TSINGHUA SCIENCE AND TECHNOLOGY
DOI: 10.26599 / TST.2025. 9010143

GTMH-TasteNet: Advanced Deep Learning for sEMG-Based Taste Sensation Recognition

Asif Ullah, Zhendong Song*, Waqar Riaz and You Wang

Abstract: Medical diagnostics and advanced virtual reality recognize and categorize human taste sensations. This research will examine the recognition of five fundamental human taste sensations—sour, sweet, bitter, salty, and umami—alongside a no-stimulation physiological condition using GTMH-TasteNet, an advanced deep learning framework. Data was obtained through the sEMG device, which can capture signals from muscle activity during taste perception. After collecting the dataset, it was preprocessed involving four steps: detrending, high-pass filtering, adaptive notch filtering, and segmenting the signal into 1-second intervals. Data augmentation methods were employed to improve the model's resilience and generalizability. The GTMH-TasteNet architecture was applied to the processed dataset to extract features, improve interpretability, and emphasize the most significant data elements. The model demonstrated commendable performance, achieving an accuracy of 85.2% in categorizing taste experiences. The results validate that GTMH-TasteNet is highly proficient in detecting taste from sEMG, with potential implications for customized medicine, virtual reality, and multisensory interaction. It presents a scalable, interpretable, and accurate system for identifying taste sensations, advancing human-computer interaction, and taste-driven healthcare.

Keywords: Taste sensations; Muscle activity; GTMH-TasteNet DL framework; Feature extraction; Human-computer interaction; Pattern recognition

1 Introduction

Taste is a vital modality in human sensory perception, playing a dual role in enabling the selection and ingestion of nutritionally valuable food while also constituting a protective mechanism to avoid ingesting toxic substances, thereby ensuring survival and quality of life. More than a biological function, taste experience is an essential dimension of food quality and food-induced emotions, thus being an active determinant in dietary choices and consumer liking [1]. However, taste disorders caused by different medical conditions like cancer, diabetes, anorexia, and COVID-19 infection have become a concern. These disorders can reduce the quality of life and may serve as an early biomarker for underlying diseases [2-5]. Early diagnosis of taste disorders using advanced methodologies may play a pivotal role in personalized medicine and preventive healthcare.

Besides its relevance to medicine, the sense of taste now finds its new frontier in virtual and other emerging environments. Virtual taste has recently gained hype, particularly for remote multisensory interaction applications that enable innovative uses in gaming, entertainment, and immersive digital

- Asif Ullah, Zhendong Song, and Waqar Riaz are with the Institute of Intelligent Manufacturing, Shenzhen Polytechnic University, Shenzhen 518055, China (email: asifkh@szpu.edu.cn, szdeer@szpu.edu.cn, and riazwaqar@szpu.edu.cn)
- You Wang is with the State Key Laboratory of Industrial Control Technology, Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou 310027, China (king_wy@zju.edu.cn)

* Zhendong Song

Manuscript received: 2025-02-19; accepted: 2025-09-08

environments [6]. With the increasing use of brain-computer interfaces (BCIs), understanding and emulating this sensation in virtual reality has gained significant importance, holding great promise for the future with path-breaking discoveries in human-computer interaction [7, 8].

The five basic human taste sensations—sour, sweet, bitter, salty, and umami—are the foundation of the gustatory experience [9]. Taste perception begins with specialized chemoreceptors located primarily on the tongue. These receptors, housed within taste buds, detect diverse taste stimuli and convey the information to the brain through intricate neural pathways [10, 11]. Electrical signals generated by taste receptor cells travel via the vagus, glossopharyngeal, and facial nerves to the brainstem, where they synapse at the nucleus solitarius [12].

The facial nerve's two branches—the chorda tympani and superficial petrosal nerves—play distinct roles. The chorda tympani innervates the anterior tongue and transmits taste-related information, while the superficial petrosal nerve activates taste responses in the soft palate [13]. The glossopharyngeal nerve carries most of the tactile and gustatory information from the posterior tongue, while the vagus nerve carries signals from the epiglottic region. These nerves send their signals to the medulla, which then sends them to the cerebral cortex, where interpretation can occur. After the brain interprets the signals, actions at the taste site are initiated through motor signals from the facial nerve, cranial nerve VII, to the facial muscles, causing expressions of liking or disliking the tasted substance.

Surface Electromyography (sEMG) is a non-invasive technique that records the electrical activity produced by skeletal muscles during contractions in response to stimulation. Electrical impulses are usually received across the skin's surface, which helps quantify and analyze the muscles' activity in eating and swallowing [14-17]. The technique has emerged as a forefront in many applications, including clinical diagnostics, sports science, and prosthetic control, among others [18-24].

The versatility of sEMG is further underscored by its applications in fields ranging from fundamental to clinical sciences. For Example, Dawes et al. applied it to measure the activity of salivary glands in response to gustatory stimulation. At the same time, Mukai et al. investigated changes in facial expressions according to different taste stimuli [18-20]. In particular, Miura et al. discussed the influence of temperature and carbonation on sEMG during swallowing [25]. Similarly, Sato et al. studied sEMG responses to various types of food, focusing on texture and flavor, and provided valuable information on taste perception and related muscle activation responses [26].

However, despite their efficacy, sEMG-based taste studies often encounter difficulties in data interpretation and classification. Conventional analysis methods are limited by manual designs of statistical techniques, which suffer from a lack of adaptability and scalability. Recently, this has shifted the focus toward machine learning (ML) and deep learning (DL) approaches that offer improved precision and flexibility in analyzing complex sEMG data.

Recent breakthroughs in ML and DL have entirely changed the analysis of sEMG signals, especially those for applications related to taste. Although powerful, traditional feature engineering methods are manual and lack the adaptability required for complex datasets [27, 28]. Several deep learning frameworks have emerged that are more effective in learning hierarchical feature representations from raw sEMG signals directly using CNN and LSTM models [29, 30].

For instance, the authors employed FNNs to classify unlabeled sEMG data and retrieved muscle-tendon parameters from muscle force records [31]. Another recent work was proposed by Zanghieri et al., which introduced a method for estimating multi-finger forces using temporal convolutional networks, allowing efficient real-time training on embedded devices [32]. Fatayer et al. proposed a novel gesture recognition framework for myoelectric prosthesis control, leveraging multiresolution sEMG signal mapping and a data-centric CNN-based method to address label noise, achieving over 95% accuracy on large-scale public datasets[33]. Zhou et al. evaluated and compared ML algorithms, including SVM, LR, and ANN, for shoulder motion pattern recognition using sEMG signals, highlighting SVM as the most effective classifier with a peak accuracy of 97.41% at a 270 ms sliding time window [34]. These studies highlight the potential of DL models in sEMG signal processing, offering new possibilities for real-time and scalable applications.

Furthermore, He et al. compared the performance of temporal convolutional networks (TCNs), CNNs, gated recurrent units (GRUs), and Transformers for gesture recognition based on sEMG signals, demonstrating that advanced architectures could effectively capture the temporal and spatial dynamics of muscle activity [35]. Vijayvargiya et al. introduced hybrid deep learning models combining CNN with LSTM or GRU for lower limb activity recognition using wearable sensor data, achieving superior performance metrics compared to individual models [36]. Sakorn et al. employed a ResNeXt-based model to achieve a remarkable 96.73% accuracy in detecting abnormalities in knee sEMG signals, showcasing the efficacy of multiscale convolution in learning complex representations [37]. While previous research on sEMG signals has predominantly focused on motion pattern and gesture recognition, challenges such as limited adaptability, scalability, and the complexity of taste-related muscle activation remain primarily unexplored, prompting our novel work that pioneers the use of advanced machine learning and deep learning techniques to classify basic taste sensations from sEMG signals—a domain scarcely addressed in the existing literature.

This study proposes **GTMH-TasteNet**, a DL architecture that leverages CNN with a multi-head attention layer and a granular encoder to recognize the basic taste sensations, as mentioned in **Fig. 1**. First, several basic tastes were placed on the tongue's tip to excite facial muscles. Secondly, the dataset was collected from facial muscles through an sEMG device. The dataset was preprocessed in four steps: detrending, high-pass filtering, notch removal, and segmenting the signal into 1-second intervals. After preprocessing, the dataset was augmented to enhance the training robustness. Finally, GTMH-TasteNet was applied to identify several taste states. The proposed model identifies physiological states and taste sensations, including sweet, sour, salty, bitter, and umami, achieving an accuracy of 85.2% and demonstrating its efficacy in recognizing taste patterns.

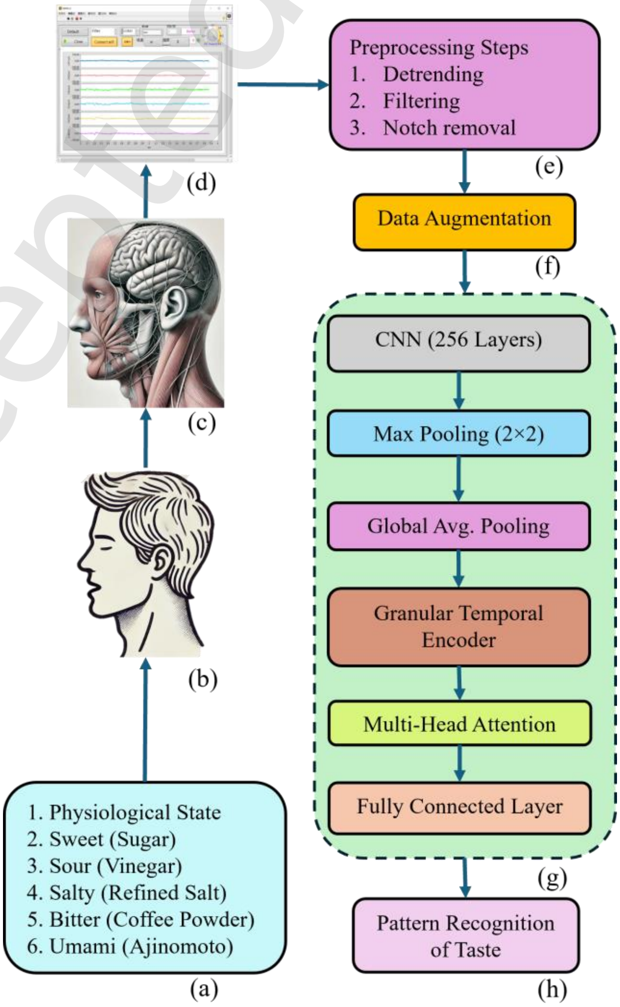

Fig. 1 Schematic diagram (a) Basic Tastes (b) Taste materials were put on the tongue's tip (c) the taste buds send commands to the brain which in return excite facial muscles (d) Facial signals were collected through an sEMG device (e) the dataset was preprocessed (f) Dataset was augmented (g) the neural network was applied on the augmented dataset (h) Basic taste recognition.

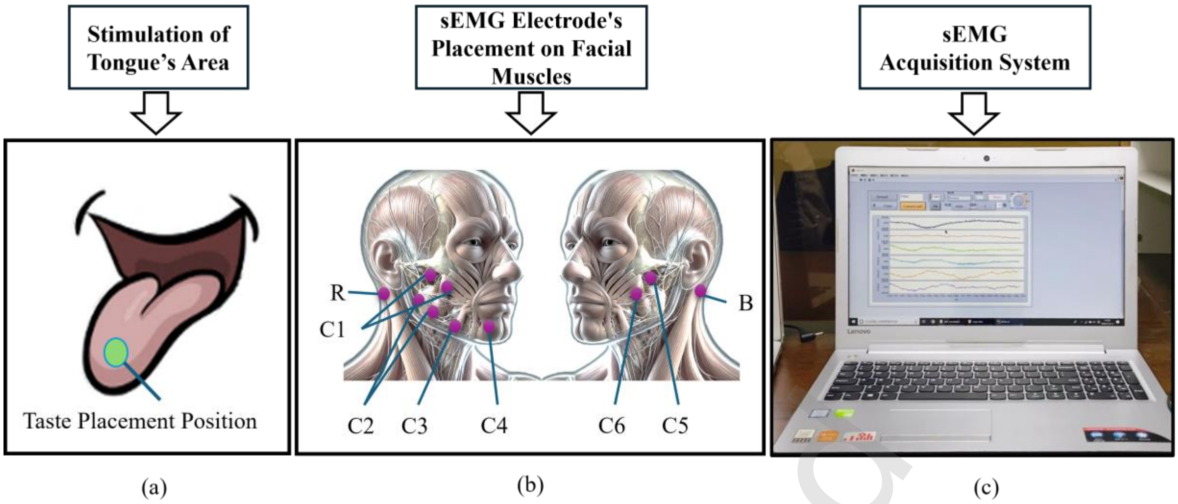

Fig. 2 Data Collection Environment (a) Tongue's stimulation area (b) Silver Electrodes placement on facial muscles (c) sEMG acquisition device

2 Data Collection

2.1 Configuration

The experimental setup and sEMG data acquisition system were carefully customized to ensure high-quality data collection. A soundproof and comfortable environment was prepared for the participants to minimize external distractions. Given the unique nature of taste perception, participants were advised to refrain from consuming food or beverages, except water, for at least two hours before the experiment and to avoid spicy foods for 24 hours preceding the session. Taste stimuli were applied to the tongue, and signals were recorded using an sEMG device equipped with six channels and ten electrodes, operating at a sampling frequency of 1 kHz. The experimental setup is illustrated in **Fig. 2**.

2.2 Materials

The five fundamental taste sensations—sour, sweet, bitter, salty, and umami—are utilized as stimuli to generate sEMG signals, providing insights into taste perception and its neural-muscular responses. Besides these taste stimuli, participants' physiological state is also explored to better understand its impact on the sEMG signals related to taste. Specific substances were carefully selected to represent each taste category: sugar for sweet, vinegar for sour, refined salt for salty, coffee powder for bitter, and monosodium glutamate (MSG) for umami. These standardized substances will ensure uniformity and dependability in eliciting the corresponding taste responses in the experiment [14].

2.3 Subjects

Eight people participated in the studies. All subjects had good physical and mental health and no history of illnesses associated with taste or the neurological system. Before the experiment, each participant underwent a pretest to assess their gustatory function, ensuring their eligibility for the investigation. No participant exhibited any illness or signs of dysgeusia throughout the research. The Ethical Committee of Zhejiang University's College of Biomedical Engineering and Instrument Science approved this study.

2.4 Data Recording Protocol

The sEMG device was examined to ensure it was functioning correctly before the experiment. Participants were instructed to cleanse their faces, sanitize their skin with alcohol wipes, and remain relaxed while electrodes were carefully placed on their faces. A circular spoon was used to ensure uniformity and convenience in presenting five distinct tastes on the tip of the tongue. The subjects were asked to close their eyes and mouths, relax their minds, and refrain from chewing during the test. sEMG data for each taste stimulus were recorded for 12 seconds.

After the 12-second recording period, participants rinsed their mouths with water and remained inactive for another two minutes before the next taste sensation. This way, after all the taste sessions had been completed, participants were afforded a 10-minute break. Following this session, a session was conducted in just the same way. All participants performed these protocols to bring consistency and reliability throughout the study.

2.5 Dataset

There were 60 experiments, resulting in 180 sessions, each with six trials spanning all six tastant conditions- a no-stimulation physiological state and the five basic tastes. sEMG data was recorded for 12 seconds during each trial. In total, 2,160 data samples were collected for each taste state, resulting in 12,960 samples across all six taste states, thereby ensuring a comprehensive dataset for analysis.

2.6 Data Processing

Fig. 3 demonstrates the sequential data preprocessing steps applied to the raw sEMG signals. The dataset was segmented into 1-second intervals to ensure a thorough analysis of the signal. The preprocessing involved four distinct stages: (a) raw signal acquisition, (b) detrending, (c) high-pass filtering, and (d) adaptive notch filtering. The raw signals (**Fig. 3a**) exhibit baseline drifts and noise components. Detrending (**Fig. 3b**) was performed to eliminate polynomial trends and DC offset, effectively reducing baseline drift and restoring signal fluctuations to within the acquisition system's acceptable range (approximately ±1mV).

A high-pass Butterworth filter with a 10 Hz cutoff frequency was then applied to suppress further low-frequency components, including motion artifacts (**Fig. 3c**). The high-pass filter cutoff of 10 Hz was selected based on literature indicating that increasing the high-pass filter cutoff frequency successfully reduces movement artifacts without significantly altering the sEMG signal. The research demonstrates that movement artifacts are significantly reduced at a 10 Hz cutoff, with diminishing results at higher frequencies [38]. Additionally, a study in the field of taste sensation recognition employed a 10 Hz high-pass filter to preprocess sEMG data, removing noise while preserving the desired muscle activity signals. Therefore, using a 10 Hz high-pass filter cutoff is physiologically and methodologically justified for analyzing face muscle activity in response to taste [6].

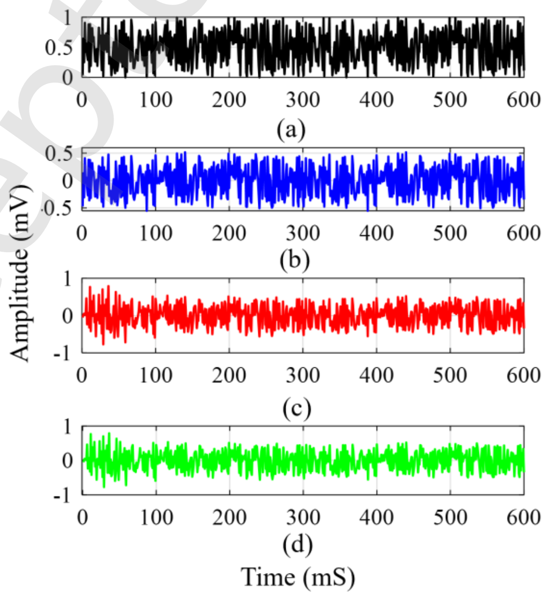

Fig. 3 Data Processing steps (a) Raw Signal (b) Detrended Signal (c) Filtered Signal (d) Signal After Notch Removal

Following this, an adaptive notch filter was employed to eliminate power line interference (50 Hz) and its harmonics (**Fig. 3d**). The implementation of a 50 Hz notch filter in our preprocessing sequence is essential to mitigate power line interference (PLI), a prevalent source of noise in sEMG recordings, especially in regions where the electrical power grid operates at 50 Hz [39]. PLI can significantly disrupt sEMG signals, resulting in poor signal quality and potential misunderstandings. At this frequency, a notch filter effectively lowers interference, improving the signal-to-noise ratio while preserving the integrity of the physiological information. These preprocessing steps collectively enhanced the signal quality, ensuring clean and interpretable sEMG waveforms. The visual comparison of the raw and processed signals highlights the efficacy of the preprocessing methods, showcasing a progression toward cleaner signals while preserving relevant physiological information.

A data augmentation approach was implemented after signal preprocessing to increase the diversity and quantity of the training dataset. Each original sEMG data was enhanced with additive Gaussian noise to replicate slight physiological differences and sensor inconsistencies. Zero-mean Gaussian noise with a standard deviation of 0.1 was added to each raw sample to create nine extra samples. The dataset size increased tenfold as a result of this procedure, which also helped enhance model generalization and more accurately depict the variability in actual sEMG signals.

3 Methodology

After the preprocessing and data augmentation steps, a neural network named GTMH-TasteNet was applied to classify the basic taste sensations. The neural network structure is explained in the following paragraphs.

3.1 Structure of GTMH-TasteNet

This section introduces a versatile and adaptive neural network architecture that combines the hierarchical feature extraction of convolutional layers with the targeted focus of granular encoder and multi-head attention mechanisms to process sEMG signals for taste sensation classification, as illustrated in **Fig. 4**. This neural network is structured to capture and analyze both spatial and temporal patterns in muscle activity corresponding to the perception of different taste sensations, such as sour, sweet, salty, bitter,

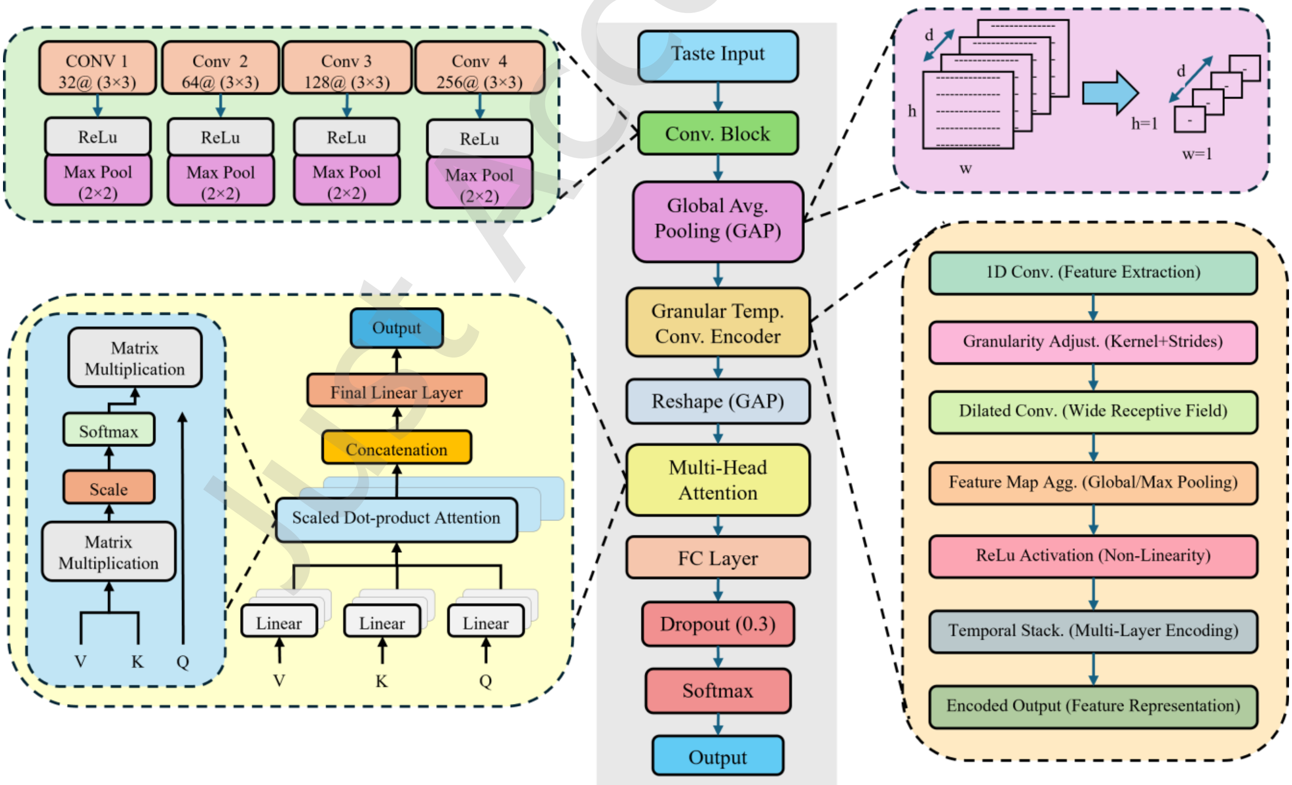

Fig. 4 The architecture of the GTMH-TasteNet neural network. The central section illustrates the main network pipeline, integrating key modules. The left side highlights the convolutional block and multi-head attention mechanism for feature extraction and refinement, while the right side details the Global Average Pooling (GAP) and Granular Temporal Convolutional Encoder (GTCE) for temporal and spatial feature representation.

umami, and the no-stimulation physiological state. This approach excels in handling complex sEMG-based taste signals, where even minor variations in features have a significant impact on classification accuracy. The following sections detail the specific adaptations made in the network to address the unique challenges of this classification task.

3.2 Taste Input

The input consists of sEMG signals recorded from facial muscles during taste stimulation. These signals encode the electrical activity patterns associated with perceiving specific tastes.

3.3 Convolution Block

The convolution operation extracts temporal features from the sEMG signals. These layers scan the input for localized patterns corresponding to distinct taste responses. Each convolutional layer applies filters that detect fine-grained temporal patterns unique to each taste, such as the sharp transitions in muscle activation for sour tastes versus the smoother activation for umami. The convolution block can be mathematically represented as

$$O[i,j,k] = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X[i+m, j+n, l] \,.\, W[m,n,l,k] + b[k] \quad (1)$$

where *O [i, j, k]* represents the feature map value at position *(i, j)* for the *k*-th filter, *X [I +m, j +n, l]* denotes the input feature map value at position *(i, j)*, *W [m, n, l, k]* signifies filter weights of size M × N for the *k*-th filter and *b[k]* is the bias term for the *k*-th filter.

3.4 Global Average Pooling (GAP)

GAP computes the mean values over the spatial dimensions of each feature map, resulting in a singular representative value of muscle activity for each feature map. It extracts the essence of muscle activation patterns most indicative of the tasted stimuli. By summarizing spatial information into a single value per feature map, GAP makes the representation more robust and compact, emphasizing consistent patterns across the facial muscle group. Mathematically,

$$F_{GAP} = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} F(i,j) \quad (2)$$

F_{GAP} denotes the feature map's global representation, and *F (i, j)* represents the activation value at position *(i, j)*.

3.5 Granular Temporal Convolution Encoder (GTCE)

The GTCE integrates sophisticated operations and architectural improvements to optimize granularity, computational efficiency, and learning temporal dependencies. It is tailored to extract temporal dependencies in the sEMG signals, reflecting the dynamic nature of muscle activity during the perception of taste. GTCE comprises the following layers.

3.5.1 1D Convolution Layer

The convolution operation extracts temporal features from the sEMG signals. This layer scans the input for localized patterns corresponding to distinct taste responses and applies filters that detect fine-grained temporal patterns unique to each taste. A 1D convolution layer can be represented as:

$$O[t,k] = \sum_{i=0}^{k-1} X[t+i] . W[i,k] + b[k] \quad (3)$$

O[t, k] denotes the output feature map at time *t* for the *K-th* filter, *and X [t+1]* represents the input signal at time *t*+*i*, where *i* ranges over the kernel size. *W [i, k]* refers to the weight at position *i* of the *K*-th 1-D convolution filter, and b*[k]* is the corresponding bias term.

3.5.2 Granularity Adjustments

Granularity adjustment ensures that the network captures patterns at varying levels of detail. Smaller kernel sizes focus on precise, localized muscle activations, while larger strides help capture broader temporal trends. This adjustment is crucial for distinguishing between overlapping tastes, such as sour and salty, which may share similar spatial features but differ in temporal dynamics. Mathematically,

$$O[t] = \sum_{i=0}^{k-1} X[t + i.s].W[i] + b \qquad (4)$$

where s denotes the strides controlling the granularity (spacing between samples).

3.5.3 Dilated Convolutions

Dilated convolutions expand the receptive field, enabling the network to analyze long-term dependencies in muscle activity. For instance, the prolonged activation patterns seen in bitter taste responses are effectively captured without increasing computational complexity. DC can be represented as:

$$O[t] = \sum_{i=0}^{k-1} X[t + r.i].W[i] \qquad (5)$$

where r is the dilation rate (the gap between elements in the input), and K represents the kernel size.

3.5.4 Feature Map Aggregation (FMA)

Features extracted from convolutional layers are aggregated using Global Average Pooling (GAP) and Max Pooling. GAP summarizes the spatial dimensions of the feature maps into a single value, emphasizing the most significant muscle activation patterns for each taste. On the other hand, Max Pooling captures the most dominant activation, which helps highlight strong responses to stimuli. FMA includes Max pooling and Global pooling. Mathematically,

$$p[k] = \max_{t \in T} X\,[t, k] \qquad (6)$$

$$p[k] = \frac{1}{T} \sum_{t=1}^{T} X[t, k] \qquad (7)$$

$P[k]$ denotes the maximum value in the k-th channel over the time window T in the above equations.

3.5.5 ReLu Activation

ReLU introduces non-linearity, enabling the model to distinguish subtle variations in muscle activity for closely related tastes. Negative values in the feature maps are replaced with zero, ensuring only meaningful activations are considered.

$$O[t, k] = \max(0, X[t, k]) \qquad (8)$$

where $O\ [t,\ k]$ is the activated output.

3.5.6 Temporal Stacking

By stacking multiple convolutional layers, the encoder progressively increases the granularity of its analysis. Each successive layer captures more complex temporal patterns, combining localized responses to form a universal representation of muscle activity. For Example, this stacking helps differentiate between the sharp, short-lived activations for sour tastes and the sustained, steady activations for umami.

$$Z = Stack\ (0_1, O_2, \ldots, O_L) \qquad (9)$$

where O_1, O_2,…, O_L show the output of each layer in the temporal stack, and Z represents the encoded temporal features.

3.5.7 Encoded Output

The final encoded output aggregates all granular features, comprehensively representing the muscle activity patterns associated with each taste. This representation effectively resolves ambiguities in overlapping responses, such as distinguishing between salty and umami, which may activate similar muscle regions but exhibit distinct temporal characteristics.

$$Y = f(w.Z + b) \qquad (10)$$

w represents the weight matrix mapping the stacked features to the output space, and f denotes the non-linear activation function (e.g., softmax or linear).

By combining the above equations, the overall operation can be represented as:

$$Y = f(w.ReLu(Pool(\sum_{i=0}^{K-1} X[t + r.i].W[i])) + b) \quad (11)$$

3.6 Reshape Layer

The output from the GTCE layer is restructured to meet the specifications of the multi-head attention mechanism, enabling the model to handle the condensed characteristics for thorough contextual analysis efficiently.

3.7 Multi-head Attention (MHA)

The multi-head attention mechanism emphasizes the most relevant features in the input by establishing relationships between different parts of the data. MHA is handy for resolving overlapping taste responses, such as distinguishing between salty and sour stimuli, which may activate similar muscle groups but with varying intensities and temporal patterns. Input features are transformed into three distinct spaces: Query (Q), Key (K), and Value (V), which serve as the foundation for attention computations. The scaled dot-product attention calculates the relevance of features by performing the dot product of Q and K, scaling the result, normalizing it with a softmax function, and weighting it with V to highlight essential features. Outputs from multiple attention heads are concatenated and processed through a final linear layer, effectively aggregating diverse perspectives and enhancing the model's representational power. Mathematically,

$$Q = XW_Q \quad (12)$$

$$K = XW_K \quad (13)$$

$$V = XW_V \quad (14)$$

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (15)$$

$$MHA(X) = Concat(head_1, head_2, \ldots, head_h)W_O \quad (16)$$

where W_Q, W_k, and W_V are the weight matrices for Query, Key, and Value, QK^T denotes the scaled dot-product between Query and Key, d_k represents the dimensionality of the Key, $head_i$ is the independent attention head, and W_O is the output projection matrix.

3.8 Fully connected/Softmax Layer

The final predictions of the model are generated through a softmax layer, which translates the processed outputs from the dense layer into probabilities corresponding to each tasting class. These probabilities indicate the likelihood that the input belongs to specific taste sensation categories, such as sour, sweet, salty, bitter, umami, or a state of no stimulation. The softmax function ensures these probabilities are normalized, making them interpretable and summing to one across all classes. This step is essential for classification tasks, enabling the model to allocate a definite class label based on the highest probability.

$$y = W^T x + b \quad (17)$$

where x represents the input vector, W is the weight matrix, and b represents the bias vector.

3.9 Output layer

The output layer is the final layer of this model, where the final features are translated into final class predictions. These predictions represent the classification of the input data into one of the predefined categories: sour, sweet, salty, bitter, umami, or no stimulation physiological state. This layer utilizes the probabilities generated by the softmax function to determine the most likely class, ensuring that the model's predictions are accurate and interpretable. The model then selects the class with the highest probability, defined as:

$$y = Softmax(W_x + b) \quad (18)$$

3.10 Cross-Entropy Loss Function

GTCE employs the Cross-Entropy Loss Function to optimize model parameters for supervised learning tasks. Mathematically,

$$L = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_i^c \,.\log(\hat{y}_i^c) \qquad (19)$$

where N specifies the number of samples, C indicates the number of classes, y_i^c represents the true label and $\hat{y}_i^c$ defines the predicted probability for class c.

Table 1 Training Parameters and Model Configuration Details to Optimize the Proposed Model's Performance

Parameters	Value
Epochs	200
Iterations	60.8 K
Batch Size	16
Validation Frequency	10
Total/Trainable Param	490,887 (1.87 MB)
Non-Trainable Param	0 (0.00 B)
Training Time Duration	64.2167 min
Early Stopping Factor	0.2
Learning Rate	0.001
Train/Test Ratio	9/1

4 Results and Discussion

4.1 Evaluation Metrics

Table 1 shows the hyperparameters of the proposed model. The proposed model was trained for over 200 epochs, comprising 60,800 iterations, with a batch size of 16. A validation frequency of every 10 iterations ensured robust model performance monitoring during training. The architecture consisted of 490,887 total parameters (1.87 MB), all of which were trainable, reflecting the absence of non-trainable parameters. The training process spanned approximately 64.22 minutes, leveraging an early stopping factor of 0.2 to prevent overfitting and optimize performance. The model was trained using a learning rate of 0.001, following a 9:1 train-test split, ensuring sufficient data for both learning and validation. These parameters underscore the training regimen's computational efficiency and balanced design, facilitating effective learning while minimizing resource utilization.

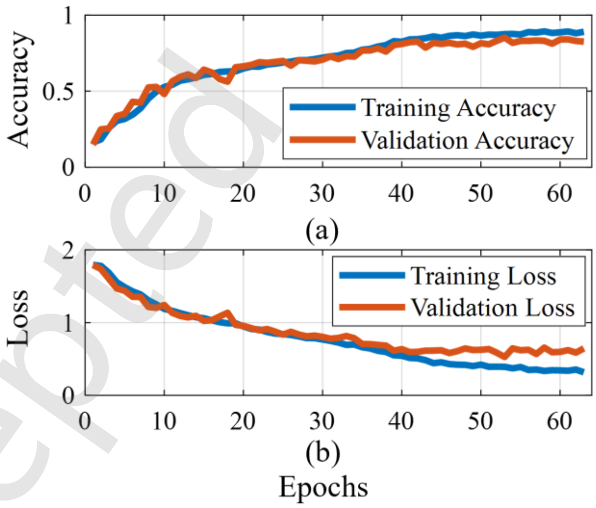

Fig. 5 Performance trends of validation (a) accuracy and (b) loss across 60 epochs, highlighting model learning dynamics.

4.2 Training and Validation Accuracy/Loss

As depicted in **Fig. 5**, the training and validation accuracy provide insights into the model's performance during the training process. **Fig. 5(a)** illustrates that the training accuracy steadily increases as the model learns from the data, demonstrating a consistent improvement in its ability to classify the training samples accurately. Similarly, the validation accuracy follows an upward trajectory, indicating that the model generalizes well to unseen data. The corresponding loss plot shows a consistent decrease in both training and validation loss, as mentioned in **Fig. 5(b)**, indicating effective learning. Minor fluctuations in the validation loss may reflect noise in the validation data or slight overfitting as the model's complexity increases. These results collectively suggest that the model strikes a good balance between learning and generalization.

Table 2 Performance metrics for each class, including TPR, TNR, FPR, FNR, and accuracy.

Class	TPR	TNR	FPR	FNR	Accuracy
S0 (Physiological State)	0.90	0.9816	0.0184	0.1000	0.9558
S1 (Sweet)	0.82	0.9697	0.0303	0.1800	0.9367
S2 (Sour)	0.86	0.9631	0.0369	0.1400	0.9408
S3 (Salty)	0.91	0.9559	0.0441	0.0900	0.9492
S4 (Bitter)	0.81	0.9464	0.0536	0.1900	0.9150
S5 (Umami)	0.80	0.9696	0.0304	0.2000	0.9458

4.3 Class Level Performance Metrics

The classification results indicate that the classes are well-separated; the model achieved an accuracy of 85.2% on the validation set. This performance evaluation is further supported by class-wise metrics and F1 scores, with detailed data visualizations using PCA and t-SNE analysis.

Table 2 summarizes the model's class-wise performance metrics, enabling a thorough examination of the model's classification ability across the six classes. It establishes that, among those, the model promises a high True Negative Rate (TNR) value, ranging between 94.64% and 98.16% for all classes, thus demonstrating its efficiency in correctly identifying class instances with a high accuracy rate. Similarly, the sensitivity measure, represented by the True Positive Rate (TPR), is robust and ranges between 80.0% and 91%, indicating robust performance in detecting samples of their true class. The highest TPR, 91%, is recorded for Class S3, while the lowest, 80.0%, is recorded for Class S5, showing slight variations in class-specific detection effectiveness.

The FPR and FNR further detail the error patterns. The values of FPR are constant and low, ranging from 1.84% to 5.36%, indicating minimal misclassification for the class samples. However, FNR, which measures missed detections, exhibited a wider fluctuation, ranging from 9.00% to 20%, indicating difficulty in distinguishing certain classes. Among them, Class S3 has the best balance, with a low FNR of 9.00% and a high TPR of 91%, while S5 has the highest FNR of 20%, which indicates room for improvement in the correct classification of this class.

Overall, the accuracy for all classes is high, ranging from 91.50% to 95.58%, which reflects the model's strong generalization capability. However, slight imbalances in class-specific performance hint at the need for possible fine-tuning to further improve the detection rates of underperforming classes. These metrics are compiled to provide valuable insights into the strengths and limitations of the model, thereby helping to optimize efforts further and enhance classification performance.

4.4 Data Visualization

Fig. 6 depicts the data visualization. The two-dimensional view of PCA **(Fig. 6(a))** is based on a projection of the data onto the first two principal components. Each dot represents a sample, while the color is classified into six classes, ranging from S0 to S5. The clusters formed from samples of the same class indicate the data's separability in the reduced-dimensional space. Although PCA captures a great deal of class-specific variance, some overlapping clusters suggest that classes "S3," "S1," "S5," and "S0" are similar, which causes problems in classification.

The 3D visualization of PCA **(Fig. 6(b))** extends the preceding plots to include the third principal component of the data, aiming to reveal class separation. Compared to the 2D PCA version, three-dimensional PCA has increased cluster separability. It shows how the third principal axis captures some aspects of the overall variance. Despite the improved separation, some overlap, such as "S1" and "S2", reflects the underlying complexity and similarity between certain classes. This further underlines the role of dimensionality reduction in understanding data structure.

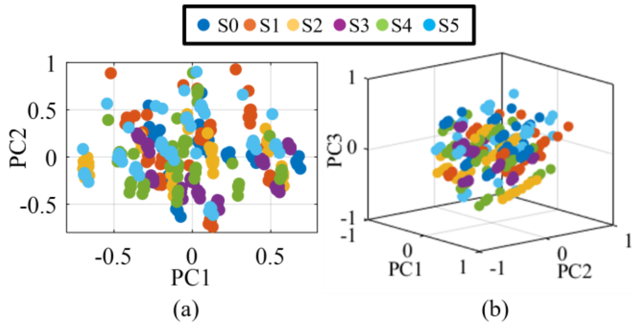

Fig. 6 2D and 3D PCA visualizations of taste states (S0-S5) showcasing class separability in reduced dimensions.

4.5 T-SNE Visualization

T-SNE visualization provides a higher-dimensional data representation by projecting it into a two-dimensional plane and attempts to preserve the structure at multiple scales, both locally and globally, as shown in **Fig. 7**. This plot clearly highlights how class-specific clusters are identified and well-defined. Nevertheless, there are still slight overlaps between classes, and a few classes exhibit similarities in their shared features. The t-SNE projection here reiterates its usefulness in capturing complex patterns, in addition to assessing class separability from high-dimensional data points, and offering insight into the data structure and relationships.

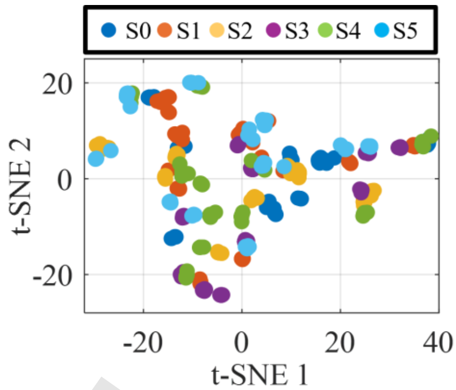

Fig. 7 t-SNE visualization of taste states (S0-S5) demonstrating class clustering in a two-dimensional space.

4.6 Evaluation of Clustering Metrics

Three widely used clustering metrics—the Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index—were applied across a range of cluster counts from 2 to 20 to identify the ideal number of clusters in our dataset, as illustrated in **Fig. 8**.

The Silhouette Score (**Fig. 8(a)**) evaluates the fit of each data point within its designated cluster relative to neighboring clusters, with values approaching 1 indicating enhanced cohesiveness and separation. The score consistently rises with the number of clusters, reaching its highest point at 14, indicating an optimal balance of compactness and distinctiveness.

The Davies-Bouldin Index (**Fig. 8 (b)**) compares intra-cluster scatters to inter-cluster separation to determine the average similarity between clusters.

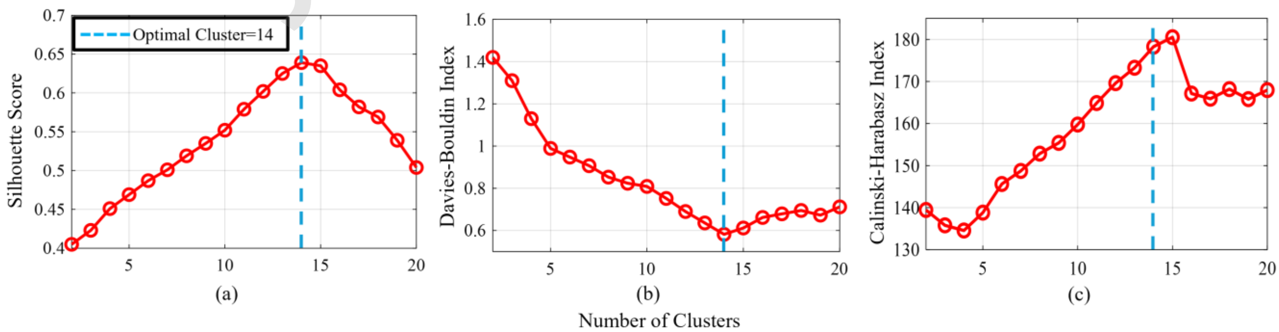

Fig. 8 Determination of the optimal number of clusters using three different clustering evaluation metrics: (a) Silhouette Score, (b) Davies–Bouldin Index, and (c) Calinski–Harabasz Index

Lower numbers indicate better clustering quality. The index declines with an increase in cluster numbers, reaching a minimum of around 14 clusters, which confirms the optimality of this cluster count through improved separation and reduced overlap.

The Calinski-Harabasz Index (**Fig. 8 (c)**) assesses the ratio between cluster variation and within-cluster variance, where elevated values indicate distinct, compact, and well-separated clusters. This parameter constantly rises and reaches its peak at approximately 15 clusters. However, cluster 14 is considered an optimal cluster due to its overall feature representation, which further validates the robustness of the clustering structure at this level.

These measurements collectively indicate that 14th is the optimal number of clusters, offering the best balance between compactness and separation. This selection ensures significant categorization, enhances interpretability, and supports the reliability of subsequent studies derived from the clustering outcomes.

4.7 Confusion Matrix

As mentioned in **Fig. 9**, the confusion matrix gives a detailed performance evaluation of the classification for the six states: sweet, sour, salty, bitter, umami, and baseline. Each entry along the diagonals of the matrix represents the TPR of a particular class, such as 0.90 for S0 and 0.91 for S3, reflecting the appropriate model classification for the requested taste states of S0 and S3. On the other hand, the off-diagonal elements represent FPR and FNR, which are the cases of misclassification for samples. For Example, S0 has a negligible FPR of 0.07 and an FNR of 0.10, indicating low error rates but still leaving room for improvement. Results vary across classes, with the lowest TPR being 0.81 for S4 and 0.80 for S5, indicating a slightly increased misclassification rate for these classes. The S4 and S5 classes may have lower identification rates than other taste states because of their more varied or mild facial muscle activation patterns, which result in less identifiable sEMG signals. Because S4 is linked to adverse reactions that vary from person to person, and the S5 class frequently evokes mild responses, the model's capacity to reliably distinguish between these classes is reduced. The findings underscore the overall robustness of the model's accuracy but highlight the need to address specific misclassification trends for enhanced reliability.

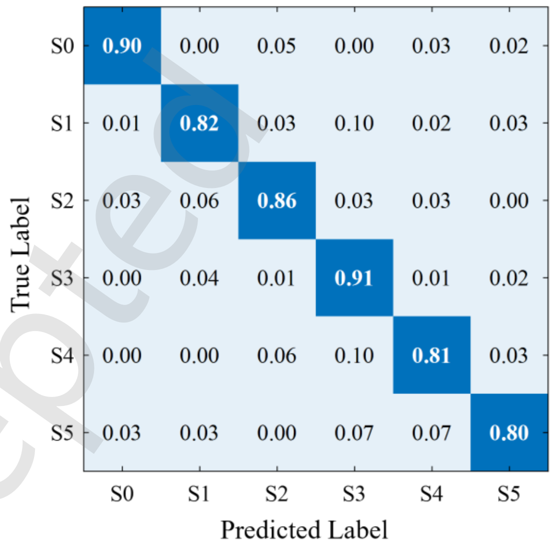

Fig. 9 Confusion matrix illustrating classification performance across six taste states (S0-S5) with normalized values.

4.8 Precision-Recall and F1-Score

The performance metrics in **Fig. 10** for the six classes demonstrate that the suggested model provides highly effective classification. The precision across all classes exhibits significant variation, ranging from 0.9464 to 0.9816, indicating the model's accuracy in correctly identifying true positives among all predicted positives. The recall values indicate the sensitivity of the classes, with S3 (Salty) at 0.91, signifying that the model effectively identified the true positives in this category. Correspondingly, the F1-scores, representing the harmonic mean of precision and recall, likewise seem elevated, balancing both metrics. Specifically, S0 exhibits a physiological condition with an F1-score of 0.9391, whereas other categories, such as Sour, Salty, and Sweet, demonstrate competitive ratings over 0.89. The results validate the model's efficacy in accurately classifying taste sensations and physiological states, enhancing its dependability and applicability in practical scenarios.

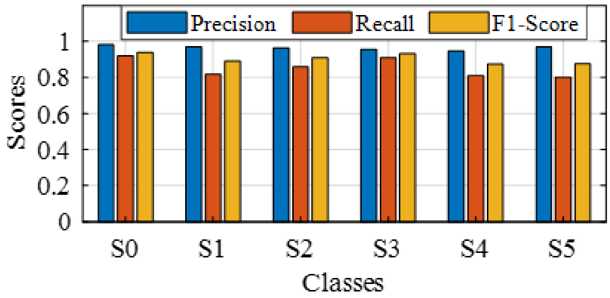

Fig. 10 Bar chart depicting precision, recall (TPR), and F1-Score for each taste state classification (Class 0 to Class 5)

4.9 Error Histogram

The error histogram in **Fig. 11** provides a more detailed representation of the prediction errors for the model in classifying the six taste states. The x-axis is the error value computed as a difference between the predicted and true class labels, while the y-axis is frequency. A dominant peak at zero error demonstrates the model's robustness, with most predictions aligning with the actual labels. This implies that the model's performance is comparatively satisfactory across most datasets. Conversely, nonzero error values do indicate misclassifications; however, they do so at a reduced frequency. These errors can be credited to overlapping signal characteristics among the six taste states, noisy sEMG signals, or an imbalance in the classes within the dataset. Errors of ±1 suggest the presence of adjacent misclassifications, where the predicted label is one class away from the actual label. The higher absolute values of errors indicate more significant deviations that may be linked to complex signal patterns or limitations in the model's ability to distinguish subtle differences between taste states fully. This analysis indicates the overall reliability of the model; at the same time, there is still room for optimization to attain probably better preprocessing, balancing between classes, or advanced means of feature extraction.

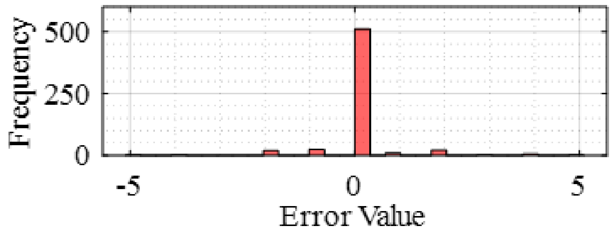

Fig. 11 Error histogram showing the frequency distribution of prediction errors, highlighting minimal deviations around zero.

4.10 Ablation Study

A systematic ablation study evaluated the individual contributions of essential components inside the GTMH-TasteNet architecture by selectively deleting each main module and assessing the impact on classification accuracy. The accuracy dropped significantly to 73.35% upon removing the convolutional block responsible for the initial spatial feature extraction from sEMG data, which underscores the block's vital role in detecting localized muscle activation patterns necessary for taste distinction. The accuracy dropped to 77.58% upon the exclusion of the Granular Temporal Convolution Encoder (GTCE), which is responsible for modeling temporal dependencies through dilated convolutions and granularity adjustments, necessitating the need for temporal feature extraction to represent dynamic muscle responses during taste stimulation. The significance of the multi-head attention mechanism in improving feature discrimination among closely related taste classes is underscored by the fact that its absence reduces accuracy to approximately 81.23%. This mechanism allows for improved feature representation by modeling long-range dependencies and highlighting prominent features.

Furthermore, epochs and train/test ratio-based experimental setups were conducted to evaluate the impact of different training configurations on the proposed model's performance. **Table 3** summarizes the configuration setups and their effects on validation accuracy and model performance.

In Setup 1, the training process was limited to 50 epochs, each consisting of 304 iterations, resulting in a total of 15,200 iterations. The training and testing data were split in a 7:3 ratio. This configuration demonstrated a gradual convergence of training and validation accuracies, achieving a maximum validation accuracy of 72.37%. The training accuracy did not fully converge, and the training loss remained relatively high throughout, indicating that the model was not sufficiently optimized. The relatively low validation accuracy suggests underfitting, where the model fails to capture the complexity of the data. Setup 1 required 16 minutes and 5 seconds, making it the fastest configuration due to the limited number of epochs.

Setup 2 was carried out with 100 epochs, maintaining the same number of iterations per epoch (304), resulting in a total of 30,400 iterations. The training and testing split was modified to 8:2. This configuration achieved a validation accuracy of 79.81%. The training accuracy approached 80%, and the training loss consistently decreased, indicating effective optimization of the training data. The validation accuracy reached 79.81%, with minimal improvement in validation loss, suggesting a potential risk of overfitting. The model performed exceptionally well on training data but struggled to generalize to unseen data. The increase in training data led to improved validation accuracy compared to Setup 1. However, overfitting became more pronounced due to the higher number of epochs and a reduced validation set.

Setup 3 employed the most extensive training configuration with 200 epochs, maintaining 304 iterations per epoch, resulting in 60,800 iterations. The training and testing split was adjusted to 9:1. This setup achieved the highest validation accuracy of 85.2%. The training accuracy rapidly converged to almost 90%, and the training loss declined steadily, indicating efficient optimization. Despite achieving the highest validation accuracy, the gap between training and validation losses suggests a significant potential for overfitting. While increased iterations and epochs improved training performance, they constrained the model's ability to generalize effectively. Setup 3 outperformed the previous configurations in validation accuracy.

Table 3 Ablation study comparing training configurations with varying epochs, iterations, and validation accuracies across three setups.

Training Configuration	Setup 1	Setup 2	Setup 3
Toral Epochs	50	100	200
Iteration/Epochs	304	304	304
Max Iteration	15,200	30,400	60,800
Training/Testing	7/3	8/2	9/1
Val; Acc; (%)	72.37	79.81	85.20

4.11 Comparison Study

Several comparison analyses were observed to evaluate the efficiency of this study, as mentioned in **Table 4**. The comparison analysis reveals significant inconsistencies that differentiate the performance of various models in recognizing taste sensations through sEMG signals. These comprise the Random Forest model, which achieves an accuracy of 74.46% [27], and the SVM model, which achieves an accuracy of 76.42% [40]. While these models exhibit reasonable robustness, they lack temporal modeling capabilities, which constrains their suitability for dynamic sEMG data. For instance, studies such as Olfactory-Gustatory sEMG and Manometry [41] achieve approximately 80% accuracy, comparable to the findings on Bitter Taste Substances and Salivation [42]. However, they are confined to specific subjects and concentrate on a limited number of categories. The Random Forest classification model, which utilized only frequency and time domain variables, achieved an accuracy of 42.42% due to its limited ability to predict temporal dependencies in this investigation [43].

Consequently, the newly designed GTMH-TasteNet neural network achieved the highest accuracy of 85.2% among all the developed models. This score indicates that the model performs better because it can effectively identify spatial and temporal sEMG patterns, capture granular and multi-head attention, and have a more profound and multi-scaled temporal

Table 4 A comparative study showing the performance of previous and current model

Model/Study	Acc; (%)	Dimensionality	Robustness	Key Characteristics
Random Forest (RF) **[27]**	74.46	Moderate	Moderate (Subject-Specific)	Handles feature sets well but lacks temporal context modeling.
Support Vector Machine (SVM) **[40]**	76.42	High	Moderate	Effective with high-dimensional features but sensitive to subject-specific variations.
Olfactory-Gustatory sEMG and Manometry **[41]**	~80	Moderate	Moderate (Subject-Specific)	Simultaneous odor and tastant stimuli enhance sEMG amplitude and swallowing biomechanics. Best for integrating sensory modalities.
Taste and Salivation Study **[42]**	Not specified	High	Moderate	Bitter taste stimuli alter salivation and swallowing processes distinctly. Practical for sensory studies, not primarily classification-focused.
RF Frequency & Time Features **[43]**	42.42	High	Moderate	Demonstrated strong feature learning capability for taste recognition, but lacked advanced temporal dependency modeling.
GTMH-TasteNet **(Current Study)**	85.2	High	High	Captures spatial and temporal sEMG patterns effectively and is robust against subject variability.

convolution. Additionally, by clarifying ambiguities from potentially overlapping responses linked to salty or umami flavors, the recognition accuracy of these complex taste categories remains high, maintaining a robust and efficient level of performance, thereby presenting a newly evolving perspective on taste recognition.

5 Discussion and Limitations

The study demonstrates the efficacy of GTMH-TasteNet in recognizing taste sensations through sEMG signals, achieving a notable classification accuracy of 85.20%. The results highlight the model's ability to extract hierarchical and fine-grained features using convolutional layers while leveraging granular encoder and multi-head attention mechanisms to focus on critical data aspects. Integrating advanced preprocessing techniques, such as detrending, high-pass filtering, and adaptive notch filtering, ensured cleaner input signals, thereby improving model performance. Additionally, data augmentation contributed significantly to enhancing model robustness and generalizability.

Nonetheless, certain limitations warrant consideration: firstly, the proposed dataset comprises only 12,960 samples across six taste states, which, while comprehensive, may not fully represent the complexity of human taste perception. Secondly, the inherent noise in sEMG signals may still compromise classification performance, even after preprocessing, particularly in instances with overlapping categories such as "salty" and "umami." Thirdly, the research is limited to six specified categories, and expanding the model to accommodate more intricate taste states may prove challenging. Finally, Setup 1, with fewer epochs and a 70:30 train-test split, demonstrated underfitting, low validation accuracy, and high training loss, suggesting a lack of model capacity or training time to capture complex sEMG patterns, while Setups 2 and 3, with more epochs and training splits, improved validation accuracy but increased the training-validation variance, indicating overfitting and poor generalization to unseen data.

Despite these challenges, the results provide a robust foundation for future research in this domain. For Example, future research should focus on increasing the dataset size to understand human taste perception better. Transfer learning and cross-validation approaches should be employed to mitigate underfitting and overfitting, and advanced regularization techniques should be implemented to achieve optimal generalization. Integrating early stopping with adaptive criteria and investigating data augmentation techniques tailored to sEMG signals may enhance the model's resilience and suitability in various real-world scenarios. Likewise, exploring other designs, such as transformers or hybrid models, and including additional modalities, such as EEG signals, could enhance the system's robustness and accuracy. Furthermore, the real-time applications of GTMH-TasteNet, particularly in virtual reality and assistive healthcare, represent promising avenues for future research.

6 Conclusion

This study presents an advanced deep learning model, GTMH-TasteNet, designed to categorize six taste states, including physiological states, utilizing sEMG data. The framework's architecture, which includes temporal convolutional layers and multi-head attention methods along with GTCE, achieved an 85.2% classification accuracy. The integration of these algorithms allows GTMH-TasteNet to extract features from localized and long-range temporal dependencies, which is essential for nuanced comprehension.

The preprocessing pipeline, which includes detrending, high-pass filtering, and notch filtering, ensures that the input signals contain no noise and are of superior quality. Meanwhile, data augmentation enhances model generalizability. The integration of these elements, along with suitable model configurations, yields strong performance across diverse training and testing contexts. The ablation study demonstrates the model's adaptability and scalability, enhancing its design for improved efficiency and accuracy.

The applications of GTMH-TasteNet extend beyond taste classification, facilitating significant advancements in healthcare, assistive technology, and multisensory interactions for diagnosing taste impairments, developing artificial taste systems, and creating immersive virtual realities. Its versatility enables its application in broader contexts, including human-computer interaction and the examination of human sensory perception. GTMH-TasteNet signifies a cutting-edge advancement in taste recognition via sEMG signals and may serve as a foundation for subsequent research in sensory perception and interface technologies.

Acknowledgment: This work was supported by the Postdoctoral Foundation (Grant no. 6024331025K) and Research Projects of Shenzhen Polytechnic University (Grant no. 6023310034K and 6023310026K).

References

[1] X. Xia, Y. Yang, Y. Shi *et al.*, "Decoding human taste perception by reconstructing and mining temporal-spatial features of taste-related EEGs," vol. 54, no. 5, pp. 3902-3917, 2024.

[2] M. van den Brink, R. C. Havermans, M. Fiocco *et al.*, "A longitudinal evaluation of smell and taste function in children with cancer during and after treatment with chemotherapy," vol. 193, pp. 107174, 2024.

[3] M. F. Asan, G. S. Babu, V. Ajila *et al.*, "Taste disorders in diabetes–An insight," vol. 29, no. 2, pp. 268-272, 2022.

[4] S. Ahmad, A. Sohail, M. A. S. Chishti *et al.*, "How common are taste and smell abnormalities in COVID-19? A systematic review and meta-analysis," vol. 17, no. 2, pp. 174-185, 2022.

[5] A. M. Chao, A. Roy, A. T. Franks *et al.*, "A systematic review of taste differences among people with eating disorders," vol. 22, no. 1, pp. 82-91, 2020.

[6] A. Ullah, F. Zhang, Z. Song *et al.*, "Surface Electromyography-Based Recognition of Electronic Taste Sensations," vol. 14, no. 8, pp. 396, 2024.

[7] R. Anbarasan, D. Gomez Carmona, and R. J. F. E. R. Mahendran, "Human taste-perception: brain computer interface (BCI) and its application as an engineering tool for taste-driven sensory studies," vol. 14, no. 3, pp. 408-434, 2022.

[8] H. Wang, X. Zhan, L. Liu *et al.*, "Unsupervised cross-user adaptation in taste sensation recognition based on surface electromyography," vol. 71, pp. 1-11, 2022.

[9] K. J. T. A. j. o. c. n. Kurihara, "Glutamate: from discovery as a food flavor to role as a basic taste (umami)," vol. 90, no. 3, pp. 719S-722S, 2009.

[10] J. Chandrashekar, M. A. Hoon, N. J. Ryba *et al.*, "The receptors and cells for mammalian taste," vol. 444, no. 7117, pp. 288-294, 2006.

[11] D. A. Yarmolinsky, C. S. Zuker, and N. J. J. C. Ryba, "Common sense about taste: from mammals to insects," vol. 139, no. 2, pp. 234-244, 2009.

[12] J. R. Gibbons and N. M. Sadiq, "Neuroanatomy, neural taste pathway," 2019.

[13] A. Komiyama, T. Kobayakawa, H. Toda *et al.*, "A high-concentration NaCl solution does not stimulate the human trigeminal nerve at the tip of the tongue," *Acta Otolaryngol,* vol. 127, no. 7, pp. 754-759, 2007.

[14] A. Ullah, Y. Liu, Y. Wang *et al.*, "Gender Differences in Taste Sensations Based on Frequency Analysis of Surface Electromyography," vol. 130, no. 3, pp. 938-957, 2023.

[15] Y. Miyaoka, I. Ashida, D. Inagaki *et al.*, "Differentiation of activity patterns in the suprahyoid muscles during swallowing of foods with five taste qualities," *Journal of Sensory Studies,* vol. 20, no. 6, pp. 473-483, 2005.

[16] R. Merletti, and P. J. Parker, *Electromyography: physiology, engineering, and non-invasive applications*: John Wiley & Sons, 2004.

[17] M. Cifrek, V. Medved, S. Tonković *et al.*, "Surface EMG based muscle fatigue evaluation in biomechanics," *Clinical biomechanics,* vol. 24, no. 4, pp. 327-340, 2009.

[18] C. Dawes, A. L. Pedersen, A. Villa *et al.*, "The functions of human saliva: A review sponsored by the World Workshop on Oral Medicine VI," *Archives of oral biology,* vol. 60, no. 6, pp. 863-874, 2015.

[19] A. Del Vecchio, F. Negro, F. Felici *et al.*, "Associations between motor unit action potential parameters and surface EMG features," *Journal of Applied Physiology,* vol. 123, no. 4, pp. 835-843, 2017.

[20] K. Mukai, and K. Watanabe, "Intentional facial expression variation per taste preference for beverages."

[21] P. Kaczmarek, T. Mańkowski, and J. Tomczyński, "putEMG—a surface electromyography hand gesture recognition dataset," *Sensors,* vol. 19, no. 16, pp. 3548, 2019.

[22] N. Nasri, S. Orts-Escolano, and M. Cazorla, "An semg-controlled 3d game for rehabilitation therapies: Real-time time hand gesture recognition using deep learning techniques," *Sensors,* vol. 20, no. 22, pp. 6451, 2020.

[23] Y. Wang, M. Zhang, R. Wu *et al.*, "Silent speech decoding using spectrogram features based on neuromuscular activities," *Brain sciences,* vol. 10, no. 7, pp. 442, 2020.

[24] N. Nasri, S. Orts-Escolano, F. Gomez-Donoso *et al.*, "Inferring static hand poses from a low-cost non-intrusive sEMG sensor," *Sensors,* vol. 19, no. 2, pp. 371, 2019.

[25] Y. Miura, Y. Morita, H. Koizumi *et al.*, "Effects of taste solutions, carbonation, and cold stimulus on the power frequency content of swallowing submental surface electromyography," *Chemical senses,* vol. 34, no. 4, pp. 325-331, 2009.

[26] W. Sato, K. Minemoto, A. Ikegami *et al.*, "Facial EMG correlates of subjective hedonic responses during food consumption," *Nutrients,* vol. 12, no. 4, pp. 1174, 2020.

[27] Y. Wang, H. Wang, H. Li *et al.*, "Qualitative recognition of primary taste sensation based on surface electromyography," vol. 21, no. 15, pp. 4994, 2021.

[28] R. Ueda, K. Umetani, F. Konishi *et al.*, "Characterization of palatability and ease of deglutition of the five basic tastes by partial least squares regression analysis using myoelectric potential parameters of the submental muscle," vol. 28, no. 1, pp. 75-86, 2022.

[29] F. Rahimi, M. A. Badamchizadeh, S. Ghaemi *et al.*, "Simultaneous Estimation of Digit Tip Forces and Hand Postures in a Simulated Real-Life Condition With High-Density Electromyography and Deep Learning," *IEEE Journal of Biomedical and Health Informatics,* vol. 28, no. 10, pp. 5708-5717, 2024.

[30] R. Gupta, V. Kavanal, and A. Joshi, "Deep Learning Models for Recognition of Hand Gestures in Basketball Sports." pp. 1-6.

[31] S. Ma, J. Zhang, C. Shi *et al.*, "Physics-Informed Deep Learning for Muscle Force Prediction With Unlabeled sEMG Signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 32, pp. 1246-1256, 2024.

[32] M. Zanghieri, P. M. Rapa, M. Orlandi *et al.*, "sEMG-Driven Hand Dynamics Estimation With Incremental Online Learning on a Parallel Ultra-Low-Power Microcontroller," *IEEE Transactions on Biomedical Circuits and Systems,* vol. 18, no. 4, pp. 810-820, 2024.

[33] A. Fatayer, W. Gao, Y. J. I. J. o. B. Fu *et al.*, "sEMG-based gesture recognition using deep learning from noisy labels," vol. 26, no. 9, pp. 4462-4473, 2022.

[34] Y. Zhou, C. Chen, M. Cheng *et al.*, "Comparison of machine learning methods in sEMG signal processing for shoulder motion recognition," *Biomedical Signal Processing and Control,* vol. 68, pp. 102577, 2021/07/01/, 2021.

[35] J. He, X. Niu, P. Zhao *et al.*, "From Forearm to Wrist: Deep Learning for Surface Electromyography-Based Gesture Recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 32, pp. 102-111, 2024.

[36] A. Vijayvargiya, B. Singh, R. Kumar *et al.*, "Hybrid Deep Learning Approaches for sEMG Signal-Based Lower Limb Activity Recognition," vol. 2022, no. 1, pp. 3321810, 2022.

[37] S. Mekruksavanich, W. Phaphan, and A. Jitpattanakul, "Deep Learning Networks for Human Knee Abnormality Detection Based on Surface EMG Signals." pp. 160-163.

[38] C. J. De Luca, L. D. Gilmore, M. Kuznetsov *et al.*, "Filtering the surface EMG signal: Movement artifact and baseline noise contamination," vol. 43, no. 8, pp. 1573-1579, 2010.

[39] M. Boyer, L. Bouyer, J.-S. Roy *et al.*, "Reducing noise, artifacts and interference in single-channel EMG signals: A review," vol. 23, no. 6, pp. 2927, 2023.

[40] H. Wang, D. Lu, L. Liu *et al.*, "Quantitatively recognizing stimuli intensity of primary taste

based on surface electromyography," vol. 21, no. 21, pp. 6965, 2021.

[41] N. A. Wahab, R. D. Jones, M.-L. J. P. Huckabee *et al.*, "Effects of olfactory and gustatory stimuli on the biomechanics of swallowing," vol. 102, no. 5, pp. 485-490, 2011.

[42] E. Oganesiants, V. Sarkisyan, A. Bilyalova *et al.*, "The Role of Bitter-Tasting Substances in Salivation and Swallowing: Results of the Pilot Study," vol. 14, no. 2, pp. 210, 2025.

[43] H. Gao, S. Zhao, H. Li *et al.*, "Feature Extraction and Feature Combination for Basic Taste Sensation Recognition Based on Facial Electromyography." pp. 5490-5495.

Asif Ullah received his PhD in Control Science and Engineering from Zhejiang University, Hangzhou, China, in 2023. He is a Postdoctoral Fellow at Shenzhen Polytechnic University, Shenzhen, China. He has contributed to various research projects, including E-Taste design and a remote-controlled wheelchair model. His research interests include intelligent manufacturing, data analysis, PCB designing, healthcare, Deep learning models, and control systems applications.

Zhendong Song (Member, IEEE) received his PhD from the Harbin Institute of Technology, Harbin, China, in 2017. He is an Assistant Professor at Shenzhen Polytechnic University, Shenzhen, China. His research interests mainly include the industry Internet, intelligent manufacturing based on big data, and the Internet of Things.

Waqar Riaz is a postdoctoral researcher at the Shenzhen Polytechnic University and the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. He received his PhD in Communication and Information Engineering from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2022. His research focuses on intelligent systems at the intersection of healthcare, computer vision, and intelligent manufacturing. His work has been supported by competitive research grants and published in high-impact journals.

You Wang received the Ph.D. degree in biomedical engineering from Zhejiang University, Hangzhou, China, in 2007. He is an Associate Professor with the College of Control Science and Engineering, Zhejiang University. His research interests include biosensors, gas sensors, and their applications in industry and medical devices.