OptiTexture: A Low-Cost, Optical-Deep Learning Solution for Objective Food Texture Assessment in Dysphagia Care

HAIYAN HU, Hong Kong University of Science and Technology, China
JUNYAO PENG, Hong Kong University of Science and Technology, China
TINGTING JIANG, The Third Affiliated Hospital of Sun Yat-sen University, China
QIANYI HUANG, Sun Yat-sen University, China
QIAN ZHANG*, Hong Kong University of Science and Technology, China

Objective food texture assessment for dysphagia care is crucial for preventing life-threatening aspiration in millions of patients. However, current clinical practice relies on subjective tests that are prone to significant error when performed by non-professionals, creating a pressing need for an objective tool. In this work, we propose a low-cost (sub-$20) multimodal system that objectively assesses food texture by jointly fusing time-domain light scattering (TDLS) and RGB imaging. The proposed system leverages the complementary strengths of the two sensing modalities: TDLS captures subsurface light scattering signatures that reflect food microstructure and mechanical properties such as hardness and particle size, while RGB imaging provides surface geometry and macroscopic visual information. A key challenge arises from the fact that low-cost TDLS hardware produces noisy signals in which texture-related information is easily obscured by chemical and environmental interference. To address this, we introduce a physics-inspired multi-scale convolutional feature extractor tailored to the temporal characteristics of light scattering, combined with a contrastive learning strategy that disentangles microstructural texture cues from confounding factors. Furthermore, to effectively fuse these subsurface scattering signals with surface image features, which may conflict, we propose a physical information attention mechanism. This mechanism can dynamically align the two modalities using texture parameters as a guide, enabling conflict-free fusion. Evaluated on 112 clinically-annotated food samples, our system achieves 91.96% accuracy in classifying The International Dysphagia Diet Standardization Initiative (IDDSI) levels (L3-L6), a 22.32% improvement over a vision-only baseline. It also maintains robust performance under varying ambient conditions like illuminations and temperatures. These results demonstrate a practical solution that enables reliable, objective assessment by non-professionals in home and point-of-care settings.

CCS Concepts: • **Human-centered computing** → **Mobile devices**; • **Computing methodologies** → *Machine learning*.

ACM Reference Format:
Haiyan Hu, Junyao Peng, Tingting Jiang, Qianyi Huang, and Qian Zhang. 2026. OptiTexture: A Low-Cost, Optical-Deep Learning Solution for Objective Food Texture Assessment in Dysphagia Care. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 2, Article 45 (June 2026), 24 pages. https://doi.org/10.1145/3810209

1 INTRODUCTION

Dysphagia, a swallowing disorder, impacts over 590 million people worldwide [14, 21]. Epidemiological studies indicate that it affects 51.14% of stroke patients, 48.3% of individuals with neurodegenerative diseases, and 19.2% of

*Corresponding author.

Authors' Contact Information: Haiyan Hu, Hong Kong University of Science and Technology, Hong Kong, China, csehaiyanhu@ust.hk; Junyao Peng, Hong Kong University of Science and Technology, Hong Kong, China, jpengau@connect.ust.hk; Tingting Jiang, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, jiangtt1229@163.com; Qianyi Huang, Sun Yat-sen University, Guang Zhou, China, huangqy89@mail.sysu.edu.cn; Qian Zhang (corresponding author), Hong Kong University of Science and Technology, Hong Kong, China, qianzh@cse.ust.hk.

ACM 2474-9567/2026/6-ART45
https://doi.org/10.1145/3810209

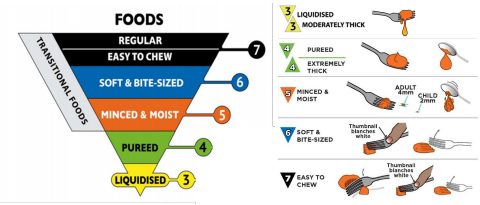


Fig. 1. IDDSI framework of foods and corresponding subjective assess methods.

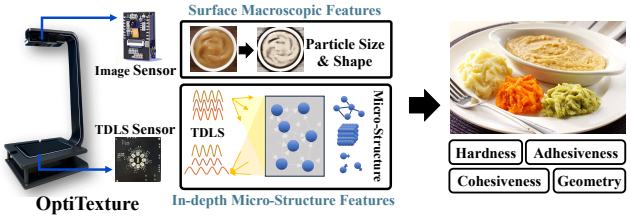


Fig. 2. System rationale: correlating multimodal optical sensing with texture parameters. (TDLS: Time Domain Light Scattering)

otherwise healthy older adults [44]. Due to the decline of swallowing capability, dysphagia significantly increases the risk of aspiration, *i.e.*, the entry of food or liquid into the airway, potentially leading to serious complications such as aspiration pneumonia and malnutrition [3]. To mitigate this risk, patients are often required to consume texture-modified foods, such as purees [31]. The International Dysphagia Diet Standardization Initiative (IDDSI) framework defines food categories from liquid (L0–L4) to food (L3–L7) based on texture properties, along with a practical assessment method using common utensils like forks and spoons (see Fig. 1) [15]. However, these tests rely on human intervention and subjective judgment, and studies have shown that non-professional assessments can have errors exceeding 30% [34]. Therefore, in non-clinical settings like homes or nursing homes where professional support is absent, the reliability of texture assessment by patients or caregivers is concerning. This uncertainty poses a significant safety concern, as inaccurate texture consistency can increase oral residue and thereby elevate the risk of aspiration [8, 24]. Hence, there is an urgent need for objective, quantitative grading methods to ensure mealtime safety.

The primary challenge in achieving objective and accurate assessment of texture-modified foods lies in the inherent complexity of defining semi-solid food textures. Unlike liquid foods, which are differentiated primarily by viscosity, the classification of solid and semi-solid foods (L3-L7) involves multiple interdependent textural parameters. Among these, **hardness, adhesiveness, cohesiveness**, and **geometric properties** are the most critical [4, 10, 14]. These parameters intuitively define characteristics like softness, smoothness, and whether a food is mashed or bite-sized. It is critical to note that accurate IDDSI level determination requires the simultaneous consideration of all parameters, as they must be evaluated jointly, not individually. For instance, a sample with low hardness, low adhesiveness, and low cohesiveness might still be classified as L6 if it contains large particles. Similarly, a puree with fine particles and low adhesion could be considered unsafe (L6) due to excessively high hardness.

Existing approaches, however, struggle to provide this holistic measurement. They are either limited to a subset of parameters (e.g., utensil-based tests) or dependent on sophisticated, costly, and often destructive mechanical instruments (e.g., texture analyzers) confined to laboratory settings [10, 13, 41]. Optical methods like mid-infrared spectroscopy offer non-contact operation but are prohibitively expensive and geared toward composition analysis, not physical microstructure [25, 26]. While portable spectrometers reduce cost, they lack the capability to probe texture-governing microstructural properties [19]. Computer vision, though accessible, is limited to surface geometry and cannot assess key mechanical properties like hardness [29]. Some recent works have successfully obtained the viscosity of liquid food by using vibration-based sensing, but these are difficult to generalize to solid foods or other texture parameters [1, 12, 38]. Consequently, a practical, non-destructive, and cost-effective solution for holistic texture assessment remains an unmet need.

This work overcomes these limitations by introducing a non-contact optical solution for food texture assessment (see Fig. 2). We demonstrate that critical food texture parameters can be accurately derived without

mechanical feedback, ensuring objectivity and convenience. **Our key insight is that by analyzing how light scatters and interacts with food over time, or Time Domain Light Scattering (TDLS), we can infer its microscopic structure.** When pulsed light interacts with food, the temporal changes in scattering intensity reveal important parameters like particle movement and structural rigidity, which directly affect adhesion, cohesion, and hardness [27, 37]. For example, the high viscosity of a thick puree results in a characteristic slow decay of the scattered light signal. Additionally, RGB images can quantify surface geometry features, like particle size and shape. The fusion of these techniques enables a breakthrough in objective food texture detection using low-cost hardware, effectively addressing the core challenge of holistic parameter evaluation.

Implementing this innovative approach with low-cost components involves two primary technical challenges. **Firstly**, the raw optical signals are contaminated with significant interference. To control costs, we use LEDs to emit high-frequency pulses and a photodiode to capture the time-domain spectrum of reflected light. However, this captured spectrum contains not only the desired light-scattering signal but also hardware noise from high-frequency sampling and absorption spectra from the food's chemical components. This interference severely distorts the target signal, compromising the accuracy of texture assessment. **Secondly**, a complex nonlinear relationship exists between the mechanical features extracted from the LED time-domain spectrum and the shape features from RGB images. These features can be complementary in some cases but contradictory in others, leading to suboptimal fusion outcomes. For instance, considering two samples with similar mechanical properties but different surface textures, the two modality will give totally conflict level results. This intermittent alignment-conflict dynamic makes naive fusion approaches unreliable for consistent texture assessment.

To overcome these challenges, we develop a dual-component learning framework, with each component designed to address one specific issue. **(1) To tackle the signal interference challenge**, the key is to robustly extract intrinsic features representing mechanical properties from the noisy data. We discover a strong correlation between the food's mechanical properties and the waveform characteristics of its light-scattering signal across multiple temporal scales. Based on this insight, we design a Multi-scale Convolutional Feature Extractor (MSCFE). By aligning the receptive fields of the convolutional kernels with the natural scales of the time-domain signal, the MSCFE effectively acts as a multi-scale filter bank, extracting robust features related to adhesiveness, cohesiveness, and hardness that are resilient to noise. Furthermore, we introduce a physics-aligned contrastive loss function to guide the model to focus on similarities in physical texture rather than being misled by chemical variations. **(2) To address the feature fusion challenge**, we recognize that cohesion serves as a natural indicator for weighting the relative importance of mechanical and shape features. Based on this insight, we design a Physical Information Attention Mechanism (PIAM). The PIAM utilizes a preliminary estimate of cohesion as a conditional input to dynamically generate sample-specific attention weights, thereby enabling the fusion network to adaptively prioritize the most relevant feature modality and resolve integration conflicts.

We implement a functional prototype using an 8-wavelength LED module and an ESP32 camera. The total hardware cost of the system is under $20. To validate the system's performance, we test it on 112 dysphagia-prone food samples and invite a professional rehabilitation physician to annotate these samples with the IDDSI framework. Given that the risk of misclassification and aspiration for Level 7 (L7) foods is minimal, we adopted a four-category classification system, focusing on classifications L3 through L6. The system achieves an accuracy rate of 91.96%, representing a 22.32% improvement over a baseline that relies solely on computer vision methods [16, 42]. These results underscore the effectiveness of our system in assessing dysphagia-prone foods. We also evaluates the robustness of the system over several environmental conditions like ambient light, temperature, *etc* .

We summarize the contributions as follows:

- We present the first end-to-end, low-cost (sub-$20) system that provides automated, objective IDDSI texture grading, offering a practical tool for caregivers beyond subjective fork tests and lab-bound instruments.

- We introduce a novel sensing paradigm by, for the first time, fusing time-domain light scattering (TDLS) with RGB imaging. This allows us to infer mechanical texture properties from optical scattering dynamics correlated with food microstructure, moving beyond surface-level analysis.
- To overcome the challenges of noisy low-cost signals and feature fusion, we design a physics-informed deep learning architecture, featuring a Multi-Scale Convolutional Feature Extractor for noise-resilient feature extraction and a Physical Information Attention Mechanism for dynamic feature alignment.
- We empirically validate our system on a clinically annotated dataset of 112 food samples, demonstrating not only high accuracy (91.96%) and a significant improvement over vision-based baselines (+22.32%) but also its robustness to environmental variations, confirming its potential for real-world deployment.

2 PRELIMINARIES

2.1 IDDSI Framework and Key Texture Parameters

The International Dysphagia Diet Standardisation Initiative (IDDSI) framework provides a standardized methods for classifying texture-modified foods and drinks. The IDDSI framework defines drinks from Levels 0 to 4, ranging from Thin (Level 0, similar to water) to Extremely Thick (Level 4, with a viscosity exceeding 1,750 mPa·s). For foods, Levels 3 to 7 encompass purees to Regular foods (Level 7) [14]. However, the IDDSI-defined food texture scale relies on subjective performance [15, 34]. To achieve a more objective IDDSI food scale, previous work has examined the correlation between IDDSI food scales and food parameters, revealing that hardness, adhesiveness, and cohesiveness, and geometry are the most important indicators [4, 10, 14].

- **Hardness** refers to the resistance to deformation under applied force, necessitating that foods be soft enough for tongue compression. For instance, foods classified as Level 6 must be crushable by fork tines, ensuring that they can be manipulated safely within the oral cavity.
- **Cohesiveness** measures the ability of food to form a bolus after chewing. Low cohesion can lead to increased pharyngeal residue, which can compromise swallowing safety. For example, crumbly foods are prohibited in Level 5 due to their inability to form a cohesive bolus.
- **Adhesiveness** indicates the stickiness of food to oral surfaces. High adhesion can impede bolus transit, making it difficult for individuals to swallow safely. Foods such as sticky rice are excluded from Level 7 because they can lead to oral and pharyngeal coating, increasing the risk of aspiration.
- **Geometry** refers to the size and shape of food particles. Inappropriate geometry, such as large chunks or sharp edges, can pose a choking hazard or be difficult to manage during swallowing. For safety, IDDSI specifies that foods at Level 6 should be offered in pieces no larger than 1.5×1.5 cm, and small, hard items like nuts are excluded from Levels 4 and 5 to prevent airway obstruction.

Clinically, deviations from these established parameters pose serious risks. Studies have shown that caregivers prepare foods with higher viscosity than standard grading, leading to increased risk of aspiration and oral residue. Although the IDDSI fork test is widely used for texture assessment, inter-rater disagreement is as high as 30% due to subjective evaluation. This highlights the urgent need for objective, quantitative texture grading methods to ensure food safety for patients with dysphagia.

2.2 Rationale of Light Scattering for Food Texture

Time-domain light scattering is a technique that probes material microstructure by analyzing how short pulses of light propagate through and scatter within a medium. When a light pulse enters a food sample, photons undergo multiple scattering events. The temporal profile of the scattered light that exits the sample, often called a temporal point spread function, encodes rich information about the medium's internal structure. The scattering dynamics are influenced by key microstructural properties [17, 39]:

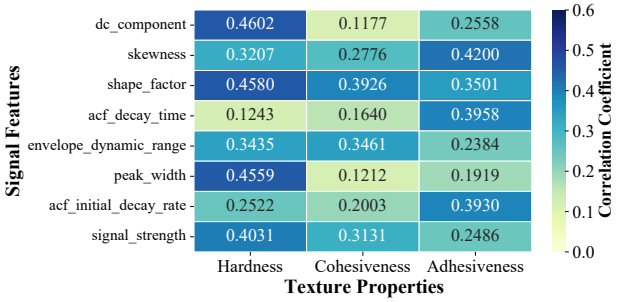


Fig. 3. Pearson correlation between time-domain light scattering features and the texture parameters. (N = 43 samples)

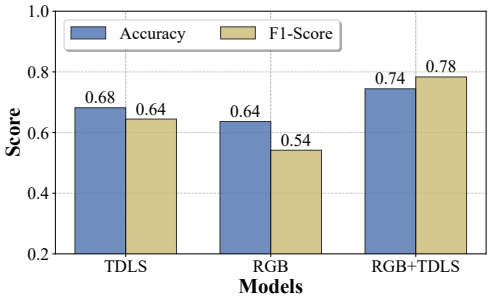


Fig. 4. Comparison of TDLS, RGB and fusion modalities for IDDSI classification. (N = 43 samples)

- **Scattering coefficient (μ_s)**: Primarily related to the density and size of particles. Larger or more numerous particles increase scattering.
- **Reduced scattering coefficient (μ_s')**: Combines the effects of scattering and anisotropy, governing the effective distance photons travel. It is directly influenced by the packing density and aggregation state of the food matrix [35].
- **Absorption coefficient (μ_a)**: Related to chemical composition (e.g., water, pigments).

Crucially, the mechanical texture parameters are themselves manifestations of the food's microstructure. Specifically, hardness and cohesiveness are determined by the strength and arrangement of the internal matrix (e.g., starch network, protein bonds) [2, 33]. A stronger, more connected network will scatter light differently than a weak, crumbly one. Particle size and distribution directly affect the scattering coefficient μ_s [22], and adhesiveness is influenced by the free versus bound water content, which affects the absorption spectrum and the mobility of scattering particles [5]. Therefore, by carefully analyzing the temporal scattering signal, such as quantifying the time decay or temporal broadening of the light pulse, it is possible to infer the very microstructural features that define the macroscopic mechanical properties of interest. This establishes the fundamental physical principle that optical scattering signals can serve as a proxy for mechanical texture.

2.3 Feasibility Validation

Building upon the theoretical rationale presented above, we hypothesize that the mechanical properties of food, governed by its microstructure, dictate its light scattering dynamics. To empirically validate this premise and demonstrate its practical utility for objective food texture assessment, we conduct a feasibility study. Experiments are performed on 43 dysphagia food samples spanning IDDSI Levels 3 to 6 using our collected LED-based time-domain light scattering (LED-TDLS) data and RGB images. From the raw temporal scattering signals, we extract a set of hand-crafted features encompassing temporal statistics, energy distribution, and autocorrelation properties. We first investigate the correlation between these LED-TDLS features and the key texture parameters. Then, we validate the effectiveness of LED-TDLS and modality fusion by using these hand-crafted features for objective IDDSI grading.

Decoding Microstructure through Optical Features. The correlation heatmap in Fig. 3 provides the initial evidence for assessing the feasibility of our approach. It reveals discernible, yet moderate, correlations between TDLS features and key texture parameters. Firstly, the observed patterns are physically interpretable and align with our theoretical framework. For Hardness, the highest correlations are with features like dc_component(0.46) and peak_width(0.46), suggesting that a harder, denser matrix increases scattering, thereby elevating the DC

offset and broadening the pulse. For Cohesiveness, the distinct correlation with shape_factor(0.39) implies that a cohesively bound microstructure creates a uniform scattering environment, resulting in a characteristically shaped waveform. For Adhesiveness, correlations with skewness(0.42) and acf_initial_decay_rate(0.39) indicate that a sticky matrix affects photon migration dynamics, skewing the temporal signal and altering its decay rate. It is important to note that this analysis is a preliminary feasibility study from the perspective of individual features. The observed correlations, while significant and interpretable, are modest because each feature captures only a part of the complex scattering process. Thus, no single feature can fully characterize texture properties, necessitating multi-feature fusion and advanced models such as deep learning to integrate the distributed information. This finding underscores the necessity of the proposed method.

IDDSI Grading Performance. The performance comparison demonstrates the translational value of these correlations, as shown in Fig. 4. All results are obtained using a Random Forest classifier with identical settings across different feature configurations. The model using only TDLS features (TDLS) achieves higher accuracy of 68% than the RGB-only baseline with only 64%. This key result confirms that TDLS features that derived from subsurface scattering dynamics provide superior predictive power for the mechanical properties (hardness, cohesiveness, adhesiveness) critical to swallowing safety, which are not fully captured by surface images. Most notably, the fusion of TDLS and RGB features (RGB+TDLS, implemented by concatenating the two feature sets and training a unified Random Forest classifier) yields a clear performance improvement, achieving an accuracy of 74% and an F1-score of 78%. Considering the limited sample size (N = 43), we assess the robustness of this improvement using bootstrap resampling. The resulting effect size estimates and bootstrap confidence intervals consistently indicate a positive performance trend in favor of the fused model (95% CI: [0.02, 0.23]). These results collectively suggest that TDLS and RGB modalities provide complementary information for IDDSI grading.

3 SYSTEM DESIGN

This section presents the design of OptiTexture, a multimodal system for objective IDDSI food texture classification. Building on our feasibility study, which establishes that TDLS signals encode key texture parameters and that fusion with RGB data improves performance, we introduce a system to overcome the core challenge: extracting robust features from noisy LED-TDLS data and effectively fusing them with RGB shape features. The design centers on two novel components: a Multi-Scale Physics Encoder (MSPE) for noise-resilient TDLS feature extraction, and a Physics-Informed Attention Mechanism (PIAM) for adaptive multimodal fusion.

3.1 Multi-Scale Physics Encoder

Although hand-designed features demonstrate a correlation between time-domain laser scattering signals and texture parameters, their practical effectiveness is limited by noise sensitivity and poor generalization. This limitation becomes particularly pronounced when using a low-cost 1kHz LED-photodiode array, where the original scattering signal $S(t)$ is severely degraded. The measured signal thus becomes: $R(t) = S(t) + A(t) + \eta_(t)$. Here, $A(t)$ represents absorption and reflection from the sample's chemical composition, while $\eta_(t)$ denotes composite noise from the low-cost sensor, including high-frequency thermal noise, shot noise, low-frequency drift, and multiplexing artifacts [18, 36]. For texture analysis, both $A(t)$ and $\eta_(t)$ act as interference.

This interference poses a dual challenge. **First**, the noise spectrum overlaps with that of the physical processes of interest, rendering conventional single-scale filtering or feature extraction ineffective. **Second**, and more critically, this noise blurs the mapping between microstructure and signal characteristics. In complex noise backgrounds, similar microstructures can produce significantly different signals, while different microstructures may yield similar ones, leading to a misalignment between the feature space and the physical property space.

To address this issue, it is essential to consider the physical origin of the TDLS signal. Food texture does not arise from a single characteristic process but from a hierarchy of microstructural dynamics occurring over distinct temporal ranges, including particle–interface interactions, adhesion events, network relaxation, and slow

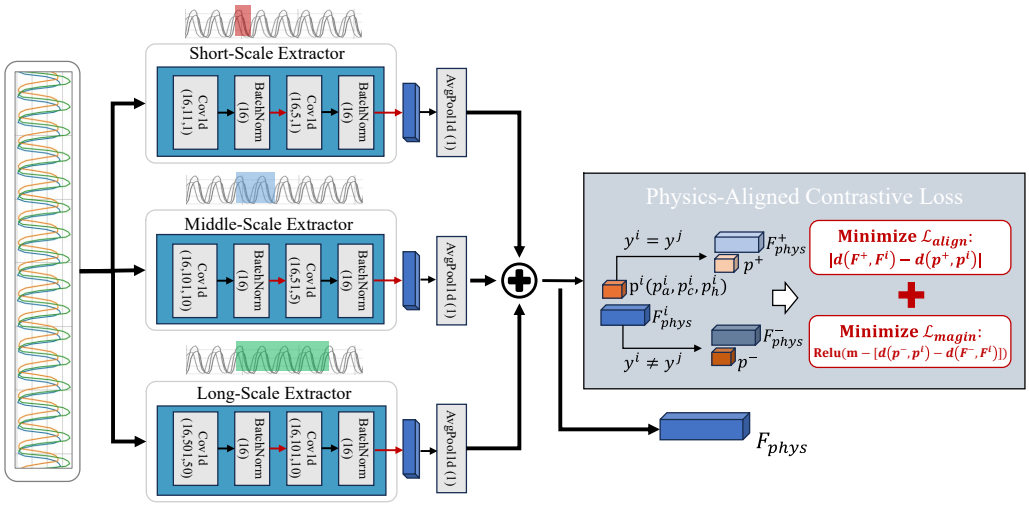


Fig. 5. Illustration of the architecture of Multi-Scale Physics Encoder, which includes a Multi-Scale Convolutional Feature Extractor and a Physics-Aligned Contrastive Loss.

structural rearrangements. As a result, texture-related information is intrinsically distributed across multiple temporal resolutions rather than concentrated at a single frequency band. This multi-scale interpretation is further supported by empirical evidence from our preliminary correlation analysis (Fig. 3). Individual handcrafted TDLS features exhibit statistically significant yet moderate correlations with different texture parameters. Notably, hardness correlates more strongly with low-frequency or global features (e.g., DC component and peak width), cohesiveness with waveform shape descriptors, and adhesiveness with higher-order temporal statistics such as skewness and autocorrelation decay rates. The moderate magnitude of these correlations indicates that no single feature or temporal scale is sufficient to fully characterize texture properties.

These observations imply that feature extraction at a single, fixed temporal scale is inherently incomplete. Capturing texture-relevant information in TDLS signals therefore requires an explicit multi-scale representation whose receptive fields align with the characteristic time scales of the underlying physical processes. Motivated by this consideration, we propose a Multi-Scale Physics Encoder (MSPE), which embeds scale awareness directly into the feature extraction process through parallel convolutional pathways operating at different temporal resolutions. The encoder further incorporates a physics-aligned contrastive constraint to structure the feature space in accordance with physical texture similarity, as detailed in the following subsections.

Multi-Scale Convolutional Feature Extractor. This component aims to address the challenge of feature extraction in noisy environments. The fundamental design principle is to explicitly align the receptive field of convolutional kernels with the feature timescales of different physical processes, thereby directly targeting the extraction of noise-robust features correlated with texture. Our approach exploits the differential sensitivity of different kernel sizes to temporal frequency components by defining $\Delta t \approx (k_s - 1) \times T_s$, where k_s is the kernel size and T_s is the sampling interval, thereby simulating filters of different scales based on the convolutional neural network kernel size. Although this expression relates Δt to k_s, the design process is conceptually reversed: the characteristic temporal ranges of interest are first identified based on the physical dynamics of food texture evolution, and the kernel sizes are subsequently chosen to approximate these ranges under a fixed T_s.

Specifically, we include three types of convolutional kernels of different scales: short-term kernels, medium-term kernels, and long-term kernels (see Fig. 5). Short-term kernels are characterized by compact receptive

fields ($k_s = 11$)), acting as matched filters sensitive to fast transients. By focusing on adjacent frequency bands and suppressing low-frequency drift, the signal-to-noise ratio of localized micro-collision and adhesion events is inherently improved. Under the adopted sampling interval T_s, this scale captures fast transient dynamics on the order of several milliseconds. Medium-term kernels have a medium receptive field ($k_s = 101$). They effectively average out the effects of high-frequency noises on adhesion while mitigating some residual shot noise. The extracted features capture the characteristic decay constant and shape parameter of the relaxation process. This scale integrates information over tens of milliseconds, corresponding to intermediate relaxation behavior. Long-term kernels have the widest receptive field ($k_s = 1001$) and are able to integrate signal behavior on timescales associated with slow structural rearrangements. They inherently filter out medium- and high-frequency noise, effectively recovering long-term trends. Features at this scale can quantify baseline shifts, low-frequency oscillation amplitudes, and trend persistence. This scale aggregates information over hundreds of milliseconds, reflecting slow structural rearrangements and baseline evolution. Mathematically, the convolution operation at each scale k can be viewed as a structured projection:

$$F_k(t) = (R \cdot K_k)(t) = \int R(\tau) \cdot K_k(t - \tau) d\tau, \tag{1}$$

where $K_k(\tau)$ is a kernel function optimized to capture the expected morphology at scale k. The multi-scale design ensures complementarity among the feature sets. Therefore, the multi-scale extractor functions as a parallel, physics-based feature extractor. Each path isolates discriminative texture patterns associated with a specific physical process while simultaneously suppressing the dominant noise type that plagues that particular temporal region. This collaborative filtering effect, enabled by the scale-specific receptive fields, is key to the robustness of our extracted features compared to single-scale feature extraction after simple denoising.

Physics-Aligned Contrastive Constraint. The Multi-Scale Convolutional Feature Extractor addresses the first challenge of disentangling multi-scale physical processes from overlapping noise. However, variations in chemical composition $A(t)$ can cause misalignment: texturally identical samples with different chemistries may have divergent features, while texturally different samples may have convergent features, confusing the model. This undermines the physical interpretability and generalizability of the model.

To directly enforce an intrinsic alignment between the feature space and the physical property space, we introduce a Physics-Aligned Contrastive Constraint. Conventional contrastive learning, which relies on fixed thresholds applied to potentially noisy texture measurements, is susceptible to the very interference we aim to overcome. A slight perturbation in the measured texture parameters of a sample pair (i, j) from noise could arbitrarily flip the similarity label, thereby misguiding the feature embedding. Let $\mathbf{p}_i$ denote the measured physical texture parameter vector of sample i, defined as $\mathbf{p}_i = [p_a^i, , p_c^i, , p_h^i]$, where p_a^i, p_c^i, and p_h^i correspond to the adhesion-related parameter, characteristic relaxation constant, and hysteresis-related descriptor, respectively. Our key insight is to leverage the semantic IDDSI class labels y_i as robust, high-level proxies for physical similarity. We posit that samples within the same IDDSI category, despite measurement noise and intra-class variance, share a common physical essence. This categorical prior provides a stable supervisory signal to structure the feature space, circumventing the fragility of threshold-based methods. The constraint includes two items to correct the misalignment: intra-class consistency loss $\mathscr{L}_{\text{align}}$ and inter-class margin loss $\mathscr{L}_{\text{margin}}$:

$$\mathscr{L}_{\text{align}} = \mathbb{E}_{(i,j)\in\mathcal{P}_{\text{same}}} \left[\left| d(F_i, F_j) - d(p_i, p_j) \right| \right], \tag{2}$$

$$\mathscr{L}_{\text{margin}} = \mathbb{E}_{(i,j)\in\mathcal{P}_{\text{diff}}} \max\left(0, m - \left[(1 - d(F_i, F_j)) - (1 - d(p_i, p_j))\right]\right), \tag{3}$$

where $\mathcal{P}_{\text{same}} = \{(i, j) | y_i = y_j\}$ and $\mathcal{P}_{\text{diff}} = \{(i, j) | y_i \neq y_j\}$, $d(F_i, F_j)$ and $d(p_i, p_j)$ indicate the cosine_similarity between features and texture parameters. Therefore, $\mathscr{L}_{\text{align}}$ ensures that within a category, the relative distances in the feature space faithfully reflect the distances in the physical texture space. It penalizes the model if the feature similarity $d(F_i, F_j)$ deviates from the ground-truth texture similarity $d(p_i, p_j)$ for samples of the same class. On

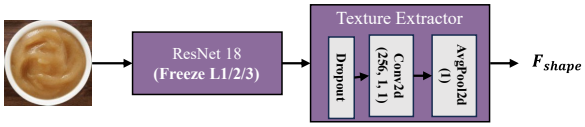


Fig. 6. The structure of surface shape feature extractor that used to process RGB images.

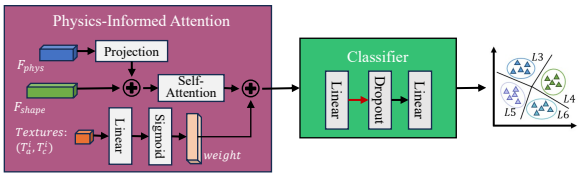


Fig. 7. The structure of Physics-Informed Attention Module and the classifier for IDDSI grading.

the other hand, $\mathscr{L}_{\text{margin}}$ acts as a regularizer to prevent the features of different categories from overlapping due to interference. It enforces a safety margin m (set as 0.2 empirically) such that the feature-space distance between samples from different classes must exceed their texture-space distance.

The complete physics-aligned loss integrates these complementary objectives: $\mathscr{L}_{\text{phy}} = \mathscr{L}_{\text{align}} + \lambda \cdot \mathscr{L}_{\text{margin}}$. In conjunction with the Multi-Scale Feature Extractor, this hierarchical constraint ensures that the final feature representation F_i is not only extracted from noise-resistant signal components but is also explicitly structured in the embedding space to reflect the true physical relationships among samples. Ablation study in § 5.3 validates the effectiveness of this module.

3.2 Physics-Informed Attention Mechanism

Our multi-scale physics encoder effectively quantifies fundamental texture parameters, such as hardness and cohesion, from LED signals. However, these physical measurements alone are insufficient for accurate IDDSI classification. The IDDSI framework incorporates not only texture but also morphological characteristics. For example, oatmeal typically exhibits the low hardness and cohesion of a L4 food, yet it must be classified as L5 if it contains particles larger than 5 mm. Conversely, cookie crumbs without large particles may qualify as L6 due to their high hardness. This discrepancy underscores the necessity of incorporating visual shape information.

To this end, we integrate a supplementary, low-cost RGB imaging module. As depicted in Fig. 6, our shape feature extractor uses a pre-trained ResNet18 backbone, which is a typical backbone in image analysis tasks [11], followed by a custom convolutional layer that projects the features into a dedicated texture-relevant space. Considering the limited size of the specialized dataset (N = 112), the RGB branch is deliberately designed with controlled trainable capacity. Specifically, the initial three residual stages of the pre-trained ResNet18 are frozen during training, and only higher-level semantic layers together with the lightweight projection head are updated. This design constrains the effective degrees of freedom of the visual branch while retaining sufficient representational power for texture-related cues. This module enables the identification of critical morphological parameters, such as particle size distribution, thereby compensating for the limitations of the physical data alone.

Nevertheless, integrating the two modalities, *i.e.*, physical LED texture parameters and visual RGB shape features, poses significant challenges. The complex and often nonlinear interactions between these feature spaces give rise to three distinct scenarios, each requiring tailored fusion strategies. In co-registration cases, like pumpkin puree, both modalities yield consistent results. However, in shape dominance conflicts, such as with oatmeal, physical parameters suggest a classification of level 4, while visual features that indicate large particles ($> 5mm$) lead to a classification of level 5. Conversely, in physical dominance conflicts, as seen with broken cookies, visual texture suggests level 4, but the excessive hardness indicates a classification of level 6. Therefore, in such complex circumstances, existing static fusion techniques are unable to adapt to these context-dependent conflicts effectively. This insight emphasizes the need for more dynamic and responsive fusion strategies that can accommodate the complexities of heterogeneous feature integration.

To solve this problem, we observe that the relative importance of texture parameters versus shape features in determining the final IDDSI classification is heavily contingent upon the inherent material properties of the sample. A key metric in this regard is the sample's measured cohesion, a fundamental textural property that reflects the strength of internal bonds. Specifically, samples exhibiting high cohesion tend to form stable structures, allowing hydration effects to manifest predictably across the entire spectral band, thereby rendering the physical parameters more reliable. In contrast, low cohesion is often associated with heterogeneous mixtures or fragile structures, where particle size and shape become critical discriminants.

This discovery forms the foundation of our Physically-Informed Attention Mechanism (PIAM), whose structure is shown in Fig. 7. PIAM leverages the measured physical parameters, notably cohesion, as a dynamic gating signal. Instead of treating all features equally, PIAM learns to assign sample-specific attention weights, effectively letting the physics of the sample itself guide the fusion process. Formally, given the LED-derived physical feature vector F_{phys} and the RGB-derived shape feature vector F_{shape}, we first map the LED features into the RGB feature space through a linear projection: $\tilde{F}_{\text{phys}} = W_p F_{\text{phys}} + b_p$, where W_p and b_p denote the learnable projection parameters. This operation aligns the physical and visual features into a common embedding space with identical dimensionality. The aligned features are then concatenated to form the joint representation $F_{\text{cat}} = [\tilde{F}_{\text{phys}}; F_{\text{shape}}]$. Then we fed the concatenated feature vector into a lightweight gating network implemented as a two-layer multilayer perceptron with a sigmoid activation, as illustrated in Fig. 7. The gating network produces a scalar weight $\alpha_i = \sigma(W_2 \cdot \tanh(W_1 \cdot F_{\text{cat}})) \in (0, 1)$, where α_i is computed independently for each sample i and controls the relative contribution of the two modalities. The final fused representation is obtained through a linear combination of the modality-specific features $F_{\text{cat}} = \alpha_i \cdot F_{\text{shape}} + (1 - \alpha_i) \cdot \tilde{F}_{\text{phys}}$. This sample-level gated fusion allows the model to adaptively emphasize either RGB-derived shape cues or LED-derived physical cues based on the underlying texture characteristics of each sample, while maintaining a fixed feature dimensionality for subsequent classification.

To dynamically adjust the relative contribution of the two modalities, PIAM introduces a physically conditioned attention module. Specifically, the physical texture parameter vector $\mathbf{p} = (p_c, p_a)$, representing cohesion and adhesion, is passed through a lightweight multilayer perceptron (MLP) with one hidden layer and a sigmoid activation function to generate a scalar attention weight:

$$\mathbf{Attn}_{weight} = \sigma(W_2 \cdot \delta(W_1 \cdot [F_{\text{cat}}; p] + b_1) + b_2), \tag{4}$$

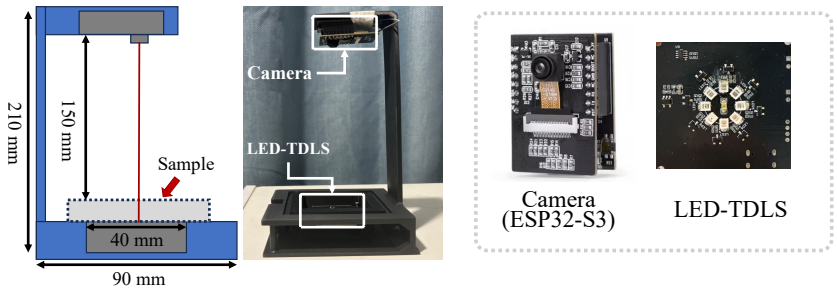
Where (W_1, b_1, W_2, b_2) are learnable parameters, $\delta(\cdot)$ is the ReLU activation function, and $\sigma(\cdot)$ is the Sigmoid activation function. Importantly, the physical features $p = (p_c, p_a)$ are integrated as explicit conditioning signals, enabling the model to dynamically recalibrate feature importance based on the inherent physical state of the sample. Therefore, for highly cohesive samples, PIAM enhances the contribution of reliable physical features, whereas for heterogeneous mixtures (i.e., low cohesion), PIAM shifts its focus to key RGB-derived shape features. During the training phase, these weights are stored and applied in the interference stage without texture input.

3.3 Objective Function

In summary, the overall loss function is defined as follows:

$$\mathscr{L}_{\text{total}} = \mathscr{L}_{\text{task}} + \alpha \cdot \mathscr{L}_{\text{phy}} + \beta \cdot \mathscr{L}_{\text{orth}}, \tag{5}$$

where $\mathscr{L}_{task}$ represents the primary classification loss, $\mathscr{L}_{\text{phy}}$ refers to the physics-aligned contrastive loss, and $\mathscr{L}_{orth} = d(F_{phys}, F_{shape})$ encourages orthogonality between the normalized physics features and shape features. By minimizing $\mathscr{L}_{orth}$, the feature vectors are driven toward orthogonality in the latent space, which enhances feature diversity. This constraint ensures that the physics feature extractor captures intrinsic material properties, while the vision feature extractor focuses on morphological aspects.



(a) System Prototype (b) Hardware Modules

Fig. 8. The prototype of OptiTexture, which cost <$20.

Table 1. Detailed Bill of Materials (BOM) for OptiTexture Prototype.

Category	Component	Model / Part No.	Qty.	Cost (USD)
MCU	Main Controller	STM32F407VGT6	1	1.44
Camera	RGB Camera Module	ESP32-Camera (Hiwonder)	1	5.56
Interface	USB-UART	CH340E	1	0.18
Optical Sensor	Photodiode	0090-3111-185	1	1.69
Analog Front-End	Op-Amp	OPA333AIDBVR	1	0.42
NIR Emitters	LEDs (850–1650 nm)	FY2502 Series	8	5.40
Power	LDO / Current Source	AMS1117-3.3, LT3092	2	1.70
Clocking	Crystal Oscillators	8 MHz, 32.768 kHz	2	0.20
Passives	Resistors & Capacitors	0402/0603 Series	–	0.60
Connectors	USB-C, Headers, Switch	ZX-XH / PH Series	–	0.60
Total				**16.4**

4 IMPLEMENTATION

Fig. 8 depicts the system prototype, a compact platform measuring 90 mm (L) × 85 mm (W) × 210 mm (H). During operation, samples are placed directly on the bottom surface of the platform, which then initiates data acquisition by automatically controlling the integrated camera and spectral modules. Each sample measurement is completed within approximately 2 seconds. The prototype integrates two core sensing subsystems: an RGB imaging module and a TDLS optical module consisting of multi-wavelength NIR LEDs, a photodiode, and dedicated driving and readout circuitry. The system is powered and accessed via an external USB connection to a host computer, without an onboard battery. To improve cost transparency, Table 1 summarizes the hardware cost based on the implemented bill of materials (BOM). The table explicitly includes the RGB camera module, the TDLS optical module, the microcontroller and interface circuitry for data acquisition and control, as well as power regulation and passive components. All components are commercial off-the-shelf parts. Based on this detailed BOM, the total cost of the system electronics is approximately $16.

Camera Module. An ESP32-S3 camera module is installed at the top of the platform, chosen for its cost-effectiveness and built-in autofocus capability. Samples are positioned directly beneath the camera on the platform surface, with a fixed distance of 15 cm maintained between the camera lens and the sample. This distance optimizes the system's footprint while providing a sufficient field of view (approximately 60 cm × 60 cm at the sample plane) to capture comprehensive images of the food.

(a) Examples of our dataset (b) GT setup (c) IDDSI Distribution

Fig. 9. The setup of the evaluation. (a) Partial examples of our dataset. (b) The setup of of ground truth. (c) The IDDSI level distribution of our dataset.

LED-TDLS Module. LED-TDLS module located at the base of the platform, is designed to capture time-resolved diffuse reflectance profiles of paste-like foods within the short-wave near-infrared (SW-NIR) region. The module employs an array of eight NIR LEDs with central wavelengths of 850 nm (FY25021904), 940 nm (FY25021905), 1050 nm (FY25021906), 1200 nm (FY25021907), 1300 nm (FY25021908), 1460 nm (FY25021909), 1550 nm (FY25021910), and 1650 nm (FY25030401). This specific spectral range is selected based on the characteristic molecular absorption bands of key components that influence food texture within the short-wave near-infrared (SW-NIR) region. For paste-like foods, critical texture attributes such as moisture content, particle size or structure exhibit strong absorption or scattering signatures in this range. Each LED operates at 0.2 W, selected for high spectral stability and consistent radiant intensity across the measurement range. Each LED is driven by a 1 kHz pulsed signal with a 50% duty cycle and activated sequentially to prevent spectral overlap. The driving circuitry adopts a 56 mA constant-current source combined with a MOS-switch driver, ensuring precise pulse control, stable optical output, and minimal thermal drift during operation. The reflected light is detected by a broadband photodiode (900–1700 nm response range), enabling high-speed, time-resolved acquisition of diffuse reflectance. For each illumination event, 10,000 temporal samples are collected over a 100 ms window, providing robust averaging for transient light propagation analysis, which correlates strongly with key texture attributes such as moisture and fat distribution. Additionally, system control and data acquisition are managed by an STM32F103C8T6 microcontroller, responsible for LED switching, timing synchronization, and data transfer. Its integrated 12-bit ADC and DMA interface provide efficient, low-latency sampling performance.

5 PERFORMANCE EVALUATION

5.1 Study Setup

5.1.1 Data Collection. We assemble a dataset comprising 112 distinct commercially available food items categorized for use in dysphagia diets, representing International Dysphagia Diet Standardisation Initiative (IDDSI) levels 3 to 6. This dataset includes 19 pre-formulated ready-to-eat dysphagia foods, while the remaining 93 items consist of prepared food products (e.g., cereal powders, oatmeal) commonly recommended for middle-aged and elderly populations. Following the sample preparation protocol detailed in reference [10], each food powder is reconstituted into test samples at 4 to 5 distinct concentrations to achieve a range of texture levels consistent with the target IDDSI grades. Fig. 9(a) displays a representative subset of the prepared food samples.

All samples undergo a single scanning session using our prototype device, which generates the following data outputs: (1) A 320×320 pixel RGB image. (2) A set of time-domain spectral data, structured as (8 channels × 10,000 data points), corresponding to the sequential excitation of the eight LEDs. Furthermore, we characterize the textural properties of the samples using instrumental Texture Profile Analysis (TPA) with a TA.XT Plus texture analyzer (Stable Micro Systems, UK). A 20-mm cylindrical probe compresses each sample twice to 50% of

its original height, and key parameters (hardness, cohesiveness, adhesiveness) are derived from the resulting force-time curve.[1]

5.1.2 Ground Truth. The IDDSI levels of the samples are determined by a rehabilitation therapist using standardized tests from the International Dysphagia Diet Standardisation Initiative (IDDSI) framework to ensure clinical validity. We conduct two major IDDSI tests:

(1) **Fork Pressure Test**: This test is used to assess the sample's ability to hold its shape. A standard stainless steel fork with four 4-mm gaps between the tines is used. A 10-mL slip-tip syringe is employed to deposit the sample onto the fork. The test is considered passed (indicating thicker, IDDSI Level 4 consistency) if the sample does not drip through the prongs within 10 seconds after the syringe is emptied.
(2) **Spoon Tilt Test**: This test is used to evaluate the sample's cohesiveness and tendency to flow. A standard teaspoon is filled with the sample and tilted sideways. The behavior of the sample (e.g., whether it poured easily, fell as a single blob, or held its shape on the spoon) is observed and recorded against the official IDDSI criteria to determine the appropriate level.

All samples are tested within 2 hours of preparation to prevent changes in physical properties. The testing environment is strictly controlled at a temperature of 25 ± 1° C and a relative humidity of 50 ± 5% to minimize external variability. Fig. 9(c) illustrates the distribution of all samples across the IDDSI framework. To assess labeling consistency, an independent rehabilitation therapist re-labeled all samples based on recorded IDDSI testing procedures following the same standardized criteria. The inter-rater agreement between the two therapists achieved an exact agreement rate of 83.0%, indicating substantial inter-rater consistency and supporting the reliability of the ground truth annotations.

5.1.3 Metrics. To comprehensively evaluate the performance of our IDDSI-level grading system, we employ a set of metrics that assess both technical accuracy and clinical reliability: **accuracy, precision, recall, macro-averaged F1-score,** and **Cohen's Kappa**. **Accuracy** indicates the overall classification correctness. **Precision** shows the risk of falsely recommending an unsafe consistency, crucial for patient safety. **Recall** refers to the detection of genuinely hazardous samples. **Macro-averaged F1-score** balances precision and recall across all IDDSI levels. Meanwhile, **Cohen's Kappa** measures clinical reliability by assessing agreement with an expert therapist beyond chance.

For the multi-class classification task across IDDSI Levels 3–6, all metrics are first computed on a per-class basis. Overall performance is then summarized using a **sample-size–weighted average**, defined as:

$$M_{\text{weighted}} = \sum_{i=1}^{C} \frac{N_i}{\sum_{j=1}^{C} N_j} \cdot M_i, \tag{6}$$

where M_i denotes the class-wise metric value for the i-th IDDSI level, N_i is the number of samples in that level, and $C = 4$ corresponds to IDDSI Levels 3–6. Unless otherwise specified, all reported overall performance metrics in this study (*i.e.*, score) refer to the sample-size–weighted averages across IDDSI Levels 3–6.

5.1.4 Baselines. Given the lack of low-cost solutions for IDDSI food grading, we establish baselines using previously employed methods for food texture detection:

- **RGB-RandomForest**: Inspired by [16], which introduces visual features such as Gray-Level Co-occurrence Matrix (GLCM) and color-based descriptors for food texture analysis, we extract a total of 547 handcrafted features from the RGB data. These features include color distribution and color moment features, statistical

[1](1) Hardness is the peak force (N) during the first compression cycle. (2) Cohesiveness is calculated as the ratio of the positive work area of the second compression to that of the first compression. (3) Adhesiveness is determined as the negative work area (N·s) representing the force required to withdraw the probe from the sample after the first compression.

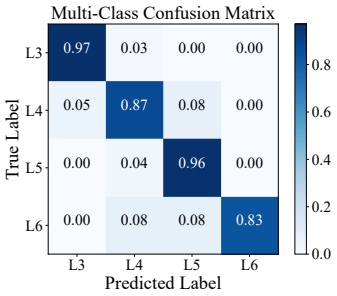


Fig. 10. Confusion matrix of OptiTexture's IDDSI classification.

Table 2. OptiTexture's IDDSI classification results. (N = 112 Samples)

IDDSI Levels	Accuracy	Precision	Recall	F1-Score
Level 3 (N=35)	97%	94%	97%	96%
Level 4 (N=38)	87%	92%	87%	89%
Level 5 (N=27)	96%	87%	96%	91%
Level 6 (N=12)	83%	100%	83%	91%
Weighted Average	**92%**	**93%**	**91%**	**92%**

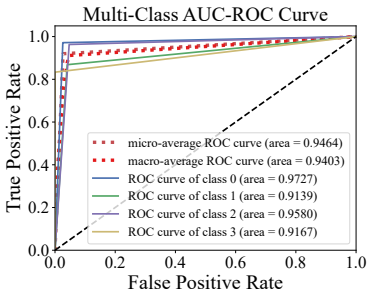


Fig. 11. AUC-ROC curve of OptiTexture's classification results. (N = 112 Samples)

Table 3. 10-fold cross-validation performance of OptiTexture and baseline models on IDDSI level classification. Values are reported as weighted score ± standard deviation across folds. (N = 112 Samples)

Models	Accuracy	F1-Score	Kappa
RGB-RandomForest	0.67 ± 0.11*	0.61 ± 0.14	0.53 ± 0.15
RGB-ResNet18	0.70 ± 0.09*	0.61 ± 0.11	0.57 ± 0.14
RGB+TDLS	0.72 ± 0.14*	0.71 ± 0.21	0.61 ± 0.23
OptiTexture	**0.92 ± 0.06**	**0.92 ± 0.10**	**0.89 ± 0.09**

* indicates $p < 0.01$.

texture descriptors, as well as simple structural statistics related to particle and region distributions. [2] These features are then fed into a Random Forest classifier to form our first baseline.

- **RGB-ResNet18**: Following [42], where computer vision and deep learning models are used to analyze meat texture, we employ a pre-trained ResNet18 model from torchvision.models, which has been trained on the ImageNet dataset. The model is fine-tuned on the region of interest (ROI) of our RGB data for the specific task of food texture analysis.
- **RGB+TDLS**: We further extend the first baseline by integrating both RGB and TDLS data. This multimodal setup simulates the best achievable performance of existing solutions when multiple data sources are available.

5.2 Overall Performance

We first evaluate OptiTexture's performance in IDDSI levels classification by conducting stratified 10-fold cross-validation.

5.2.1 OptiTexture Performance. The OptiTexture system achieves an overall accuracy of 92% across 112 mixed food samples, demonstrating robust classification capability. Performance consistency across folds is further

[2]Due to space limitations, we summarize the feature categories here, while providing the complete feature definitions and extraction pipeline in our open-sourced code at https://github.com/hyanhu/Optitexture.

supported by a low standard deviation (0.06), indicating stable generalization despite the limited dataset size. Performance variation across texture levels provides important practical insights. The system achieves exceptional performance for Level 3 paste-like foods, with accuracy and recall exceeding 97% and an F1-score of 0.96. This high reliability is particularly valuable for individuals with severe dysphagia who require precise texture identification for safe consumption. Level 5 foods also show strong results with accuracy and recall above 96% and an F1-score of 0.91. In comparison, Level 4 foods exhibit a higher rate of missed detections with 87% accuracy, indicating increased classification ambiguity for intermediate texture levels. Level 6 foods achieve perfect classification accuracy but a lower recall of 83%, suggesting that a subset of Level 6 samples is misclassified as adjacent lower texture levels. The confusion matrix in Fig. 10 further illustrates OptiTexture's discriminative capability across all categories, showing clear separation between Level 3 and Level 4 samples. This performance stems from the strong discriminative power of time-domain LED spectral features incorporated in the model. Most misclassifications occur between adjacent texture levels, particularly between Level 5 and Level 6, reflecting the subtle physical differences between these categories. However, classification of Level 5 and Level 6 samples shows room for improvement, potentially due to limitations in the camera module's spatial resolution when distinguishing subtle textural variations. In addition, Fig. 11 presents the multi-class ROC curves obtained under 10-fold cross-validation. OptiTexture achieves a macro-average AUC of 0.940 and a micro-average AUC of 0.946, with consistently high class-wise AUCs across all IDDSI levels. These results indicate strong discriminative performance and provide further evidence that the model does not overfit the limited training data.

5.2.2 Baselines Comparison. Fig. 3 compares the classification results of our OptiTexture system with several baseline systems and highlights three key findings. First, the baseline that utilizes only RGB features performs poorly, achieving an accuracy of just 67%. This result underscores the limitations of using basic image features in effectively detecting texture parameters. Second, the computer vision-based baseline using a ResNet18 backbone achieves a slightly higher accuracy of 67%. This improvement indicates that more advanced feature extraction from food surface images is beneficial for IDDSI food texture classification. However, the limited availability of surface texture information restricts the system's ability to differentiate between L3 and L4 food samples, as these two categories exhibit highly similar visual characteristics when static. Third, by integrating time-domain spectroscopy with image analysis techniques, the classification performance of the system improves to 72%. This enhancement demonstrates the complementary roles of the two modalities in IDDSI food texture classification.

OptiTexture significantly outperforms all these baselines, achieving an accuracy of 92%, along with an F1-score of 92% and a Kappa value of 0.89 in the four-class classification task. Notably, OptiTexture also exhibits the lowest performance variance across folds, indicating improved generalization and reduced overfitting compared to baseline models. To further verify that the observed performance improvements are not due to random variation, we conducted statistical significance analysis on classification accuracy using McNemar's test. As indicated in Table. 3, the accuracy improvement of OptiTexture over all baseline methods is statistically significant ($p < 0.01$). The superior performance of OptiTexture can be attributed to two design enhancements: the advanced feature extraction capabilities based on LED time-domain spectra and the improved utilization of modalities achieved through a texture-guided modality fusion strategy.

5.2.3 Human Results Comparison. To further evaluate the practicality of OptiTexture, we compare its performance against objective human labels. We recruit 10 participants representing the profile of non-professional but educated caregivers (e.g., graduate students and research staff), each of whom receives standardized training: a 20-minute self-study session using the IDDSI framework guidelines and instructional videos, followed by a 5-minute guided classification of three exemplary samples. Finally, each user independently classifies 20 unknown samples within 30 minutes.

As illustrated in Fig. 12, individual human performance varies substantially, with accuracies ranging from 45% to 75% and F1-scores from 0.40 to 0.75. This variability reflects the inherent subjectivity and difficulty

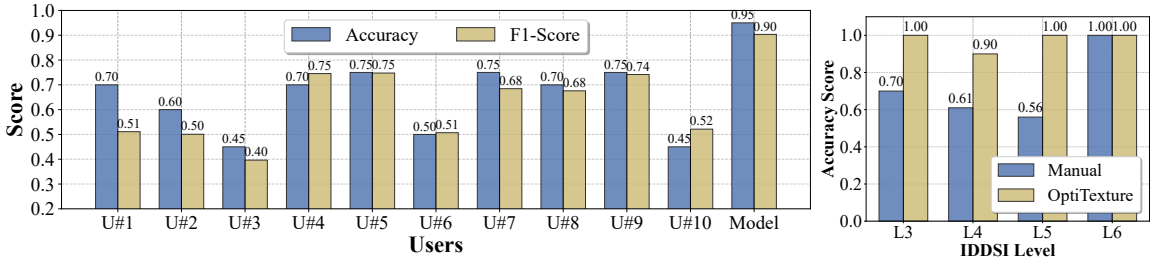


Fig. 12. OptiTexture's performance against subjective human labels. (N = 20 Samples)

Table 4. OptiTexture's performance compared with ablation models and the best baseline.

Model	Accuracy	F1-Score
RGB+TDLS	72%*	71%
OptiTexture (w/o MSPE)	83%*	82%
OptiTexture (w/o PIAM)	81%*	79%
OptiTexture (ours)	**92%**	**92%**

* indicates $p < 0.01$.

of manual IDDSI texture classification, even among trained individuals. In comparison, OptiTexture achieves an accuracy of 95% and an F1-score of 0.90 on the same set of samples, demonstrating stable and consistent performance across all test cases. However, when statistically comparing OptiTexture with the aggregated human annotations using McNemar's test, the difference does not reach statistical significance (p = 0.2187), likely due to the limited sample size (N = 20) and inter-rater variability among human annotators. To further contextualize this variability, we analyze inter-rater agreement among human annotators using Cohen's Kappa. The average human–human Kappa is 0.46, indicating moderate agreement and confirming that IDDSI texture classification is prone to subjective interpretation. In comparison, the agreement between OptiTexture and the ground-truth labels yields a Cohen's Kappa of 0.93, reflecting a strong and consistent alignment with the objective standard.

Finally, we examine classification accuracy across different IDDSI levels. As shown in Fig. 12, human annotators achieve near-perfect accuracy on Level 6 (Soft & Bite-Sized) foods due to their distinctive shape features. However, their performance decreases markedly for the more challenging and texturally ambiguous Levels 3 to 5, with average accuracies of 70%, 61%, and 56%, respectively. In contrast, OptiTexture maintains consistently high accuracy across all IDDSI levels, particularly for Levels 3 to 5, highlighting its potential to provide objective and reliable decision support in scenarios where human judgments are more variable.

5.3 Ablation Study

We then investigate the modules of OptiTexture by conducting the ablation study to demonstrate the effectiveness of the system design.

Effectiveness of Multi-Scale Physics Encoder (MSPE). The results of our ablation study indicate a significant 9% drop in accuracy and a 10% degradation in F1-score when the Multi-Scale Physics Encoder is disabled (as shown in Table 4). The improvement in accuracy score is statistically significant, as confirmed by a non-parametric bootstrap test with 10,000 resamples (p < 0.01). This decline clearly demonstrates the critical role that MSPE plays in extracting physically meaningful features from time-domain spectral data. The multi-scale convolutional

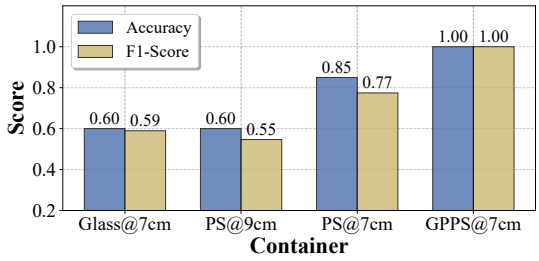


Fig. 13. Performance of OptiTexture for different containers.

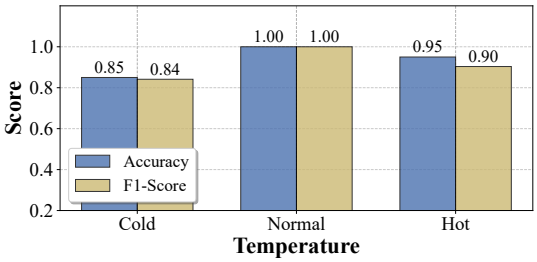


Fig. 14. Performance of OptiTexture for different food temperatures.

kernels within the MSPE effectively decode texture-relevant waveform patterns across various temporal scales. Furthermore, the physics-aligned contrastive learning approach employed by MSPE significantly improves clustering consistency for samples that are texturally similar but chemically diverse. This ability to maintain coherent groupings among varied samples is essential for accurate classification in complex datasets, highlighting the importance of integrating physical principles into the feature extraction process.

Effectiveness of Physics-Informed Attention Mechanism (PIAM). The removal of the Physics-Informed Attention Mechanism results in even more pronounced performance degradation than the removal of MSPE. Specifically, the accuracy falls by 11%, while the F1-score declines by 13%. This improvement is statistically significant, as confirmed by a non-parametric bootstrap test with 10,000 resamples ($p < 0.01$). These findings underscore the pivotal function of PIAM in facilitating adaptive feature fusion. By dynamically reweighting the contributions of different modalities based on texture parameters, such as cohesion and adhesion, PIAM utilizes these features as conditional priors to enhance the overall classification performance. This adaptive mechanism allows the system to prioritize the most relevant features from each modality, thus improving its ability to accurately assess and classify food textures.

5.4 Differing Conditions

We then evaluate OptiTexture under various experimental and environmental conditions to show its robustness. We train the model using the dataset collected under default settings in §5.1.1. Then we collect data of 20 samples, including 10 commercial food samples and 10 brewing food samples, under various setups to evaluate OptiTexture's performance across these different scenarios.

5.4.1 Varying Containers. Container material and size may influence light absorption and scattering properties in time-domain spectroscopy measurements. To evaluate these effects, we test our system with four container configurations: 7 cm diameter glass, 9 cm diameter polystyrene (PS), 7 cm diameter polystyrene, and 7 cm diameter general-purpose polystyrene (GPPS). The GPPS container, which is used for our baseline dataset collection, represents our reference configuration. As shown in Fig. 13, OptiTexture achieves perfect performance in the GPPS container with both accuracy and F1-score reaching 1.00. The 7 cm polystyrene container yields good results with 0.85 accuracy and 0.77 F1-score, while both the glass container and 9 cm polystyrene container show relatively lower performance, with accuracy around 0.60 and F1-scores between 0.55-0.59. We attribute performance variations mainly to the containers' optical properties and geometry. Glass containers have high surface reflectivity, leading to significant specular reflections that lower the signal-to-noise ratio. The larger 9 cm container increases optical path length, potentially causing greater signal attenuation than the standard 7 cm containers. To address container-dependent performance variations, we explore model fine-tuning using 25% of our full robust set, which successfully boost performance above 90% for new container types. This is because

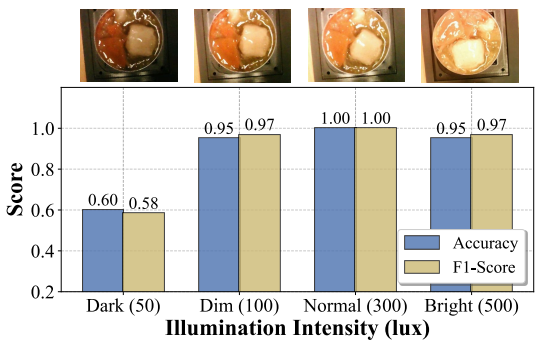


Fig. 15. Performance of OptiTexture for different illumination intensities.

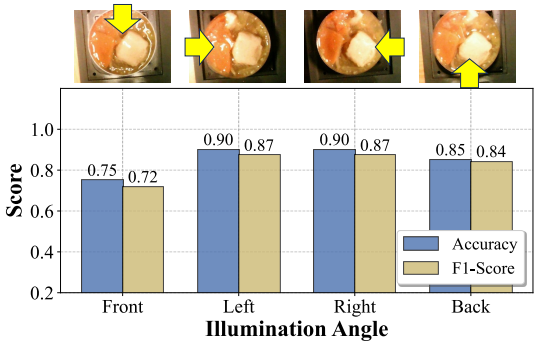


Fig. 16. Performance of OptiTexture for different illumination directions.

that container effects represent a systematic bias that can be readily compensated, unlike the more complex scattering variations caused by food texture differences themselves. This enables practical deployment with minimal adaptation effort.

5.4.2 Varying Sample Temperatures. Temperature is a critical factor that may influence the spectral and textural properties of food samples. In real usage scenarios, foods may be served at various temperatures, such as heated preparations, which could affect the measurement consistency of our system. To evaluate the robustness of OptiTexture under different thermal conditions, we test 20 food samples under three temperature regimes: Cold (samples refrigerated for over 12 hours), Normal (room temperature), and Hot (10 brewing samples prepared with water above $75°C$). As illustrated in Fig. 14, OptiTexture achieves perfect accuracy and F1-score under Normal temperature conditions. Performance remain strong under Hot conditions, with 95% accuracy and 0.90 F1-score, while Cold samples show a slight decrease to 85% accuracy and 0.84 F1-score. These results indicate that our system maintains reliable performance across a wide temperature range. The minor performance variation under non-ideal temperatures may be attributed to changes in the sample's surface texture or moisture distribution, which can influence optical properties. Nevertheless, the stable performance demonstrates the robustness of our approach, making it suitable for real-world applications where food temperature may vary.

5.4.3 Varying Ambient Illumination. Ambient lighting conditions significantly influence optical measurement systems through two primary interference mechanisms: RGB camera readings are directly affected, particularly in environments reliant on natural or artificial illumination, while intense ambient light may also interfere with the spectral module's measurements despite its built-in lighting system.

We evaluate system performance under controlled variations in illumination intensity and direction. **(1) Illumination Intensities.** The experiment is conducted under four illumination levels: dim (50 lux), low light (100 lux), normal light (300 lux), and bright light (500 lux). As shown in Fig. 15, under dim light conditions of 50 lux, system performance significantly degrade, with accuracy and F1 score reaching 60% and 0.58, respectively. This result is expected, as the signal-to-noise ratio of images acquired by the RGB camera is significantly reduced under low illumination. When the light intensity increases above 100 lux, the system performs stably, with both accuracy and F1 score remaining above 95%. At 300 lux, it even achieves perfect performance. This result demonstrates that our system maintains excellent performance under typical indoor lighting conditions. **(2) Illumination Directions.** Experiments test four typical lighting directions: front, left, right, and back. As shown in Fig. 16, the side lighting (left and right) performs best, achieving both accuracy and F1 score of 90%. Back lighting performs second best with accuracy of 85% and F1 score of 0.84, while front lighting performs relatively

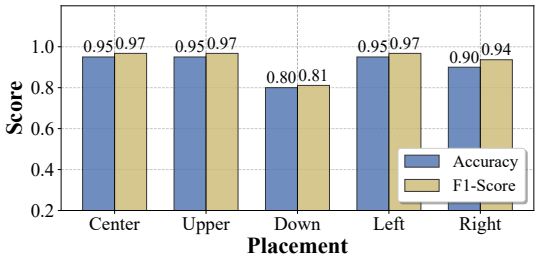


Fig. 17. Performance of OptiTexture for different food placement positions.

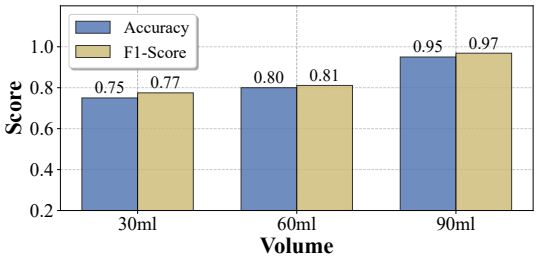


Fig. 18. Performance of OptiTexture for different food volumes.

poorly with accuracy of 75% and F1 score of 0.72. Our analysis suggests that front lighting tends to produce strong specular reflections on the food surface. These highlights obscure the food's texture features, thus affecting classification accuracy. In contrast, side lighting produces more uniform diffuse reflections, which facilitates feature extraction.

Notably, the RGB and spectral modules demonstrate excellent complementarity under varying lighting conditions. When ambient light is insufficient, the spectral module's self-luminous source provides stable illumination. In bright light environments, the spectral module is less susceptible to interference from ambient light, providing reliable supplementary data. This dual-module design effectively enhances the system's adaptability to diverse lighting environments, providing strong support for its application in real-world scenarios.

5.4.4 Varying Sample Placements. In real-world scenarios, users may not always be able to precisely place the sample in the center of the detection area. To evaluate the impact of sample position on measurement results, we test system performance at five different positions, including center, top, bottom, left, and right. As illustrated in Fig. 17, the experimental results show that the system maintains good performance stability in most positions. When the sample is placed in the center, top, and left positions, the system achieves an accuracy of 95% and an F1 score of 0.97. Performance degrades slightly at the right position with 90% accuracy and 0.94 F1 score, while performance is relatively low at the bottom position with only 80% accuracy and 0.81 F1 score. These data demonstrate that despite some deviation in sample placement, the system maintains good recognition performance in the five cardinal positions, *i.e.*, center top, left, and right. The performance degradation at the bottom position may be related to uneven sample distribution or varying lighting conditions. Overall, the system demonstrates strong tolerance to sample position variations, making it suitable for deployment in real-world scenarios.

5.4.5 Varying Sample Volumes. In actual application scenarios, user-prepared sample volumes may vary. To evaluate the impact of volume variations on measurement results, we test the system performance at three volume levels: 30 ml, 60 ml and 90 ml. The experimental results show a clear correlation between volume and performance (in Fig. 18). The system performs best at 90 ml, achieving an accuracy of 95% and an F1 score of 0.97. Performance remain strong at 60 ml, with an accuracy of 80% and an F1 score of 0.81. Performance declines slightly at 30 ml, achieving an accuracy of 75% and an F1 score of 0.77. This performance difference primarily stems from the optical measurement principle. That is when the sample volume is too small, the insufficient sample height within the container may cause some near-infrared light to be directly transmitted rather than reflected, reducing effective signal intensity. Furthermore, the small sample surface area makes it difficult to fully cover the detection area, resulting in inadequate texture feature extraction. It is important to note that 30 ml is significantly below the recommended single meal volume for dysphagia patients (typically 100-200 ml) and

is rarely used in practice. Therefore, it is recommended that users prepare a sample volume of at least 60ml to ensure the reliability of the system in practical applications.

6 RELATED WORK

In this section, we briefly review existing works related to food texture detection schemes, low-cost food sensing systems, and optical technologies in food assessment.

6.1 Food Texture Detection Schemes

Standard food texture assessment methods can be divided into two main categories: instrumental analysis and subjective clinical assessments [13]. Instrumental methods, primarily texture profile analysis (TPA) using universal testing machines, are considered the gold standard for quantifying mechanical properties such as hardness, cohesiveness, and adhesiveness [10, 20]. These methods simulate oral processing by applying controlled compression and tension to food samples, providing reproducible and quantitative data [41]. For instance, Hadde et al. established strong correlations between TPA parameters and the IDDSI levels, demonstrating its efficacy for dysphagia diet standardization [10]. Despite their accuracy, these methods are inherently destructive, require standardized sample preparation, and are confined to laboratory settings, making them unsuitable for routine meal preparation or point-of-care testing. Subjective methods, such as the International Dysphagia Diet Standardisation Initiative (IDDSI) fork test [15], evaluate texture grades based on visual observations of food. However, human judgment can lead to viscosity estimation errors in 30-40% of pasty foods, increasing the risk of aspiration [8, 24, 34]. In contrast to these methods, our work seeks to achieve the accuracy of instrumental analysis for non-professionals with low-cost form factor, directly addressing the practicality limitations of existing gold-standard techniques.

6.2 Low-cost Food Sensing Systems

To address the cost barrier, computer vision has emerged as a promising low-cost solution. Pfisterer et al. attempted to classify pureed foods using smartphone images and convolutional neural networks, achieving moderate success but struggling with texturally similar samples due to limited spectral information [28]. Other studies have explored using smartphone sensors for viscosity estimation in liquids. For example, An et al. developed Viscocam, which measures liquid flow on a smartphone screen to estimate viscosity [1]. Huang et al.'s use the phone's accelerometer to analyze surface waves induced by tapping the container [12]. While innovative, these methods are difficult to generalize to solid or semi-solid foods where complex mechanical properties beyond viscosity are critical. In summary, existing methods present a trade-off between accuracy, practicality, and cost. There is currently no low-cost, non-destructive solution capable of directly quantifying the key mechanical texture parameters (hardness, cohesiveness, adhesiveness) necessary for reliable IDDSI classification of solid and semi-solid foods. Our work fills this gap by introducing a novel system that synergistically combines low-cost time-domain scattering and computer vision to translate optical properties into actionable texture grades. Our system advances beyond these low-cost sensing approaches by integrating optical scattering spectroscopy, enabling it to probe sub-surface microstructural properties critical for assessing the mechanical texture (hardness, cohesiveness) of solid and semi-solid foods, not just viscosity or surface appearance.

6.3 Optical Techniques for Food Assessment

Optical techniques offer non-contact alternatives for food quality evaluation. Hyperspectral and Multispectral Imaging methods combine spatial and spectral information to map component distribution. For example, Wang et al. utilized hyperspectral imaging to predict the moisture and fat content in meat, which indirectly relates to texture [40]. However, the high dimensionality of the data requires complex processing and the equipment

cost remains prohibitive for widespread deployment [43]. In addition, Terahertz Time-Domain Spectroscopy (THz-TDS) probes molecular vibrations and is sensitive to water content. Researchers like Sun et al's team have used it to monitor moisture loss in beef during drying, correlating it with texture hardening [7, 32]. A key limitation, besides high cost, is the strong absorption of THz waves by water, which can necessitate controlled environments and limits its application to high-moisture foods [6, 9]. Furthermore, Near-Infrared Time-Domain Spectroscopy (NIR-TDS), using pulsed lasers and time-correlated single-photon counting (TCSPC), can non-invasively probe scattering coefficients related to microstructure [30]. It has been successfully applied to quantify composition in homogeneous materials like powders and grains [45]. For instance, Tsuta et al. used time-resolved spectroscopy to assess the sugar content in fruits by analyzing light propagation [23]. However, its application has overwhelmingly focused on chemical composition rather than physical texture. The high cost and complexity of laser-based TCSPC systems have prevented the exploration of NIR-TDS for direct, mechanical texture assessment of complex, semi-solid foods like purees [30]. However, our work bridges this gap by proposing a low-cost alternative that utilizes high-frequency pulsed LEDs and a photodiode to obtain temporal profiles of reflected light. Although this approach provides a lower temporal resolution than high-end systems, we demonstrate that the captured dynamics contain rich information about the food's physical microstructure which can support roughly food texture classification for the IDDSI levels.

7 DISCUSSION

In this section, we shall discuss the limitations and future extensions of OptiTexture.

Dataset Scale and Diversity. This study verifies the system's performance on 112 food samples, covering the main food types of the IDDSI L3-L6 levels. While our system demonstrates promising performance across various experimental conditions, we acknowledge limitations in dataset scale and diversity. Especially for foods with special components or non-traditional preparation methods, the system's performance still needs further verification. To address this issue, we consider including food samples from different regions and cultural backgrounds in the future. At the same time, we can develop data augmentation algorithms based on physical parameters to simulate the effects of different raw material ratios and processing methods on the optical properties of food, thereby enhancing the model's generalization ability for unknown foods.

Complex Environmental Factors The experiments show that the system exhibits good robustness to changes in a single environmental factor (such as container type, temperature, etc.). However, in real scenarios, multiple environmental factors may coexist and produce a synergistic effect. This combined influence still requires in-depth study. For example, specific containers in a low-temperature environment may amplify optical measurement errors. To address this challenge, we suggest adding a domain-adaptive method based on transfer learning in the next-generation system. This method can effectively adapt the models developed in the laboratory environment to new usage scenarios without the need for a large amount of re-labeled data.

Practical Clinical Deployment. The proposed system is designed as a rapid, non-contact tool for relative texture screening rather than a replacement for standardized mechanical measurements. Its simple workflow enables use in practical care settings such as hospital wards or long-term care facilities, where conventional rheological testing is impractical. Nevertheless, a key limitation of the current system is its reduced recall (83.33%) for Level 6 (Soft & Bite-Sized) foods. This level typically requires higher swallowing capability and is excluded from diets for severe dysphagia. Misclassifying a Level 6 food as a softer level could therefore lead to inappropriate, potentially risky dietary recommendations. Error analysis indicates that most misclassifications occur at the boundary with Level 5, primarily due to ambiguity in distinguishing food piece size, which is a feature that human assessors can reliably judge. Therefore, to mitigate this risk in practical use, the system's predictions should be paired with confidence estimates, with low-confidence or borderline (e.g., Level 5/6) cases flagged for secondary human review. Looking forward, improving sensitivity for Level 6 is a priority, and we plan to incorporate active

learning to identify such challenging cases for expert annotation, thereby enhancing robustness in real-world scenarios.

Extended Applications. The integration of near-infrared spectroscopy and computer vision provides a new technical approach for food texture analysis. Besides the basic IDDSI grade classification, this system architecture can be extended to derivative applications such as nutritional component analysis and spoilage detection. Future work will explore a multi-task learning framework, based on shared feature extractors, to simultaneously complete tasks such as texture grading and nutritional assessment. Additionally, through integration with electronic health record systems, this technology is expected to provide personalized dietary recommendations and long-term nutritional monitoring for patients with swallowing disorders.

8 CONCLUSION

This work addresses the critical need for objective food texture assessment in dysphagia care by introducing a novel paradigm that decodes mechanical properties from optical signatures. We develop a low-cost system that fuses time-domain light scattering with RGB imaging to enable non-contact probing of both subsurface microstructure and surface texture. To overcome the key challenges of this approach, we design a multi-scale convolutional feature extractor that mirrors the physics of light scattering to distill clear texture signals from noisy hardware data, complemented by a physical information attention mechanism that dynamically aligns subsurface and surface modalities to resolve conflicts. Evaluated on 112 clinically annotated samples, our sub-$20 system achieves 91.96% accuracy in IDDSI level classification, outperforming a vision-only baseline by 22.32%. This study establishes a practical and accessible tool for point-of-care and home use, effectively mitigating aspiration risk for individuals on texture-modified diets while also offering a promising methodology for non-contact material characterization beyond food science.

Acknowledgments

We are grateful to all the anonymous reviewers for their insightful comments. This research is supported in part by RGC under Contract CERG 16206122, 16204523, 16205824, AoE/E-601/22-R, 2023B03J1234, HB016, RG702, and Contract R8015.

References

[1] Kecheng An, Qian Zhang, and Elaine Kwong. 2021. Viscocam: Smartphone-based drink viscosity control assistant for dysphagia patients. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–25.

[2] Sylvie Bureau. 2009. The use of non-destructive methods to analyse fruit quality. *Fresh produce* 3, 1 (2009), 23–34.

[3] Silvia Carrión, Alicia Costa, Omar Ortega, Eric Verin, Pere Clavé, and Alessandro Laviano. 2018. Complications of oropharyngeal dysphagia: malnutrition and aspiration pneumonia. In *Dysphagia: Diagnosis and treatment*. Springer, 823–857.

[4] Julie AY Cichero, Catriona Steele, Janice Duivestein, Pere Clavé, Jianshe Chen, Jun Kayashita, Roberto Dantas, Caroline Lecko, Renee Speyer, Peter Lam, et al. 2013. The need for international terminology and definitions for texture-modified foods and thickened liquids used in dysphagia management: foundations of a global initiative. *Current physical medicine and rehabilitation reports* 1, 4 (2013), 280–291.

[5] DG Dalgleish and FR Hallett. 1995. Dynamic light scattering: applications to food systems. *Food Research International* 28, 3 (1995), 181–193.

[6] Chao-Hui Feng and Chiko Otani. 2021. Terahertz spectroscopy technology as an innovative technique for food: Current state-of-the-Art research advances. *Critical reviews in food science and nutrition* 61, 15 (2021), 2523–2543.

[7] Ying Fu, Yuqiao Ren, and Da-Wen Sun. 2023. Investigating the dehydration feature of beef products with different pre-treatment during microwave vacuum dehydration using NIR-HSI and terahertz spectral analyses. *Biosystems and Food Engineering Research Review* 28 (2023), 23.

[8] R Goulding and A MO Bakheit. 2000. Ealuation of the benefits of monitoring fluid thickness in the dietary management of dysphagic stroke patients. *Clinical rehabilitation* 14, 2 (2000), 119–124.

[9] Aoife A Gowen, Créidhe O'Sullivan, and Colm P O'Donnell. 2012. Terahertz time domain spectroscopy and imaging: Emerging techniques for food process monitoring and quality control. *Trends in Food Science & Technology* 25, 1 (2012), 40–46.

[10] Enrico Karsten Hadde, Sangeeta Prakash, Wei Chen, and Jianshe Chen. 2022. Instrumental texture assessment of IDDSI texture levels for dysphagia management. Part 2: Texture modified foods. *Journal of Texture Studies* 53, 5 (2022), 617–628.
[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
[12] Yongzhi Huang, Kaixin Chen, Yandao Huang, Lu Wang, and Kaishun Wu. 2021. Vi-liquid: unknown liquid identification with your smartphone vibration. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 174–187.
[13] Francisco C Ibanez, Gorka Merino, María Remedios Marín-Arroyo, and María José Beriain. 2022. Instrumental and sensory techniques to characterize the texture of foods suitable for dysphagic people: A systematic review. *Comprehensive Reviews in Food Science and Food Safety* 21, 3 (2022), 2738–2771.
[14] IDDSI. 2025. History. https://www.iddsi.org/about-us/history
[15] IDDSI. 2025. Testing Methods. https://www.iddsi.org/standards/testing-methods
[16] Patrick Jackman and Da-Wen Sun. 2013. Recent advances in image processing using image texture features for food quality assessment. *Trends in food science & technology* 29, 1 (2013), 35–43.
[17] Steven L Jacques. 2013. Optical properties of biological tissues: a review. *Physics in Medicine & Biology* 58, 11 (2013), R37.
[18] Elijah Robert Jensen. 2018. High dynamic range optical devices and applications. (2018).
[19] Weijie Lan, Vincent Baeten, Benoit Jaillais, Catherine MGC Renard, Quentin Arnould, Songchao Chen, Alexandre Leca, and Sylvie Bureau. 2022. Comparison of near-infrared, mid-infrared, Raman spectroscopy and near-infrared hyperspectral imaging to determine chemical, structural and rheological properties of apple purees. *Journal of Food Engineering* 323 (2022), 111002.
[20] Seo Yeon Lee, Seo-Jin Chung, Jeong-Won Lim, Mi Young Kim, and Chung Hong Ha. 2025. Texture characterization of semi-solid foods for older adults: Sensory analysis and instrumental approaches. *LWT* 221 (2025), 117582.
[21] Julia Leira, Ana Maseda, Laura Lorenzo-López, Nuria Cibeira, Rocío López-López, Leire Lodeiro, and José C Millán-Calenti. 2023. Dysphagia and its association with other health-related risk factors in institutionalized older people: A systematic review. *Archives of gerontology and geriatrics* 110 (2023), 104991.
[22] Renfu Lu. 2017. *Light scattering technology for food property, quality and safety assessment*. Crc Press.
[23] Te Ma, Tetsuya Inagaki, and Satoru Tsuchikawa. 2023. Validation study on light scattering changes in kiwifruit during postharvest storage using time-resolved transmittance spectroscopy. *Scientific Reports* 13, 1 (2023), 16556.
[24] Sonja M Molfenter and Catriona M Steele. 2013. The relationship between residue and aspiration on the subsequent swallow: an application of the normalized residue ratio scale. *Dysphagia* 28, 4 (2013), 494–500.
[25] Amparo Moret-Tatay, Julia Rodríguez-García, Ezequiel Martí-Bonmatí, Isabel Hernando, and María Jesús Hernández. 2015. Commercial thickeners used by patients with dysphagia: Rheological and structural behaviour in different food matrices. *Food Hydrocolloids* 51 (2015), 318–326.
[26] Lucian Daniel Olaru, Oana-Viorela Nistor, Doina Georgeta Andronoiu, Ioana Otilia Ghinea, Viorica Barbu, and Elisabeta Botez. 2022. Effect of added hydrocolloids on ready-to-eat courgette (Cucurbita pepo) puree ohmically treated. *Journal of Food Science and Technology* 59, 1 (2022), 388–396.
[27] Subhash Pawde and Jaydeep Dave. 2025. Applications of soft matter physics in food science: from molecular interactions to macro-scale food structures. *Sustainable Food Technology* 3, 4 (2025), 979–1004.
[28] Kaylen J Pfisterer, Robert Amelard, Audrey G Chung, and Alexander Wong. 2018. A new take on measuring relative nutritional density: The feasibility of using a deep neural network to assess commercially-prepared puréed food concentrations. *Journal of Food Engineering* 223 (2018), 220–235.
[29] Kaylen J Pfisterer, Robert Amelard, Braeden Syrnyk, and Alexander Wong. 2019. Towards computer vision powered color-nutrient assessment of puréed food. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 0–0.
[30] Antonio Pifferi, Davide Contini, Alberto Dalla Mora, Andrea Farina, Lorenzo Spinelli, and Alessandro Torricelli. 2016. New frontiers in time-domain diffuse optics, a review. *Journal of biomedical optics* 21, 9 (2016), 091310–091310.
[31] Dele Raheem, Conrado Carrascosa, Fernando Ramos, Ariana Saraiva, and António Raposo. 2021. Texture-modified food for dysphagic patients: A comprehensive review. *International Journal of Environmental Research and Public Health* 18, 10 (2021), 5125.
[32] Yuqiao Ren, Ying Fu, and Da-Wen Sun. 2023. Analyzing the effects of nonthermal pretreatments on the quality of microwave vacuum dehydrated beef using terahertz time-domain spectroscopy and near-infrared hyperspectral imaging. *Food Chemistry* 428 (2023), 136753.
[33] G Romano, M Nagle, D Argyropoulos, and J Müller. 2011. Laser light backscattering to monitor moisture content, soluble solid content and hardness of apple tissue during drying. *Journal of Food Engineering* 104, 4 (2011), 657–662.
[34] David W Rule, Lisa Kelchner, Ashley Mulkern, Sarah Couch, Noah Silbert, and Kathy Welden. 2020. Implementation strategies for the International Dysphagia Diet Standardisation Initiative (IDDSI), part I: quantitative analysis of IDDSI performance among varied participants. *American Journal of Speech-Language Pathology* 29, 3 (2020), 1514–1528.
[35] Wouter Saeys, Maria A Velazco-Roa, Suresh N Thennadil, Herman Ramon, and Bart M Nicolaï. 2008. Optical properties of apple skin and flesh in the wavelength range from 350 to 2200 nm. *Applied optics* 47, 7 (2008), 908–919.

[36] MJ Sáiz-Abajo, B-H Mevik, VH Segtnan, and TJACA Næs. 2005. Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data. *Analytica chimica acta* 533, 2 (2005), 147–159.
[37] Tanmay Sarkar, Molla Salauddin, Kohima Kirtonia, Siddhartha Pati, Maksim Rebezov, Mars Khayrullin, Svetlana Panasenko, Lyudmila Tretyak, Marina Temerbayeva, Nadezhda Kapustina, et al. 2022. A review on the commonly used methods for analysis of physical properties of food materials. *Applied Sciences* 12, 4 (2022), 2004.
[38] Wei Sun, Ishaan Chansarkar, and Kannan Srinivasan. 2023. Short: Liquid thickness sensing with backscattered signals for dysphagia. *Smart Health* 28 (2023), 100399.
[39] Valery V Tuchin. 2016. Polarized light interaction with tissues. *Journal of biomedical optics* 21, 7 (2016), 071114–071114.
[40] Caixia Wang, Songlei Wang, Xiaoguang He, Longguo Wu, Yalei Li, and Jianhong Guo. 2020. Combination of spectra and texture data of hyperspectral imaging for prediction and visualization of palmitic acid and oleic acid contents in lamb meat. *Meat Science* 169 (2020), 108194.
[41] Ashley K Young, Jean Ne Cheong, Kylie D Foster, Duncan I Hedderley, Marco P Morgenstern, and Bryony J James. 2016. Exploring the links between texture perception and bolus properties throughout oral processing. Part 1: Breakdown paths. *Journal of Texture Studies* 47, 6 (2016), 461–473.
[42] Chunjuan Zhang, Dequan Zhang, Yuanyuan Su, Xiaochun Zheng, Shaobo Li, and Li Chen. 2022. Research on the authenticity of mutton based on machine vision technology. *Foods* 11, 22 (2022), 3732.
[43] Mingyang Zhang, Maoguo Gong, and Yongqiang Chan. 2018. Hyperspectral band selection based on multi-objective optimization with high information and low redundancy. *Applied Soft Computing* 70 (2018), 604–621.
[44] Mengqing Zhang, Chao Li, Fang Zhang, Xiaoxiao Han, Qinglu Yang, Tuo Lin, Huichang Zhou, Min Tang, Jungui Zhou, Hongling Shi, et al. 2021. Prevalence of dysphagia in China: an epidemiological survey of 5943 participants. *Dysphagia* 36, 3 (2021), 339–350.
[45] Qibing Zhu, Yongchun Xing, Renfu Lu, Min Huang, and Perry KW Ng. 2017. Visible/shortwave near infrared spectroscopy and hyperspectral scattering for determining bulk density and particle size of wheat flour. *Journal of Near Infrared Spectroscopy* 25, 2 (2017), 116–126.