Vol. 16 (2026): Proceedings of the 2026 International Conference on AI and Technological Innovation Applications (AITIA 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/aitia.2026293299

AI-Driven Implicit Consumer Emotion Measurement in Food Sensory Evaluation: Research Progress and Future Directions

Tianyang Pei*

Guangdong Technion-Israel Institute of Technology, Shantou, China

**Corresponding author: Tianyang Pei.*

Abstract

Traditional food sensory evaluation relies heavily on explicit judgments such as descriptive analysis, check-all-that-apply questions, and hedonic ratings. These methods remain indispensable, yet they are vulnerable to memory limits, social desirability, fatigue, contextual bias, and the difficulty of verbalizing weak or rapidly changing affective states. This review examines how artificial intelligence is accelerating implicit consumer emotion measurement in food sensory evaluation. We synthesize recent progress in facial expression analysis, eye tracking, and electroencephalography, with attention to the affective signals each modality can capture, the signal-processing and machine-learning methods that make those signals usable, and the role of multimodal fusion in more robust preference prediction. The review argues that implicit measurement should be treated as a complementary layer rather than a substitute for explicit sensory science. Its future value depends on calibrated multimodal protocols, transparent models, ecological test environments, and strict ethical governance of biometric data.

Keywords

artificial intelligence, implicit emotion measurement, facial expression analysis, eye tracking, electroencephalography

1. Introduction

Food sensory evaluation has long depended on explicit reports: trained panels describe attributes, consumers mark liking scales, and rapid profiling methods such as CATA translate perception into selected terms. These tools remain the backbone of product development because they are interpretable, inexpensive compared with clinical instrumentation, and directly connected to consumer language. Their weakness is equally familiar. Food preference is not only a verbal judgment after tasting; it is also a rapid affective and bodily reaction shaped by attention, memory, expectation, culture, hunger, packaging, and the testing context. A participant can underreport dislike, overthink a rating, become fatigued, or fail to name a fleeting response that occurred during the first seconds of tasting. Recent reviews therefore describe a methodological shift from purely self-response-centered sensory tests toward technology-supported assessments that combine explicit ratings with biometric, behavioral, and digital measurements [1, 2, 3].

This paper reviews that shift through the lens of AI-driven implicit consumer emotion measurement. Here, implicit measurement refers to techniques that infer affective or cognitive states from spontaneous behavior or physiological signals rather than from direct self-report. In food sensory evaluation, the most visible examples are facial expression analysis, eye tracking, and electroencephalography. These modalities do not measure the same construct. Facial expression analysis is closest to momentary emotional valence and visible motor response. Eye tracking captures attention allocation, visual search, and gaze-choice relations before and during selection. EEG captures neural dynamics related to arousal, attention, flavor representation, and preference. AI enters the workflow by detecting faces and action units, extracting eye movement features, cleaning high-dimensional EEG signals, learning predictive representations, and fusing multiple streams into preference or liking models [4, 5].

The contribution of this review is a compact framework for food science researchers who need to understand what these methods can and cannot provide. Instead of treating implicit technologies as a single replacement for traditional sensory panels, we organize them as complementary layers in an intelligent sensory evaluation system. Section 2 reviews the core modalities. Section 3 discusses AI-based signal processing and prediction. Section 4 synthesizes multimodal fusion strategies and maps each modality to practical sensory questions. Section 5 discusses calibration, validity, standardization, and ethics.

The scope is deliberately practical. Prior reviews have already catalogued many explicit and implicit tools for food-evoked emotion, including self-report emotion lexicons, physiological measures, behavioral tasks, and facial or body responses [6, 7]. This paper narrows the discussion to the three modalities most directly connected to AI-enabled sensing pipelines and most likely to be adopted in food R&D laboratories: face video, gaze data, and EEG. The review also distinguishes sensory evaluation from adjacent marketing use cases. Package attention, purchase intention, and product liking are related, but they are not identical outcomes. A robust AI-assisted sensory protocol should therefore define whether the target is attribute intensity, hedonic liking, emotional valence, choice, or post-consumption memory before selecting sensors and models.

2. Literature Review

2.1 Facial Expression Analysis

Facial expression analysis is attractive in sensory evaluation because cameras are easy to deploy and facial reactions often occur within the time window in which tasting, smelling, or viewing a product unfolds. Modern automated facial expression analysis is built on the older Facial Action Coding System tradition, but uses computer vision to detect face landmarks, infer action units or basic emotion categories, and summarize temporal changes. Protocol work in sensory booths emphasizes lighting, camera placement, participant instructions, occlusion by cups or hands, and time alignment with tasting events [8]. Reviews of AFEA in beverages and consumer products show that the method has been used for product differentiation, packaging evaluation, sensory-specific satiety, cross-cultural comparison, and hedonic inference [9, 10].

For food evaluation, the strongest use case is not reading a stable inner emotion from the face. It is detecting short-lived affective cues that add information to explicit liking. Facial expressions may be more sensitive to dislike than to pleasure, because aversive tastes can produce faster and clearer motor reactions. Empirical work using FaceReader has shown that facial emotion scores can predict perceived hedonic ratings of drinks, but also reports limitations related to time-window selection, participant heterogeneity, and weak expressions for highly palatable stimuli [11]. The practical implication is that FEA is useful when protocols carefully separate product-evoked reactions from task-evoked smiles, embarrassment, chewing artifacts, or experimenter effects.

2.2 Eye Tracking

Eye tracking measures where, when, and for how long consumers visually attend. In food sensory and consumer science, it has been applied to packages, nutrition labels, menus, shelves, food images, and question formats. The core measures include fixation duration, fixation count, first fixation, scan path, saccades, and sometimes pupil diameter. A review by Motoki et al. emphasizes that visual attention is shaped by both bottom-up features, such as salience and size, and top-down factors, such as goals and task instructions [12]. This distinction matters for sensory evaluation because a package claim, a product image, or a CATA term may receive more gaze not because it is preferred, but because it is visually prominent, unfamiliar, or difficult to process.

Eye tracking is therefore best interpreted as an attention and information-processing layer. It can explain why a consumer noticed or ignored sensory cues before forming a rating. It can also reveal mismatch between what participants claim to have used and what they actually inspected. In food contexts, however, naturalistic stimuli are noisy: labels differ in color, size, familiarity, and semantic density. Good studies must control stimulus features or use enough randomized stimuli to reduce confounding. Eye tracking becomes especially powerful when paired with explicit choice tasks or biometric measures, because gaze alone rarely proves liking.

2.3 Electroencephalography

EEG provides high-temporal-resolution neural signals and is increasingly used to study food-related attention, arousal, flavor perception, and preference. Compared with cameras or eye trackers, EEG is more intrusive and technically demanding. It requires electrode preparation, artifact rejection, stimulus timing, and careful interpretation. Nevertheless, recent studies show why it is appealing. EEG with machine learning has been used to predict coffee sensory attributes from professional panel data [13], and deep learning has been applied to classify sweetness concentration from taste-evoked EEG signals [14]. In broader consumer neuroscience, systematic evidence suggests that preference prediction improves when EEG markers are modeled with machine learning and combined with other physiological or behavioral signals [15].

EEG is most valuable when the research question concerns neural processing rather than visible behavior. It can capture changes that are not easily verbalized, such as arousal, attentional engagement, or differences in flavor representation. Yet it is also the modality most vulnerable to over-interpretation. Frequency bands or scalp regions should not be treated as direct labels for pleasure without validation against sensory ratings, product variables, and behavioral outcomes. For food industry applications, EEG is promising but still better suited to targeted studies than routine high-throughput screening.

2.4 AI in Signal Processing and Sensory Prediction

AI contributes to implicit sensory evaluation in three connected stages. The first is preprocessing. Video-based FEA requires face detection, landmark tracking, frame quality control, expression coding, and temporal smoothing. Eye tracking requires fixation and saccade parsing, interpolation for lost gaze samples, mapping of gaze to areas of interest, and removal of blinks or calibration errors. EEG requires filtering, artifact removal, segmentation around stimulation events, channel selection, and feature normalization. These preprocessing steps are not neutral technical details; they determine what part of the consumer response is visible to downstream models.

The second stage is representation. Traditional feature engineering remains common: facial emotion intensity, valence indices, fixation counts, dwell time, first fixation, spectral EEG power, event-related potentials, entropy, and Hjorth parameters. Recent work increasingly combines these handcrafted features with learned representations from convolutional networks, recurrent models, attention mechanisms, and boosted-tree ensembles. The coffee EEG study by Bilucaglia et al. [13], for example, uses spectral and temporal EEG features with boosted-tree regression to predict sensory attributes, while Wang et al. propose a neural feature calculation module for sweetness concentration classification [14]. These examples show a move from descriptive biometrics toward predictive sensory modeling.

The third stage is interpretation. In food science, a black-box classifier is rarely enough. Product developers need to know whether a response reflects taste intensity, packaging attention, dislike, novelty, fatigue, or the laboratory setting. AI models should therefore report uncertainty, support model explanation, and be validated against explicit sensory scales, behavioral choices, and repeated sessions. The most useful models are not necessarily the most complex models; they are the models that remain stable across consumers, foods, contexts, and time. This is why transparent regression, tree ensembles, and interpretable fusion layers remain competitive with deep neural networks in small-sample sensory studies.

A fourth stage is deployment design. Models built for sensory science should be evaluated under the constraints of the intended use. A laboratory screening model can tolerate slower setup and richer instrumentation; an industry panel tool needs repeatable calibration, rapid cleaning, and clear exclusion rules;

a remote consumer test must work with webcams, variable lighting, and weaker control over tasting behavior. AI systems should therefore be judged not only by accuracy but also by failure mode. For FEA, failures include face occlusion, poor lighting, false smiles, and chewing. For eye tracking, failures include calibration drift and uncontrolled visual salience. For EEG, failures include swallowing artifacts, muscle noise, and overfitting in small participant pools. Reporting these failures is scientifically useful because it tells product developers when implicit measurement can support a decision and when a conventional sensory test is more reliable.

Another practical issue is temporal alignment. Food perception is dynamic: expectation begins before tasting, aroma and first taste arise quickly, texture and mouthfeel evolve during oral processing, and aftertaste may dominate final memory. AI pipelines should therefore preserve event timing rather than averaging all signals into a single score. A better protocol marks product viewing, first odor, first sip or bite, swallow, aftertaste, and final rating. This temporal structure allows researchers to ask whether a modality predicts immediate aversion, delayed liking, or final purchase intention.

3. Key Finding

Single modalities provide partial views of consumer experience. Facial expressions may miss positive liking, eye tracking may confuse attention with preference, and EEG may capture arousal without explaining its source. Multimodal fusion attempts to reduce these blind spots. Data-level fusion synchronizes raw or lightly processed streams; feature-level fusion concatenates modality-specific features before learning; decision-level fusion combines predictions from separate models. In food sensory evaluation, feature-level and decision-level fusion are often more practical because sampling rates, noise structures, and missing data differ across cameras, eye trackers, and EEG systems.

Table 1 summarizes the synthesis used in this review. The table is not an experimental result; it is a design guide for choosing modalities according to the sensory question. FEA is well matched to rapid hedonic valence and dislike detection. Eye tracking is strongest for visual attention, package processing, and choice architecture. EEG is strongest for time-sensitive neural responses and predictive modeling of sensory attributes. AI links these modalities by transforming raw signals into time-aligned, interpretable features and then learning relationships with explicit ratings, choices, and product variables.

Table 1: Synthesis of implicit modalities for AI-assisted food sensory evaluation.

Modality	Main signal	Best food-sensory use	Key limitation
Facial expression analysis	Visible affective motor response during viewing, smelling, or tasting	Rapid valence cues, dislike detection, beverage and packaging studies	Chewing, occlusion, social display, and weak positive expressions
Eye tracking	Fixations, saccades, dwell time, scan path, pupil response	Attention to labels, menus, packages, and choice cues	Attention is not equal to liking; visual confounds must be controlled
EEG	Time-varying neural activity after taste, odor, or visual stimulation	Arousal, attention, flavor representation, sensory attribute prediction	High artifact load, small samples, and risk of reverse inference
Multimodal fusion	Synchronized behavioral, visual, and neural features	Robust preference prediction and context-aware product testing	Requires shared timing, calibration, missing-data handling, and privacy controls

Recent multimodal consumer choice work illustrates the direction of travel. Usman et al. combine EEG and eye-tracking features to predict buy versus not-buy decisions, showing that neural and attentional streams can be integrated in a predictive model [15]. Food sensory science can adapt this logic, but should avoid importing neuromarketing claims without validation in tasting contexts. A useful future framework would collect explicit liking, FEA, gaze, and EEG in the same session; segment signals around product viewing, first smell, first sip or bite, aftertaste, and final rating; and evaluate whether each modality improves prediction beyond simpler baselines. Such a framework would also help distinguish three questions that are often blended: what the consumer attended to, what the consumer felt, and what the consumer finally chose.

The most defensible fusion strategy is incremental. A study should first establish that explicit ratings and standard sensory descriptors are reliable. It should then test whether one implicit modality adds predictive or explanatory value beyond those ratings. Only after that should researchers combine modalities. This sequence prevents a common mistake: adding more sensors before showing that any sensor improves the decision. In a product development setting, the value of multimodal fusion may be highest for ambiguous cases, such as products with similar liking scores but different emotional trajectories, products where consumers report acceptance but show strong avoidance cues, or packages that attract attention without improving purchase intent. In these cases, multimodal data can reveal why a simple hedonic mean is insufficient.

For review papers and early-stage industry adoption, the key outcome is not a universal model but a reusable protocol logic. A compact protocol could use eye tracking before tasting to identify information search, FEA during and immediately after tasting to capture affective response, and EEG only for focused studies where neural dynamics are central to the question. This tiered design keeps cost proportional to the research need. It also leaves room for lower-burden physiological signals, such as skin conductance or heart rate, when EEG is unnecessary.

4. Discussion

4.1 Challenges

The first challenge is individual calibration. Consumers differ in expressiveness, gaze strategy, taste sensitivity, cultural display rules, and baseline neural activity. A model trained on young laboratory participants may not generalize to older consumers, children, clinical populations, or cross-cultural markets. Future work should report demographic coverage, test-retest reliability, and participant-level variance rather than only average accuracy.

The second challenge is ecological validity. Traditional sensory booths control noise, light, odor, serving order, and temperature, but real food consumption is social, contextual, and often distracted. XR and remote video methods may improve realism, but they also introduce novelty effects and device artifacts [3]. A balanced protocol should combine laboratory control with realistic consumption scenarios, especially when industry decisions depend on packaging, menu, or retail context.

The third challenge is standardization. The field still lacks shared datasets, common event markers, reporting standards, and benchmarking tasks for implicit food emotion measurement. Studies should describe acquisition hardware, sampling rates, preprocessing, exclusion criteria, synchronization, model validation, and code availability. Without these details, AI performance numbers cannot be compared across products or laboratories.

The fourth challenge is ethics. Facial video, gaze behavior, and EEG are biometric data. They can reveal sensitive information beyond product liking, and consumers may not understand how such data are processed. Responsible food sensory research should use clear consent, data minimization, secure storage, bias audits, and transparent explanations of what the model can infer. Industry use should be especially careful: implicit measurement should help products better fit consumer needs, not manipulate consumers by exploiting unreported vulnerabilities.

4.2 Future Directions

Several future directions follow from these challenges. First, public benchmark datasets are needed for food-specific implicit measurement, because models trained on general emotion or neuromarketing data may not transfer to tasting behavior. Second, studies should compare modality combinations under the same product, participant, and context conditions, so that researchers can estimate the marginal value of each sensor. Third, model reports should include calibration error, uncertainty, and subgroup performance rather than a single accuracy number. Fourth, interdisciplinary teams are necessary. Food scientists understand sensory protocols and product constraints, psychologists understand affective interpretation, engineers understand signal quality, and ethicists can help define acceptable use. Without this collaboration, AI-driven measurement may become technically impressive but scientifically fragile.

5. Conclusion

AI-driven implicit emotion measurement is reshaping food sensory evaluation by adding behavioral and physiological evidence to explicit ratings. Facial expression analysis, eye tracking, and EEG each reveal a different layer of consumer response, and AI makes these high-dimensional signals usable for prediction and interpretation. The main lesson from recent literature is not that machines can replace human sensory judgment. Rather, the best systems combine explicit methods with calibrated implicit signals to obtain a richer, more reliable view of food experience. Progress will depend on multimodal fusion, transparent modeling, ecological validation, and ethical governance. For food companies and researchers, the near-term opportunity is a hybrid evaluation framework: explicit scales for interpretable preference, implicit measures for moment-by-moment affect and attention, and AI models that connect both layers without overstating what any single signal means.

References

[1] Rodrigues, Sandra S. Q., Luis G. Dias, and Alfredo Teixeira. 2024. "Emerging Methods for the Evaluation of Sensory Quality of Food: Technology at Service." Current Food Science and Technology Reports 2: 77–90. https://doi.org/10.1007/s43555-024-00019-7.

[2] Torrico, Damir D., Annu Mehta, and Amalia Bernardes Borssato. 2023. "New Methods to Assess Sensory Responses: A Brief Review of Innovative Techniques in Sensory Evaluation." Current Opinion in Food Science 49: 100978. https://doi.org/10.1016/j.cofs.2022.100978.

[3] Lee, Dongju, Hyemin Jeon, Yoonseo Kim, and Youngseung Lee. 2025. "Integrating Cutting-Edge Technologies in Food Sensory and Consumer Science: Applications and Future Directions." Foods 14 (24): 4169. https://doi.org/10.3390/foods14244169.

[4] Cong, Lei, Siqiao Luan, Erin Young, Miranda Mirosa, Phil Bremer, and Damir D. Torrico. 2023. "The Application of Biometric Approaches in Agri-Food Marketing: A Systematic Literature Review." Foods 12 (16): 2982. https://doi.org/10.3390/foods12162982.

[5] Wagner, J., and J. Hort. 2025. "A Practical Evaluation of Biometric Measures for Understanding the Consumer Experience During Direct Product Evaluation: Current and Future Perspectives." Current Opinion in Food Science 63: 101311. https://doi.org/10.1016/j.cofs.2025.101311.

[6] Lagast, S., X. Gellynck, J. J. Schouteten, V. De Herdt, and H. De Steur. 2017. "Consumers' Emotions Elicited by Food: A Systematic Review of Explicit and Implicit Methods." Trends in Food Science and Technology 69: 172–89. https://doi.org/10.1016/j.tifs.2017.09.006.

[7] Kaneko, Daisuke, Alexander Toet, Anne-Marie Brouwer, Victor Kallen, and Jan B. F. van Erp. 2018. "Methods for Evaluating Emotions Evoked by Food Experiences: A Literature Review." Frontiers in Psychology 9: 911. https://doi.org/10.3389/fpsyg.2018.00911.

[8] Crist, Courtney A., Susan E. Duncan, and Daniel L. Gallagher. 2016. "Protocol for Data Collection and Analysis Applied to Automated Facial Expression Analysis Technology and Temporal Analysis for Sensory Evaluation." Journal of Visualized Experiments, no. 114: 54046. https://doi.org/10.3791/54046.

[9] Kessler, Samuel J., Funan Jiang, and R. Andrew Hurley. 2020. "The State of Automated Facial Expression Analysis (AFEA) in Evaluating Consumer Packaged Beverages." Beverages 6 (2): 27. https://doi.org/10.3390/beverages6020027.

[10] Clark, Elizabeth A., J'Nai Kessinger, Susan E. Duncan, et al. 2020. "The Facial Action Coding System for Characterization of Human Affective Response to Consumer Product-Based Stimuli: A Systematic Review." Frontiers in Psychology 11: 920. https://doi.org/10.3389/fpsyg.2020.00920.

[11] Matsufuji, Yasuyo, Kayoko Ueji, and Takashi Yamamoto. 2023. "Predicting Perceived Hedonic Ratings Through Facial Expressions of Different Drinks." Foods 12 (18): 3490. https://doi.org/10.3390/foods12183490.

[12] Motoki, Kosuke, Toshiki Saito, and Takuya Onuma. 2021. "Eye-Tracking Research on Sensory and Consumer Science: A Review, Pitfalls and Future Directions." Food Research International 145: 110389. https://doi.org/10.1016/j.foodres.2021.110389.

[13] Bilucaglia, Marco, Mara Bellati, Alessandro Fici, Vincenzo Russo, and Margherita Zito. 2025. "Tuning into Flavor: Predicting Coffee Sensory Attributes from EEG with Boosted-Tree Regression Models." Frontiers in Human Neuroscience 19: 1661214. https://doi.org/10.3389/fnhum.2025.1661214.

[14] Wang, He, Hong Men, and Yan Shi. 2025. "Adaptive Identification of Food Sweetness Concentration: An Electroencephalogram Feature Classification Network Under Taste Stimulation." Foods 14 (22): 3855. https://doi.org/10.3390/foods14223855.

[15] Usman, Syed Muhammad, Shehzad Khalid, Aimen Tanveer, Ali Shariq Imran, and Muhammad Zubair. 2025. "Multimodal Consumer Choice Prediction Using EEG Signals and Eye Tracking." Frontiers in Computational Neuroscience 18: 1516440. https://doi.org/10.3389/fncom.2024.1516440.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.