementation of the envisioned full application. Its primary purpose is to validate the core sensing and inference pipeline, laying a solid foundation for the integration of more advanced features and richer data modalities in future development stages.

While the initial classification accuracy of approximately 33.33% on a rigorously partitioned session-level dataset may appear modest, it provides a realistic and unbiased benchmark of the system's true generalization ability. Higher accuracy scores obtained via less conservative epoch-level splits, although encouraging, reveal the risks of information leakage and highlight the critical importance of carefully designed evaluation protocols in this domain.

Looking ahead, future work will aim to expand the dataset with a larger and more diverse participant pool, enhance the temporal modelling of physiological responses, and explore richer output formats such as free-text review generation. Additionally, adaptive models personalized to individual users could improve predictive performance and user experience. The application could also be improved by enabling the saving of reviews, allowing all users to view them later.

REFERENCES

[1] Z. Y. B. Z. B. G. C. C. a. Z. Y. Zhu Wang, "EmotionSense: An Adaptive Emotion Recognition System Based on Wearable Smart Devices.," *Association for Computing Machinery,* p. 17, 2020.

[2] S. L. a. H. D. S. a. S. G. a. S. H. a. W. S. a. K. V. a. P. B. a. X. G. a. V. D. Herdt, "Heart rate, electrodermal responses and frontal alpha asymmetry to accepted and non-accepted solutions and drinks," *Food Quality and Preference,* vol. 82, 2020.

[3] Flask. [Online]. Available: https://flask.palletsprojects.com/en/stable/.

[4] Kotlin. [Online]. Available: https://kotlinlang.org/.

[5] Mysql. [Online]. Available: https://www.mysql.com/it/.

[6] Python, [Online]. Available: https://www.python.org/.

[7] J. W. Token. [Online]. Available: https://jwt.io/.

[8] Google,[Online].Available: https://developers.google.com/ android/reference/com/google/android/gms/wearable/Wearable.

[9] TensorFlow. [Online]. Available: https://www.tensorflow.org/?hl=it.

[10] A. J. S. N. R. W. S. M. G. C. P. H. B. J. L. Vernon J. Lawhern, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces.," *CoRR,* vol. abs/1611.08024, p. 30, 2016 .

[11] Mindrove, [Online]. Available: https://mindrove.com/

[12] Google, "Google Pixel Watch 2," [Online].Available: https://store.google.com/it/product/pixel_watch_2_specs?hl=it&pli=1.

[13] LG.[Online].Available: https://www.lg.com/us/support/products/documents/ LG_G8_ThinQ_Tech_Specs_ATT_PDF_4-10-19_LMG820UM1.pdf.