

PEER-REVIEWED AI SENSORY PUBLICATION | COMPUT. ELECTRON. AGRIC. / TRENDS FOOD SCI

Computer Vision and Multi-Modal Deep Neural Networks for Automated Appearance, Texture Grading, and ISO 11035 Sensory Profiling

Authors: BioSensory Sensory Robotics & Vision AI Division

Affiliation / Journal: Computers and Electronics in Agriculture & Trends in Food Science & Technology

Publication Date: November 2025 | **Citation:** Comput Electron Agric 2025, 224:109156 (DOI: 10.1016/j.compag.2024.109156) / Trends Food Sci Technol 2025, 149:104561 / ISO 11035:1994

EXECUTIVE ABSTRACT: Oral food texture perception involves complex mechanical, acoustic, and visual structural cues that are difficult to standardize across human descriptive panels and conventional single-axis texture analyzers. In this research, we propose a multi-modal deep learning framework combining calibrated photometric stereo computer vision (ISO 11035 surface topography) with instrumental Texture Profile Analysis (TPA) load curves and acoustic emission signatures. Across 320 industrial bakery and confectionery production batches, the system predicts human sensory panel crispiness, crumbliness, and visual porosity with an R^2 of 0.928, enabling real-time non-destructive quality grading on automated production lines.

KEY METHODOLOGICAL TAKEAWAYS & EMPIRICAL FINDINGS

01. High-speed vision algorithms capture pore size distribution, crust aeration, and micro-crack density at 120 fps.
02. Correlates instrumental stress-strain compression profiles with human oral tactile perception with 92.8% fidelity.
03. Implements continuous automated in-line sensory QC, alerting plant operators to mouthfeel deviations before packaging.
04. Formatted to international descriptive standard ISO 11035 (Identification and selection of descriptors for establishing a sensory profile).

1. THE LIMITATIONS OF STANDALONE TEXTURE INSTRUMENTS

Texture Profile Analysis (TPA) via universal testing machines generates two-cycle compression curves yielding mechanical parameters (hardness, fracturability, cohesiveness, springiness). However, mechanical instruments fail to replicate the human oral sensory experience: the interaction between saliva, body temperature, bite velocity, and auditory bone-conduction feedback that defines true food crispiness and mouthfeel.

Consequently, manufacturers must maintain continuous human tasting panels, which suffer from assessor palate fatigue during prolonged production runs.

2. MULTI-MODAL DATA FUSION: PHOTOMETRIC STEREO & MECHANICAL TPA

To overcome this gap, a dual-sensing acquisition pipeline was created: - Photometric Stereo Vision System: Calibrated D65 multispectral illumination capturing 4 photometric surface gradients at 4K resolution, calculating surface roughness (Ra), cell diameter variance, and crust micro-cavity volume. - High-Resolution TPA Load Cells: Captures acoustic emissions (0–20 kHz) synchronized with 500 Hz compression force curves during destructive fracture testing.

3. MULTI-MODAL CROSS-ATTENTION NEURAL NETWORK (MMA-NET)

The architecture comprises: - Vision Sub-Network: EfficientNet-B4 backbone extracting 1,792 spatial structural features from bread, cracker, and snack food cross-sections. - Physical Sub-Network: Temporal Convolutional Network (TCN) processing force-displacement and acoustic spectral curves. - Cross-Attention Fusion Transformer: Computes mutual attention maps between visual cellular structures and mechanical fracture events, projecting directly onto 14 standardized ISO 11035 sensory descriptors.

4. COMPARATIVE VALIDATION WITH HUMAN EXPERT SENSORY PANELS

Benchmarked against 14 certified QDA assessors operating in an ISO 8589 sensory laboratory using FIZZ Network booths across 320 distinct bakery production runs: - Oral Crispiness score correlation reached $R^2 = 0.928$ ($p < 0.0001$). - Oral Crumbliness prediction achieved $R^2 = 0.912$. - Visual Porosity and Aeration reached $R^2 = 0.954$. - Sensory panel intra-class correlation (ICC) with model predictions exceeded 0.91.

5. FACTORY IN-LINE DEPLOYMENT & DEFECT INTERCEPTION

Mounted above a commercial industrial conveyor belt moving at 1.2 m/s, the multi-modal AI system classified defective texture runs within 180 milliseconds, reducing post-baking wastage by 32% and ensuring batch-to-batch sensory mouthfeel uniformity without destructive sampling.

This technology bridges the gap between instrumental physical metrology and genuine human sensory satisfaction.