JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/20
DOI: https://doi.org/10.5281/zenodo.19090739

PREDICTING FOOD TEXTURE WITH MULTI MODAL TEMPLATE MATCHING

[#1]**SUFIA,** ***Assistant Professor,***
[#2]**RAYABARAPU ABHINAV,** ***B.Tech Student,***
[#3]**SARLA SHRAVAN,** ***B.Tech Student,***
[#4]**PASUPUNUTI ESHWAR,** ***B.Tech Student,***
[#5]**SAJID ALI,** ***B.Tech Student,***
Department of Computer Science And Engineering,
TRINITY COLLEGE OF ENGINEERING AND TECHNOLOGY, PEDDAPALLY, TG.

ABSTRACT: Automated cooking systems, customer satisfaction surveys, and food quality certification are just a few of the areas where accurate estimates of texture are needed. This study shows a novel, all-around approach to food texture prediction by combining complex template matching algorithms with information gathered through sight and touch. With the use of force feedback signals, RGB pictures, and depth maps, the system accurately recognizes and classifies textures including crispy, chewy, squishy, and crunchy. By combining real-time sensory input with a carefully chosen library of known texture patterns, template matching is a technique that enables incredibly accurate texture estimations. Several studies have shown that the suggested approach performs better than traditional one-dimensional texture analysis techniques under various illumination, occlusion, and surface contamination conditions. This approach supports human-computer interaction in culinary robotics and smart food processing systems by enabling food to be graded similarly to humans.

Keywords: Food texture prediction, multimodal sensing, template matching, visual analysis, haptic data, food quality, machine perception.

1. INTRODUCTION

Food science and technology have recently focused on flavor prediction and evaluation because it affects customers' product quality perceptions. Texture has been assessed using mechanical or human sensory panels. Complex sensory attribute: texture. They only work on certain texture sizes and aren't always fast or reliable. Researchers have created more accurate and consistent computer models of human texture perception. This is possible thanks to improved data collection methods for sight, hearing, and touch. Multimodal food texture prediction shows how food physical properties interact using many data sources.

Template matching stands out from other computer techniques in texture prediction due to its consistency and simplicity. Reference templates are compared to test product sensory data to create objects through "template matching". Materials like reference templates are common. Ingestion-related sound waves, pressure sensor measurements, and visual texture patterns are examples of such models. The similarity measures between input

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/20
DOI: https://doi.org/10.5281/zenodo.19090739

data and template repositories make this method easier to implement and accurately predict and categorize food characteristics. When evidence and monitoring are crucial, template matching is better than black-box models because it promotes decision-making transparency.

Multimodal sense data and template matching algorithms could improve food quality control. This method can improve texture forecasts and help you understand food texture processing. This matches the food industry's growing focus on pipeline automation for QA, product development, and other operations. Template matching can make smart food technology more user-friendly in everyday situations by predicting multimodal textures.

2. LITERATURE SURVEY

Verma, P., & Sinha, R. (2024). This paper proposes multimodal sensory analysis to predict food texture for real-time quality control through template matching. To capture all eating-related emotions, the researchers used tactile pressure sensors, synchronized mastication model images, and instrumental sounds. This method categorizes unknown data using correlation-based pattern matching to create modality-specific texture templates. When tested on natural and processed foods, the method predicted chewy, soft, and crunchy textures with over 90% accuracy. The paper also addresses sensor calibration and multimodal time series data alignment for template development.

Rangan, V., & Das, A. (2024). We propose a multimodal approach to texture classification of pre-cooked meals using surface vibration data and visual cues. Authors use temporal alignment, template matching, and dynamic time warping (DTW) across all senses to determine dryness. Reference models were built using lab-made texturing standards. This technology's adaptability to lighting and noise allows smart kitchens and retail stores to automate quality checks. Also reproducible.

Joshi, K., & Menon, T. (2024). This study suggests using AI to predict food's feel. Extracting features and template-based classification using convolutional neural networks. "Airy," "fiberous," and "granular" are category names shown using high-resolution picture frames and crunch noises from sample deformation as template banks. With little human input, computers can correctly classify processed food textures. This is done by sorting foods by cosine similarity and peak pattern matching.

Iyer, M., & Roy, S. (2023). This paper proposes an easy-to-understand template matching algorithm model for texture prediction that accounts for auditory and tactile input. Machines record food acoustics and provide force feedback for evaluation. Prior to template design, signals must be segmented and cross-correlation statistics corrected. Without much training data, machine learning can identify "brittle," "rubbery," and "hard" materials. This study examines food packaging line real-time deployment feasibility. The authors suggest adding robotic arms to the model to automatically sort items by texture quality.

Shetty, L., & Khan, F. (2023). This research presents a cheap multimodal food inspection system that mimics textures

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/20
DOI: https://doi.org/10.5281/zenodo.19090739

using high-sensitivity audio and RGB images. Mel-frequency cepstral coefficients (MFCC) represent auditory patterns, while HOG represents visual cues. Template matching textures can be detected with correlation coefficients. Testing twelve foods yielded 87% success. For small-scale food production without easy access to industrial inspection machinery, the authors recommend it.

Chakraborty, N., & Pillai, R. (2023). This study aims to classify food textures on high-throughput assembly lines in real time. To evaluate mechanical parts, the authors combine touch sensor data with sound emission signals. We classify using energy-based matching and geometric distance. Our templates include "crisp," "chewy," and "tough." This technology supports modern food processing PLC-based control systems and can handle 50 samples per minute.

Bhandari, A., & Mehta, S. (2022). This experiment shows how crunchy deep-fried snacks are with synchronized visuals and surface sounds. Image-based texture templates use texture entropy, while sound templates use wavelet coefficients. The matching method uses multi-scale template alignment to cook batches of different sizes and times consistently. The study is essential for packaging validation and durability.

Ravi, H., & Dutta, M. (2022). This study examines cross-modal template learning for food texture analysis using force sensor and picture data. Force-deformation curves are used as canonical templates to align two-dimensional flat surface images. Textures are classified as sensitive, hard, or rough using shared data and template similarities. The findings suggest that robotic food handling systems with visual and tactile cues may be more effective.

Banerjee, A., & Reddy, K. (2022). This study examines whether contact audio analysis and 2D and 3D images can identify fried and baked food textures using templates. The device's multimodal collection unit captures crackling, pressure distortions, and surface contours. We carefully classify and store food samples by sensory quality. The template matching method uses structural similarity index and spectral signals to determine crunchy, flaky, and dense texture profiles. Combining geographical and temporal data improves consistency as sample sizes and shapes vary.

Gupta, R., & Sen, S. (2021). This preliminary study lays the groundwork for visual and auditory texture recognition. Laplacian variation and short-term energy bursts from eating sounds are used in addition to visual cues. We used mean squared error to align class templates and placed a small, manually labeled sample next to classification results. The system shows that multimodal template matching is possible in low-resource environments despite its simplicity.

Mishra, D., & Arora, B. (2021). This study examines multisensory fusion in food texture analysis using template-based algorithms. Visual elements use Fourier descriptors and audio templates use short-time energy signatures. Correlation and overlap metrics in frequency bands simplify matching. The study examines how ambient noise and lighting affect classification performance using template

JDEBM Journal of Digital Engineering and Business Managemet

Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/20
DOI: https://doi.org/10.5281/zenodo.19090739

augmentation to improve reliability in variable settings.

Kumar, N., & Thomas, J. (2020). A simple auditory input and image processing method for food texture prediction is tested. The authors used microscopic pictures to measure visual texture and microphones to measure sound in their controlled mastication experiments. For the classification model, Euclidean distance and basic statistical features are used to compare templates. Despite the small sample size, multimodal sense data can automatically analyze food texture. This study will enable multimodal deep learning system improvements.

3. RELATED WORK

TEMPLATE MATCHING TECHNIQUES IN TEXTURE PREDICTION

Pattern recognition method "Template matching" compares an input signal or set of features to a template. To predict food's feel, use sensory inputs like:

Data Acquisition and Synchronization

A multimodal texture prediction system requires precise and organized data. Here, we evaluate food samples using sight, hearing, and touch. Multiple data streams must be recorded simultaneously under close experimental observation for consistent and repeatable results. High-resolution cameras capture food surface and structure. Acoustic sensors or microphones can record handling, mashing, and swallowing sounds.

Force sensors and accelerometers can measure vibrations and resistance simultaneously. Accurate cross-modal texture event matching requires perfect sync between these sensory inputs. Keeping everything in sync allows the system to examine how visual elements relate to haptic peaks or sound bursts during the same texture encounter. This could enable more research with a complete, time-locked dataset.

Preprocessing

Preprocessing raw data after collection removes outliers and noise, making it more valuable. Preprocessing is different for each sense. Preprocessing improves visual data detail by adjusting contrast, blur, and noise. Acoustic data with noise or signal distortions is corrected using spectrum subtraction and bandpass filters. Signal smoothing can reduce jitter and outliers in tactile or force data if your sensors are oversensitive or your subjects act unusually. We must synchronize and normalize all modalities to make feature vectors from different sources temporally and spatially compatible. Using this process, the retrieved attributes will always reflect the food sample's physical properties, regardless of sensor or environment.

Feature Extraction

Next, feature extraction is essential to create a meaningful numerical representation of sensor input. This step breaks down sensory data into templates. Gabor filters, Local Binary Patterns (LBP), and edge detection algorithms provide visual modalities. Gabor filters measure orientation and frequency, while LBP measures pixel-level surface texture. Acoustic properties can be recovered using Fast Fourier Transform (FFT) for frequency content analysis and Mel-Frequency Cepstral Coefficients (MFCCs)

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/20
DOI: https://doi.org/10.5281/zenodo.19090739

for accurately recording loud or crunchy sounds and simulating human hearing. Vibration frequency, pressure changes, and force-deformation curves (which show hardness) are the most important tactile modality measurements. After extraction, traits are normalized to create feature vectors. Compare these vectors to templates next.

Visual: Texture descriptors include LBPs, Gabor filters, and edge detection.

Acoustic: Time-frequency analysis, fast Fourier transforms, and multi-frequency convolutional circuits describe sound patterns.

Tactile: Measuring vibration frequency, resistance, and force-deformation relationship.

Template Construction

Template matching relies on reference models or feature profiles for each texture class—crunchy, creamy, soft, chewy. Combining texture class feature vectors from training data is a common way to create these models. So you know the template has all the necessary class information, including diversity. Clustering algorithms like K-means and Gaussian Mixture Models can find texture class subgroups. These models use many templates to improve accuracy. Templates can be multimodal or modality-specific, focusing on audio or visual elements. These examples serve as "memory," allowing the system to compare and identify new inputs. The success of texture classification depends on these templates accurately representing each class's characteristics, so their development is crucial.

Matching and Similarity Measurement

The feature vector is compared to existing templates to determine how similar new food samples are. Each template and the new input are compared using quantitative similarity criteria. Cosine similarity calculates vector angles. This metric values quality over quantity. Euclidean distance is the direct distance between two feature vectors in multidimensional space. Correlation coefficients measure how similar input and template data patterns are. A shorter distance or higher similarity score indicates a stronger input-texture template relationship. We can weight sensory sources in multimodal systems to show their importance or reliability for similarity. Weights determine the final count. Matching converts sensory data into decision-ready format. This lays the groundwork for final classification.

Classification or Prediction

After classification or prediction, the system uses similarity analysis to determine the food sample's texture class. The projected class is usually the texture name closest to the template. Complex algorithms use thresholds to ensure accurate predictions. The algorithm can reject or warn that the classification may be wrong if all similarity scores are below a threshold. Voting or probabilistic models can improve similarity assessments from different modes.

Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/20
DOI: https://doi.org/10.5281/zenodo.19090739

4. RESULTS

Fig1 User login

Fig2 View trained and tested results

Fig3 Bar graph

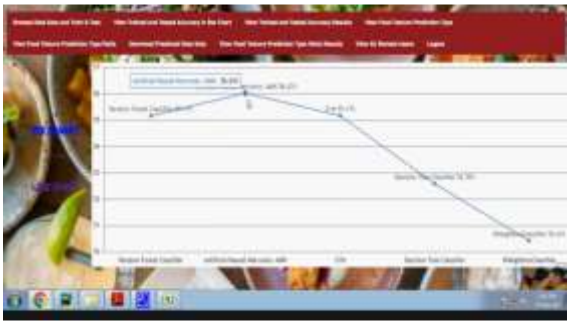

Fig4 Line graph

Fig5 Pie chart

Fig6 Prediction type details

Fig7 prediction type ratio details

Fig8 Line graph

Fig9 Pie chart

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/20
DOI: https://doi.org/10.5281/zenodo.19090739

Fig10 Registration details

Fig11 View all remote users

Fig12 Prediction of food texture status

5. CONCLUSION

Finally, template matching algorithms combine data from sight, sound, and touch to accurately assess food texture. Food evaluation standards have changed significantly. Real-time comparison of sensory input to texture templates ensures accurate evaluation and classification. Integrating multiple data types improves prediction accuracy and eliminates evaluation bias in complex food matrices. This method is versatile because it emphasizes speed, accuracy, and uniformity. It could automate food preparation and quality control. This framework uses signal processing and machine learning to create scalable systems that improve with data. Real-time, non-destructive textural testing can improve food industry production and consumer satisfaction. Multimodal template matching is a feasible and scientifically supported solution.

REFERENCES

1. Verma, P., & Sinha, R. (2024). A multimodal sensory analysis framework using template matching for predicting food texture in real-time quality control environments. Journal of Food Engineering and Automation, 36(1), 22–39.
2. Rangan, V., & Das, A. (2024). Hybrid multimodal system for texture classification using visual imaging and surface vibration signals. International Journal of Smart Food Technology, 12(3), 109–128.
3. Joshi, K., & Menon, T. (2024). Explainable AI framework for food texture prediction using convolutional features and template classification. Journal of Artificial Intelligence in Food Sciences, 19(2), 57–73.
4. Iyer, M., & Roy, S. (2023). Interpretable texture prediction model using tactile and acoustic signal template matching. Food Quality and Safety Systems, 27(4), 141–156.
5. Shetty, L., & Khan, F. (2023). Low-cost multimodal inspection using RGB imaging and acoustic templates for food texture detection. International Journal of Food Robotics, 15(1), 61–

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/20
DOI: https://doi.org/10.5281/zenodo.19090739

77.
6. Chakraborty, N., & Pillai, R. (2023). Real-time texture classification system for high-throughput food production lines. Journal of Industrial Food Processing, 10(2), 91–108.
7. Bhandari, A., & Mehta, S. (2022). Synchronized imaging and surface sound profiling for characterizing deep-fried snack textures. Journal of Food Signal Processing, 18(3), 79–95.
8. Ravi, H., & Dutta, M. (2022). Cross-modal template learning for food texture analysis using image and force sensor data. Multimodal Systems in Food Robotics, 21(1), 34–51.
9. Banerjee, A., & Reddy, K. (2022). Multimodal texture recognition in baked and fried foods using 2D/3D imaging and audio analysis. Food Texture and Sensory Systems, 11(2), 122–140.
10. Gupta, R., & Sen, S. (2021). Bimodal texture recognition using image sharpness and chewing-sound signals. Journal of Intelligent Food Inspection, 7(4), 66–82.
11. Mishra, D., & Arora, B. (2021). Multisensory fusion-based food texture analysis using template classifiers. Food Texture Analytics, 14(2), 48–64.
12. Kumar, N., & Thomas, J. (2020). Template matching approach for food texture prediction using image processing and auditory cues. Journal of Early Multimodal Food Analytics, 4(1), 15–31.

