

MACHINE LEARNING-BASED CLASSIFICATION OF SMILE AND WRY SMILE TYPES IN YOUNG JAPANESE ADULTS USING NON-CONTACT FACIAL SURFACE DATA

SOU KAMADA AND YOICHI KAGEYAMA*

Graduate School of Engineering Science
Akita University

1-1 Tegatagakuen-machi, Akita-shi, Akita 010-8502, Japan
sou1107.bell@gmail.com; *Corresponding author: kageyama@ie.akita-u.ac.jp

Received March 2026; revised July 2026

ABSTRACT. *Advances in information devices and the impacts of COVID-19 have accelerated research on emotion estimation using non-contact information. However, voluntary responses, such as facial expressions and vocal cues, can be intentionally controlled and may not reflect an individual's true emotions. Consequently, involuntary physiological signals, such as facial skin temperature, have attracted attention. Previous studies have reported skin temperature changes in the nasal region and both cheeks during joy; nonetheless, emotion recognition methods based on such signals remain underexplored when voluntary responses are inconsistent with an individual's true psychological state. This study aims to develop and evaluate a new method for classifying smile types using non-contact physiological facial surface information, addressing the challenge of distinguishing genuine from non-genuine emotional states based on visible facial expressions. The proposed method focuses on non-contact facial surface data obtained from 17 Japanese university students in their 20s, using synchronized thermal and visible-light video recordings during emotion-eliciting tasks. A method is proposed for identifying expressed emotional states when facial expressions vary from the actual psychological states. Specifically, non-contact facial surface information is extracted as features and used as inputs to a LightGBM model to classify smile, wry smile, and non-emotional states. The results show that the proposed method outperforms comparative approaches that classify smile and forced smile using voluntary features, such as facial expressions and vocal tone. It achieves a macro-average F1-score of 0.729, outperforming all comparative methods.*

Keywords: Emotional arousal estimation, Machine learning, Smile, Skin temperature, Non-contact information, Japanese people

1. Introduction. In recent years, advances in information devices and the impact of COVID-19 have spurred active research into emotion detection and affect estimation using non-contact information. Among various emotional states, positive emotions, especially joy, are recognized as essential determinants of subjective well-being [1]. Psychological factors play an important role in quality of life (QoL), and experiencing positive emotions is associated with improved QoL and better mental health [2,3]. These findings indicate that objectively determining whether joy has been elicited is important for QoL assessment and promoting well-being. However, a smile does not necessarily reflect a positive emotional state; depending on interpersonal considerations and social context, it may instead convey underlying negative emotions. If emotion estimation does not account for such discrepancies between expressed facial cues and actual psychological states, communication may be misinterpreted, potentially leading to incorrect assessments of QoL

and well-being. Therefore, distinguishing between smiles that reflect different underlying psychological states is essential for accurate QoL evaluation.

Facial expression recognition (FER) using visible-light cameras is a widely adopted method in emotion estimation research owing to the ease of collecting non-contact facial data [4,5]. Studies on FER with visible images often use Ekman's basic emotion classification (joy, sadness, anger, surprise, fear, and disgust), classifying emotions based on facial expression patterns for each category [6,7]. As regards the relationship between facial expressions and internal emotional states, studies have sought to distinguish between spontaneous smiles (Duchenne smiles) and non-spontaneous smiles (non-Duchenne smiles), primarily based on differences in periocular muscle activity and facial features [8]. Furthermore, research on multimodal emotion recognition has increasingly integrated complementary sources of information such as facial expressions, speech, and physiological signals [9]. However, many of these approaches rely on the use of multiple sensors and contact-based measurements; scenarios wherein the expressed facial cues and internal states are incongruent remain underexplored, particularly in non-contact settings.

Emotion estimation based on visible images is affected by external factors, including lighting conditions, gender differences, and camera angles [10]. Additionally, naturally expressed facial movements linked to emotions do not always exhibit consistent patterns for the same emotion, which reduces their reliability for distinguishing specific emotions. Facial expressions and vocal tone can be voluntarily controlled, highlighting the importance of involuntary physiological changes such as facial skin temperature and electrodermal activity [11].

Related studies have found that when joy is elicited, skin temperature changes occur in the nasal region and both cheeks [12]. Our previous study showed that during e-sports gameplay, an increase in the temperature of one cheek is often accompanied by a decrease in the temperature of the other. Asymmetry in temperature responses between the left and right cheeks was characteristic of non-emotional arousal and may serve as a useful marker for differentiating non-emotional arousal from emotional arousal [13]. Furthermore, another method employing skin temperature and eye-opening degree achieved higher accuracy in emotional estimation than methods based on facial expressions [14]. These results indicate that physiological information independent of external factors, such as facial expressions, may be a useful indicator for emotion estimation.

However, many previous studies have focused on either binary estimation of emotion elicitation or on the classification of emotional states as positive, negative, or non-elicited. As a result, methods for distinguishing emotions in situations wherein facial expressions and actual psychological states are inconsistent (e.g., genuine laughter versus forced laughter) have not been sufficiently studied [14,15].

This study aims to develop a method for classifying smile types expressed by individuals when their facial expressions are incongruent with their actual psychological states. We examine a smile classification method using a light gradient boosting machine (Light-GBM) model that focuses on the nasal region, left and right cheeks, and left and right eye regions as regions of interest (ROIs). The method uses physiological, non-contact facial surface information, specifically for cases where facial expressions and psychological states are inconsistent. The approach is novel in that it uses facial skin temperature and body movement features, which individuals cannot consciously control, to improve the differentiation between genuine and non-genuine emotional states. The novelty of this study lies not in individual techniques but in the integration of methods and features. Specifically, we present 1) a hierarchical classification approach focusing on subtle variations in smile, 2) a feature extraction method that integrates short-term skin temperature changes with body motion within a limited time window, and 3) an analysis of temporal delays in

physiological responses. Consequently, the proposed framework enables the non-contact differentiation of the smile types using physiological signals.

The remainder of this paper is organized as follows. Section 2 describes the dataset used in this study. Section 3 explains the face detection method applied to the acquired data. Section 4 details the feature extraction method for the segments detected in Section 3. Section 5 presents the proposed smile classification method, and Section 6 discusses the experimental results and analysis. Section 7 concludes the paper and outlines future work.

2. Data Acquisition. Data collection experiments for this study were conducted in November 2024 and September 2025. A thermal infrared thermography camera (R500EW-S, Nippon Avionics Co., Ltd. [16]) and a 4K video camera (HC-VX2M, Panasonic Corporation [17]) recorded participants' facial expressions as they watched videos intended to elicit smiles and wry smiles. Recordings included thermal video images (640×480 pixels) and visible video images (1920×1080 pixels), both at 30 fps. Figure 1 shows the data acquisition environment.

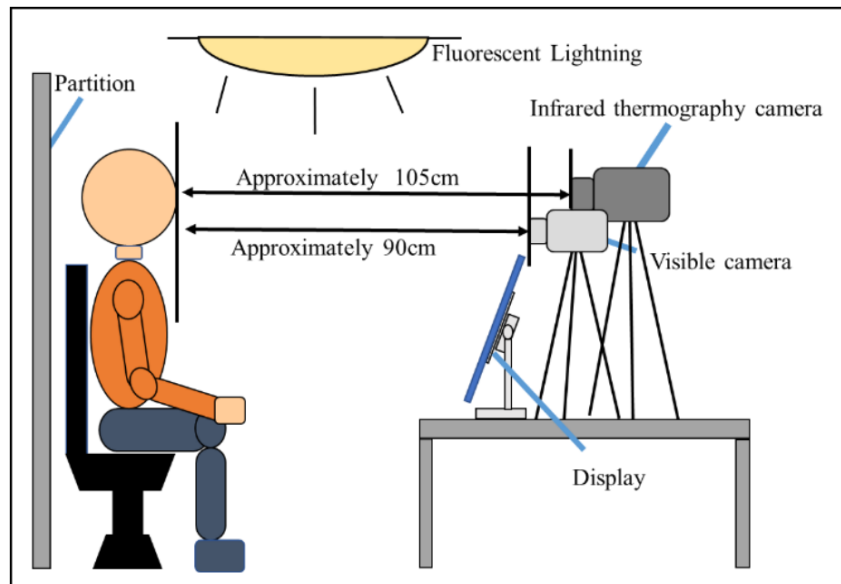


FIGURE 1. Experimental setup for data acquisition

To provide a baseline for comparison with emotion arousals, a 10-second resting state (steady arousal) was recorded before and after each video viewing session. The study included 17 university students as participants (A-Q: 11 males and 6 females), in their twenties, who did not wear glasses. Data were collected indoors under fluorescent lighting.

Because the recording ranges of the thermal and visible cameras differed, a partition was placed behind the participants, and four markers were positioned at the corners of the thermal image acquisition area to align the recording regions.

The experimental conditions for the thermal and visible video recordings at the time of data acquisition are summarized below.

- Indoor temperature ranged from 19.3°C to 24.0°C in November 2024 and from 24.8°C to 26.7°C in September 2025.
- Relative humidity was 34.5% to 41.8% in November 2024 and 58.2% to 60.3% in September 2025.
- Illuminance above the participants was 825 to 875 lx in November 2024 and 726 to 769 lx in September 2025.

- Illuminance in front of the participants was 714 to 766 lx in November 2024 and 628 to 712 lx in September 2025.

The data used in this study were acquired in accordance with ethical regulations for human research at Akita University, Japan. Data were collected after obtaining informed consent from all participants.

Facial expressions of smile and wry smile were recorded as visible-light and thermal video images while participants watched video clips (approximately 10 min each) designed to elicit these responses.

Reisenzein and Junge [18] reported that emotions can vary in intensity (e.g., strong, moderate, or weak) and that comparing emotions allows for stable intra-individual evaluation of emotional intensity. According to Amvadar et al. [19], a smile can result from positive emotions, such as joy, and negative emotions, such as embarrassment and tension, and that observers perceive certain smiles, despite the upward movement of the mouth corners, as different from positive smiles.

In this study, participants watched recorded footage of themselves and completed a questionnaire indicating whether they were smiling in each segment. The type and timing of emotion arousal were recorded based on this self-report questionnaire. Smile was categorized into four types according to the questionnaire based on the duration of emotion arousal segments, with definitions provided in Table 1.

TABLE 1. Definition of smile types

Type of smile	Description
Strong smile	The corners of the mouth are raised widely, and the person laughs audibly.
Moderate smile	A smile that is intermediate in intensity between a strong smile and a weak smile.
Weak smile	The person smiles with the mouth closed or with the corners of the mouth slightly raised.
Wry smile	The corners of the mouth are raised, but the person appears confused or perplexed.

Only “strong smile” was considered as “smile” for subsequent analysis. Based on a preliminary experiment, wry smile was defined as laughter accompanied by feelings of “confusion/embarrassment”, as these emotions were strongly identified in the questionnaire for segments labeled as wry smile. Segments that did not correspond to either smile or wry smile were classified as non-emotion arousal.

3. Face Detection Method for Facial Thermal Images. To analyze temperature changes in specific facial ROIs (defined in Section 4), it is necessary to examine temporal temperature variations within each area. The face detection process includes three main steps: preprocessing visible and thermal images, temporally aligning frames between modalities, and extracting facial landmarks for ROI analysis. This pipeline ensures accurate mapping of physiological features across synchronized video data. The flow of the face detection method is shown in Figure 2.

3.1. Preprocessing for visible images and thermal images. Face detection required the use of combined images. To convert temperature data from the thermal video into image format, normalization was applied so that the temperature range of 29.0°C-37.0°C matched grayscale intensity values from 0 to 255. This process generated thermal grayscale images.

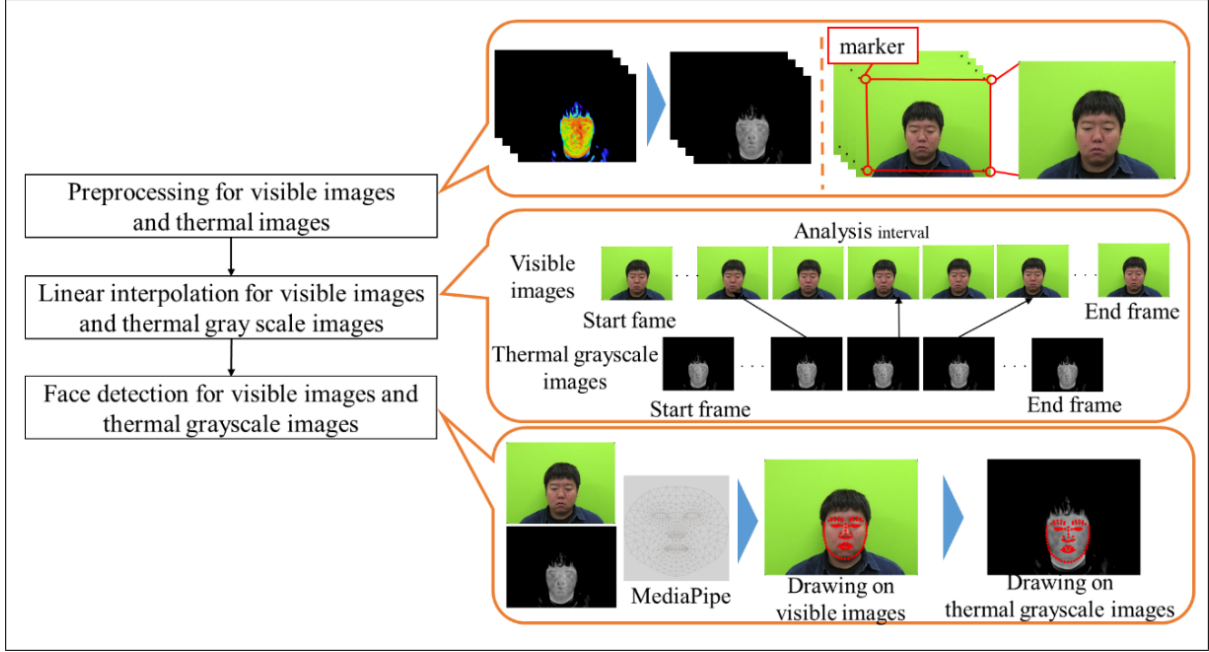


FIGURE 2. Workflow of the proposed estimation method

3.2. Linear interpolation for visible and thermal grayscale images. Because the total number of frames between the visible-light video and the thermal grayscale video differed, linear interpolation based on Equation (1) was performed using the starting and ending frames of both videos. This process calculated the frame in the thermal grayscale video that temporally corresponded to each frame within the temperature analysis interval of the visible video.

$$Temp_{out} = \frac{Temp_s - Temp_e}{Web_e - Web_s}(Web_{in} - Web_s) + Temp_s \quad (1)$$

Here, $Temp_s$ denotes the starting frame number of the thermal grayscale video, $Temp_e$ denotes the ending frame number of the thermal grayscale video, Web_s denotes the starting frame number of the visible video, and Web_e denotes the ending frame number of the visible video. Web_{in} represents an arbitrary frame number in the visible video, and $Temp_{out}$ represents the frame number in the thermal grayscale video that temporally corresponds to Web_{in} .

3.3. Face detection for visible and thermal grayscale images. First, 478 facial landmark coordinates were extracted from the visible video using the face detection library MediaPipe [20]. Next, these coordinates were superimposed onto the temporally corresponding thermal grayscale images within the analysis interval.

4. Feature Extraction in ROIs. This section describes the sequential process for extracting physiological features from defined facial regions. The steps include selecting ROIs, calculating body movement, measuring skin temperature differences, determining time delays in physiological changes, and assessing short-term temperature change.

4.1. Setting the ROIs. For the thermal grayscale images with facial landmarks detected using MediaPipe, five ROIs were defined: the nasal region, left and right cheek regions, and left and right eye regions. Figure 3 shows an example of the ROI settings with blue rectangles. The ROI sizes were 10×30 pixels for the nasal region, 30×30 pixels for the cheek regions, and 30×20 pixels for the eye regions.

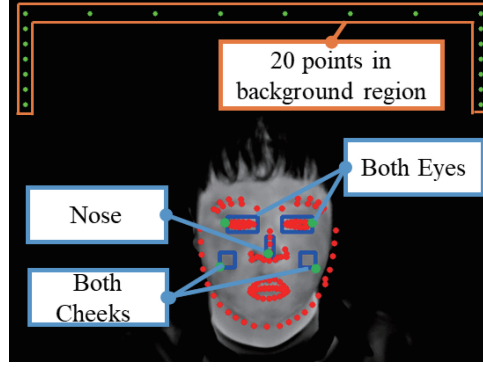


FIGURE 3. Regions of interest (ROIs) for feature extraction

4.2. Calculation of body movement. First, the x - and y -coordinates of the feature points (indicated by the green dots) in Figure 3 were calculated from the thermal grayscale images obtained using MediaPipe. Next, as defined by Equations (2) and (3), the coordinate differences between the frame of interest and the frame acquired 30 frames earlier were calculated from the time-series coordinates of each feature point and defined as body movement (BM). In this study, BMs in the x - and y -directions are denoted as DX and DY , respectively.

$$DX = M_x(n) - M_x(n - 30), \quad (n \geq 31) \quad (2)$$

$$DY = M_y(n) - M_y(n - 30), \quad (n \geq 31) \quad (3)$$

Here, M_x and M_y represent the x - and y -coordinates of the feature points, and n denotes the frame of interest.

4.3. Calculation of skin temperature difference. To account for the influence of room temperature, the temperature difference between the skin temperature of each ROI and the average temperature of 20 background points (indicated by the red rectangle in Figure 3) was calculated. The average temperature within each ROI, followed by the average temperature of the 20 background points (background skin temperature), was computed from the thermal grayscale images. The temperature difference between the average temperature of each ROI and the background skin temperature was then defined as the skin temperature difference (STD). Smoothing was applied to the calculated STD using a moving filter that included the frame of interest and one frame before and after.

4.4. Calculation of time delay in physiological changes. Yamada and Kageyama [12] used the temperature difference between the frame of interest and the frame acquired 30 frames earlier, suggesting that different types of joy could be distinguished by analyzing frames near the onset of emotion. However, physiological changes such as skin temperature and heart rate have been reported to show a temporal delay between the onset of emotion and the corresponding physiological response [21-23]. In this study, we focused on the nasal region, which is more sensitive to autonomic nervous system activity than other facial regions. We defined latency as the interval between emotional onset and the first point at which a sustained change occurred. Using this definition, we calculated latency for all 17 participants (A-Q) to extract features based on short-term temperature changes.

4.4.1. Examination of time delay in physiological changes. First, the temperature difference during emotional arousal was calculated relative to the average temperature recorded during the 3 seconds before emotion onset and normalized by the corresponding standard deviation. Next, we identified the first maximum absolute value in the time-series data of

the temperature difference as the peak. We then examined the frames preceding the peak to identify the frame where the temperature difference exceeded a predefined threshold. Finally, we determined the difference in frame numbers.

The latency was calculated as the time difference between the frame exceeding the threshold and the frame corresponding to the onset of emotion arousal.

In this study, a threshold of 0.05°C was used for the temperature difference, matching the temperature resolution of the thermal infrared camera and remaining independent of measurement error.

Table 2 presents the median latency for each participant. For 13 participants, latency was 10 frames or longer, with a mean of 11.0 frames and a median of 11.6 frames. These results show an approximate delay of 0.37 seconds between the onset of emotion arousal and the actual temperature change.

TABLE 2. Results of latency estimation for physiological changes

		Latency (frames)
Participant	A	9.5
	B	12.0
	C	10.0
	D	7.0
	E	10.5
	F	12.0
	G	11.0
	H	16.0
	I	12.5
	J	16.0
	K	8.0
	L	11.0
	M	11.0
	N	14.0
	O	12.5
	P	9.0
	Q	16.0
Average		11.0
Median		11.6

Note: Participants with a latency of 11 frames or greater are highlighted in bold in the original table.

4.4.2. *Calculation of temperature change.* Based on the 11-frame delay time calculated in Section 4.4.1, short-term temperature changes were analyzed by computing the temperature difference between the frame of interest and the frame acquired 11 frames earlier from the time-series data of the STD, as shown in Equation (4). This value was defined as the temperature change (TC). Figure 4 presents an example of the calculated TC.

$$\Delta t = T_n - T_{n-11} \quad (n \geq 12) \quad (4)$$

where Δt represents the TC, T represents the STD at the frame of interest, and n is the number of the frames of interest.

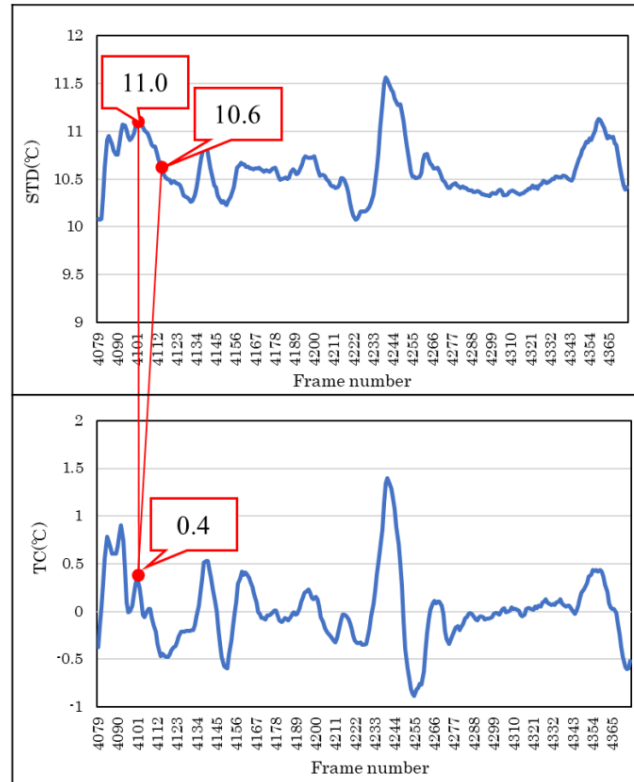


FIGURE 4. Example of calculated TC

5. Classification Method for Smile Types.

5.1. Experimental overview. According to the cognitive appraisal theory of emotion [24], emotions result from an individual's appraisal of the meaning of the event rather than from the event itself. In this process, primary appraisal determines whether the event is significant to the individual, while secondary appraisal assesses the individual's ability to cope with the event, which together define the type and intensity of emotion.

Based on this theory, this paper examines a hierarchical classification method that uses features such as skin temperature and body movement, recorded during video viewing. In this approach, smile segments were labeled as 2, wry smile segments as 1, and non-emotion arousal segments as 0. The skin temperature and body movement features were input into a LightGBM model [25] to classify smile segments. The classification process is illustrated in Figure 5.

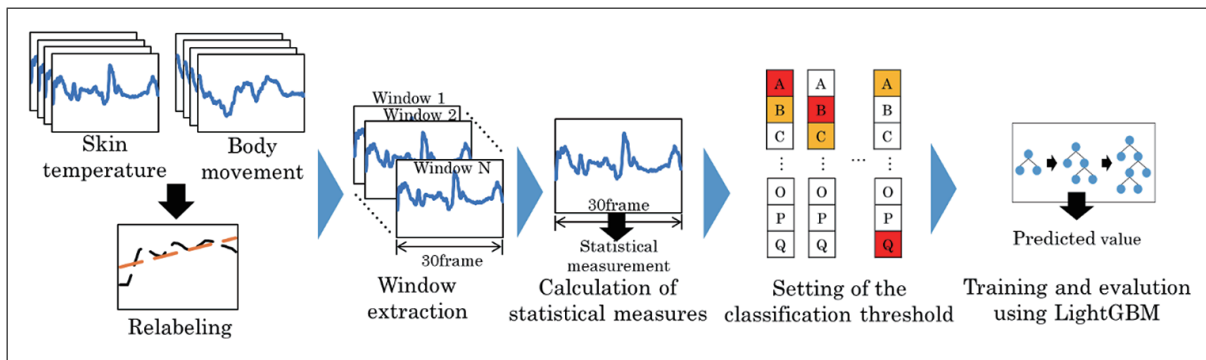


FIGURE 5. Hierarchical classification process for smile types

5.2. Labeling of time-series data. In this study, smile segments and non-emotion arousal segments were defined using questionnaires completed by participants after viewing the videos. Because these responses are subjective, the defined segments may not align precisely with the actual timing of emotional arousal. Accordingly, changes in skin temperature were observed in some segments that the participants identified as non-emotional arousal, indicating that weak emotional responses may have occurred. To address this, the proposed method included a relabeling process to reduce such noise.

In a previous study [26], linear approximation was applied to skin temperature data related to smile-associated emotions, and the slope of the resulting regression line was calculated. The results showed that temperature change tendencies varied depending on the type of smile and the ROI, suggesting that the temporal rate of change in skin temperature may help distinguish between types of smiles. Based on this, a relabeling procedure focusing on the slope of the linear approximation of the skin temperature data was applied to all non-emotional arousal segments identified in the questionnaire data, only in the training data.

Specifically, a linear approximation, as described in Equations (5) and (6), was applied to a 30-frame window (15 frames before and 15 frames after the frame of interest). The window was shifted one frame at a time, and the slope of the regression line was calculated for each position. If the mean slope in the non-emotional arousal segment was greater than the mean slope in the steady-state segment, the label of the non-emotional arousal segment was reassigned to wry smile. If the mean slope in the non-emotion arousal segment was greater than the mean slope in the smile segment, the label was reassigned to smile.

$$y - \bar{y} = a(x - \bar{x}) \quad (5)$$

$$a = \frac{S_{xy}}{S_x^2} \quad (6)$$

Here, y is the estimated value of the skin temperature difference in the target segment; $\bar{y}$ is the mean skin temperature difference in the target segment; x is the frame number in the target segment; $\bar{x}$ is the mean frame number in the target segment; S_{xy} is the covariance between x and y ; and S_x squared is the variance of x .

5.3. Feature selection using statistical measures. From the relabeled time-series data, 30-frame windows were extracted from segments annotated as smile or wry smile. To reduce short-term noise, a sliding window with a shift of 5 frames was applied. For non-emotional arousal segments, windows were randomly sampled so that the number of extracted windows matched that of the larger of the smile or wry smile classes.

For each window, the statistical measures listed in Table 3 were calculated using the time-series data of STD, TC, and BM, which were served as physiological features.

During training, feature importance was calculated for each window and then averaged across participants to obtain participant-independent feature importance. In 20 iterations,

TABLE 3. Overview of extracted physiological and statistical features

	Feature name	Description
Magnitude of change	Delta	Mean of the absolute gradient (i.e., average absolute change) within the window
Magnitude of value	Max	Maximum value within the window
	Min	Minimum value within the window
	RMS	Root mean square of the amplitude within the window

the lowest 10 percent of features by importance were removed at each step. The macro-average F1-score (Macro F1) was calculated at each iteration, and the iteration with the highest Macro F1 was selected. The feature set from this iteration was used as the optimal set for model training.

5.4. Classification using LightGBM. The constructed datasets for each label were used as input to train a LightGBM model. LightGBM is a decision tree-based model that learns both individual features and nonlinear relationships among multiple features. In this study, we conducted a hierarchical classification process.

In the first stage, the model distinguished between non-emotional arousal and joy-related emotional arousal. In the second stage, data classified as emotional arousal were further divided into smile and wry smile. This hierarchical approach first distinguished emotional from non-emotional states, which reduced class imbalance and improved the accuracy of the subsequent classification between smile and wry smile. Data from 16 participants were used for training, and data from the remaining participant were used as test data for training and evaluation. In this study, the F1-score and Macro F1, defined by Equations (7) and (8), were used as evaluation metrics. TP , FP , and FN represent the number of true positives, false positives, and false negatives, respectively.

$$F1\text{-score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (7)$$

$$Macro\ F1 = \frac{\sum_{i=1}^3 F1\text{-score}(class(i))}{3} \quad (8)$$

To determine the prediction probability threshold, subject-wise cross-validation was performed by dividing the data into training, validation, and test datasets. Specifically, data from one subject served as the test set, data from a second subject served as the validation set, and data from the remaining subjects served as the training set. This process was repeated, sequentially changing the combination of data splits, and the prediction probability threshold was selected based on performance on the validation dataset.

To assess the effectiveness of the proposed method for classifying the types of smile, results were compared with previous studies that classified smile and wry smile using features that can be voluntarily controlled by humans, such as facial expressions and speech [27,28]. Evaluation results were calculated as the Macro F1 based on the confusion matrices reported in these studies. Table 4 presents an overview of the proposed method and the comparison methods.

TABLE 4. Summary of the proposed and comparison methods

Author	Model	Dataset	Features
Proposed method	LightGBM	Self-collected	Skin temperature Body movement
	Atten-BiLSTM		
	RNN		
Nguyen et al.	VGG-face + BiLSTM	MATHNOB	Facial expressions
	VGG-face + BiLSTM		
	VGG-face + BiLSTM		
Inoue et al.	RNN	Self-collected	Speech

Additionally, machine learning models described in the previous studies were trained and evaluated under the same conditions as the proposed method, and the Macro F1 was calculated. As the VGG-Face model used by Ke et al. [25] is designed for visible video data and could not be applied to the data used in the proposed method, it was excluded

from the comparison. An overview of the comparison methods is shown in the table. In this study, an F1-score of 0.700 or higher was set as the criterion for good classification performance.

6. Experimental Results and Discussion.

6.1. Experimental results. Table 5 presents the class-wise F1-scores and Macro F1 at each stage. In the first stage, F1-scores above 0.900 were achieved for non-emotional arousal in 17 participants and for emotional arousal in 16 participants. In the second stage, although some participants exhibited F1-scores below 0.700, F1-scores above 0.700 were achieved for wry smile in 10 participants and for smile in 14 participants. These results suggest that the proposed hierarchical framework effectively differentiates emotional from non-emotional arousal, while the relatively moderate F1-scores for smile and wry smile in the second stage reflect the greater difficulty in distinguishing these subtle expressions. These findings provide a quantitative basis for further analysis of feature effectiveness.

TABLE 5. Classification results for each stage of the hierarchical model

		First stage			Second stage		
		Neutral	Emotion	Macro F1	Wry smile	Smile	Macro F1
Participant	A	0.988	0.990	0.989	0.844	0.865	0.854
	B	0.900	0.917	0.908	0.970	0.973	0.972
	C	0.996	0.996	0.996	0.837	0.732	0.785
	D	0.952	0.952	0.952	0.656	0.720	0.688
	E	0.988	0.990	0.989	0.586	0.726	0.656
	F	0.948	0.965	0.957	0.959	0.969	0.964
	G	0.988	0.989	0.989	0.657	0.126	0.392
	H	0.985	0.985	0.985	0.787	0.910	0.848
	I	0.974	0.976	0.975	0.867	0.921	0.894
	J	0.972	0.978	0.975	0.850	0.709	0.779
	K	0.985	0.986	0.986	0.449	0.853	0.651
	L	0.984	0.985	0.984	0.187	0.790	0.489
	M	0.939	0.933	0.936	0.717	0.930	0.823
	N	0.739	0.688	0.713	0.293	0.525	0.409
	O	0.952	0.961	0.957	0.732	0.749	0.741
	P	0.995	0.996	0.996	0.505	0.856	0.680
	Q	0.785	0.794	0.790	0.882	0.667	0.775
Average		0.945	0.946	0.946	0.693	0.766	0.729

Bold: Subjects with F1-score and Macro F1 greater than 0.700

6.2. Discussion. The selected features were organized into two groups: ROI and statistical measures, and physiological features and statistical measures, and the results are shown in Tables 6 and 7, respectively. From both perspectives, Delta and RMS were selected for more than half of the ROI and physiological features. This indicates that features representing the magnitude of response relative to the decision threshold are effective for differentiating non-emotional arousal from emotional arousal. Accordingly, the high F1-scores obtained for non-emotional arousal and emotional arousal in the first stage indicate that strong separation of values within each window, such as RMS and Delta, effectively captured differences in smile intensity.

These features represent the magnitude of the waveform amplitude and may not capture the subtle temporal changes associated with facial expressions, such as a wry smile. To

TABLE 6. Selected features by region of interest (ROI) and statistical measure

	Nose	Right cheek	Left cheek	Right eye	Left eye
Delta	○	○		○	
Max	○	○			
Min				○	
RMS	○	○	○		

 : Selected features

TABLE 7. Selected features by physiological feature and statistical measure

	STD	TC	DX	DY
Delta	○	○		
Max	○	○		
Min				○
RMS	○	○	○	

 : Selected features

capture temporal variation, frequency analysis is necessary because it can characterize the speed and periodicity of waveform fluctuations. Subtle facial expressions, such as smile and wry smile, likely differ in movement intensity and temporal characteristics; incorporating frequency-based feature components may improve the differentiation between smile and wry smile.

In addition, feature importance was calculated for each participant with respect to the selected features. Table 8 presents an excerpt of these results. For participants with low scores, the selected features may not necessarily have been effective. For example, in the case of participant L, features around the eyes showed relatively high importance, suggesting that facial expressions were predominantly driven by eye movements. However, as periocular features exhibited large inter-subject variability, they are less likely to serve as reliable discriminative indicators, limiting the ability of the proposed model to sufficiently capture these characteristics. Furthermore, participant G exhibited pronounced left-right facial asymmetry, whereas participant N showed subtle overall feature variations. The potentially large discrepancy between the participant-specific features and those in the training data may have contributed to decreased classification performance. These results suggest that while the proposed method is effective for typical laughter expressions, its performance tends to degrade for individual-specific expression styles.

Furthermore, based on the feature importance values obtained from each participant in the cross-validation, the mean importance, standard deviation, feature ranking, and the number of participants for whom each feature was selected were calculated. The results are presented in Table 9. Focusing on the mean importance, Nose_dy30_Delta (Importance.mean = 11593.493) and RightCheek_ma_Delta (Importance.mean = 9505.549) substantially outperformed the other features, exhibiting values approximately twice those of ranked third or lower (Importance.mean $\leq$ 4619.702). In addition, these two features exhibited the lowest ranking variability among all features, with Rank.std values of 0.529 and 1.054, respectively. The consistency between their high importance and low stable ranking suggests that they are robust features for differentiating between smile and wry smile. These features correspond to changes in the central facial region extending towards

TABLE 8. Feature importance for each participant (an excerpt)

		Nose_dy30_ Delta	RightCheek_ ma_Delta	LeftCheek_ ma_RMS	LeftCheek_ tc11_RMS	LeftEye_ ma_Min	RightEye_ dy30_Delta	RightEye_ ma_Min
Subject	A	12996.42	8328.92	4262.26	5422.94	3331.40	3297.78	5482.51
	B	11764.30	10776.84	5493.51	4548.91	3958.30	3649.17	4824.40
	C	8210.52	7552.21	3733.39	4910.67	1568.83	2661.00	2720.13
	D	13402.24	10211.63	3654.79	5005.71	3713.39	3385.70	4976.42
	E	16150.12	4423.38	6171.11	4854.45	4433.03	2189.82	4376.18
	F	10525.61	10015.87	4284.78	4658.77	3471.54	4260.59	4082.79
	G	8642.42	9646.79	4669.27	3367.43	3816.49	2371.44	4720.71
	H	11782.12	10319.46	4847.31	4512.82	3319.71	2691.06	5117.88
	I	11006.19	9978.80	4897.08	4223.33	3615.85	2396.87	4318.53
	J	13676.42	9651.35	4829.51	3849.86	4670.75	3220.20	3760.43
	K	5307.91	10196.81	3183.98	4361.03	4713.57	2799.21	4910.45
	L	14132.95	9693.70	4039.54	4447.49	2982.14	3266.59	6187.39
	M	12662.94	10032.07	4870.29	3496.47	3442.37	3461.13	5608.28
	N	12286.79	10356.79	5155.84	4440.98	3949.80	3582.08	4543.44
	O	9857.78	9536.71	3733.28	5125.91	2587.57	7757.70	3186.01
	P	11200.33	9998.96	4437.79	2923.75	4000.52	2565.61	4726.23
	Q	13484.33	10874.06	4462.12	4104.37	3850.69	3705.68	4993.17
Ave.		11593.49	9505.55	4513.29	4367.93	3613.29	3368.33	4619.70

TABLE 9. Feature importance of the selected features

Feature	Importance_ mean	Importance_ std	Rank_ mean	Rank_ std	Overall_ rank	Selected- subjects
Nose_dy30_Delta	11593.493	2578.383	1.176	0.529	1	17
RightCheek_ma_Delta	9505.549	1540.364	2.118	1.054	2	17
RightEye_ma_Min	4619.702	854.713	5.824	3.504	3	17
LeftCheek_dx30_Delta	4589.126	720.939	5.824	3.147	4	17
LeftCheek_ma_RMS	4513.286	737.456	5.765	2.728	5	17
LeftCheek_tc11_RMS	4367.934	660.518	6.294	2.995	6	17
RightCheek_ma_Max	4080.919	562.989	7.471	3.243	7	17
Nose_ma_Max	3919.775	684.015	8.647	2.396	8	17
RightCheek_tc11_RMS	3907.575	479.141	8.235	2.412	9	17
LeftEye_ma_Min	3613.290	763.869	10.176	3.610	10	17
LeftCheek_ma_Mean	3476.886	587.357	11.294	3.158	11	17
RightEye_dy30_Delta	3368.331	1264.934	12.412	3.759	12	17
RightCheek_dx30_RMS	3218.305	416.772	12.647	2.206	13	17
LeftCheek_ma_Delta	3102.274	516.698	13.412	2.694	14	17
Nose_ma_RMS	2962.686	621.902	13.353	1.801	15	17
LeftCheek_dy30_RMS	2880.845	1043.258	14.000	4.690	16	17
RightCheek_ma_RMS	2858.807	469.655	14.353	2.572	17	17

the cheeks and are likely to reflect muscle activity associated with smiling, such as contractions of the cheek muscles and the area around the nose. The consistent changes observed in these regions across participants led to their high importance and stable rankings.

In contrast, features such as RightEye_dy30_Delta, LeftEye_ma_Min, RightEye_ma_Min, and LeftCheek_dy30_RMS exhibited Rank_std values exceeding 3.5, indicating substantial variability in feature ranking across participants. These features also exhibited a relatively

large variability in importance, reflected in Importance_std and Importance_mean, compared to other features. Although these features were selected across all participants, their relative importance was not consistent. This variability may be attributed to the fact that features around the eyes, as well as those based on RMS and minimum values, are dependent on expression intensity and individual expression styles. In particular, periorcular changes during smiling vary considerably among individuals, including differences in the degree and timing of eye narrowing; this likely contributed to the observed variability in both importance and ranking.

6.3. Comparison with previous studies. Table 10 summarizes the results from the proposed method and the comparison methods. The proposed method, which uses the LightGBM model, achieved a higher macro-average F1-score than all comparison methods. This indicates that the proposed method is effective in differentiating among smile, wry smile, and non-emotional arousal.

TABLE 10. Comparison of classification performance between the proposed and existing methods

Author	Model	Features	Macro F1
Proposed method	LightGBM	Skin temperature Body movement	0.729
	Atten-BiLSTM		0.668
	RNN		0.675
Nguyen et al.	VGG-face + BiLSTM	Facial expressions	0.475
	VGG-face + BiLSTM		0.560
	VGG-face + BiLSTM		0.520
Inoue et al.	RNN	Speech	0.675

Because this method extracts features using a 30-frame window, the need for the model to learn long-term temporal dependencies is reduced. As a result, the LightGBM model with statistical feature-based inputs achieved sufficient accuracy compared with models used in previous studies. Increasing the number of frames in the window may further improve performance, even when using models from previous studies.

7. Conclusion. In this paper, we proposed a method for distinguishing types of smiles using non-contact facial surface information as features, focusing on Japanese individuals in their twenties. Specifically, facial skin temperature and body movement were used as input features, and cross-validation was conducted to evaluate the ability of the model to differentiate the smile types in participants excluded from the training data. The main findings are summarized as follows.

- 1) Under the experimental conditions of this study, there was a delay of approximately 0.37 seconds between the onset of emotional arousal and the actual temperature change.
- 2) The hierarchical classification method was effective for distinguishing among smile, wry smile, and non-emotional arousal.
- 3) The proposed method achieved higher accuracy than other comparative methods that use features voluntarily controlled by humans, such as facial expressions and voice, for classifying smile and wry smile.

Because the proposed method extracts features using a 30-frame window, the need for the model to learn long-term temporal dependencies is low. However, focusing on short-term features may not capture subtle or longer-term temporal variations in facial expressions, such as a wry smile. The relatively small and homogeneous participant group may have also limited the generalizability of the findings. Future work should focus on

increasing the number of frames in the window and investigating longer-term temporal dependencies. In summary, several issues remain for future work.

- 1) The participants were limited to Japanese university students in their twenties, resulting in a small and homogeneous sample. Therefore, future work should include participants with more diverse age groups and backgrounds.
- 2) Segments that did not correspond to either smile or wry smile were defined as “non-emotional arousal”. However, these segments may still contain subtle emotional responses. Therefore, the definition of this category and the associated labeling methodology should be refined in future studies.
- 3) The evaluation of the proposed method was based on cross-validation using a single, relatively small dataset, which may lead to dependency on a specific experimental setting, limiting its generalizability. Therefore, further validation using external datasets and subject-independent evaluation is warranted.
- 4) Future work should focus on developing models that consider individual differences, as well as exploring methods that enable stable differentiation even for individual-specific expression styles.

Acknowledgment. This work was supported by the JSPS KAKENHI [grant number JP25K15297].

REFERENCES

- [1] E. Diener, Subjective well-being, *Psychological Bulletin*, vol.95, no.3, pp.542-575, 1984.
- [2] The Whoqol Group, The World Health Organization quality of life assessment (WHOQOL): Development and general psychometric properties, *Social Science & Medicine*, vol.46, no.12, pp.1569-1585, 1998.
- [3] B. L. Fredrickson, The role of positive emotions in positive psychology, *American Psychologist*, vol.56, no.3, pp.218-226, 2001.
- [4] X. Fan, M. Jiang, A. R. Shahid and H. Yan, Hierarchical scale convolutional neural network for facial expression recognition, *Cognitive Neurodynamics*, vol.14, pp.847-858, 2022.
- [5] L. Liao, Y. Zhu, B. Zheng, X. Jiang and J. Lin, FERGCN: Facial expression recognition based on graph convolution network, *Machine Vision and Applications*, vol.33, no.40, 2022.
- [6] H. Huo, Y. Yu and Z. Liu, Facial expression recognition based on improved depthwise separable convolutional network, *Multimedia Tools and Applications*, vol.82, pp.18635-18652, 2023.
- [7] P. Ekman and W. Friesen, Constants across cultures in the face and emotion, *Personality and Social Psychology*, vol.17, pp.124-129, 1971.
- [8] P. Ekman and W. V. Friesen, The Duchenne smile: Emotional expression and brain physiology: II, *Journal of Personality and Social Psychology*, vol.58, no.2, pp.342-353, 1990.
- [9] Y. Wu, Q. Mi and T. Gao, A comprehensive review of multimodal emotion recognition: Techniques, challenges, and future directions, *Biomimetics*, vol.10, no.7, 418, 2025.
- [10] S. Li and W. Deng, Deep facial expression recognition: A survey, *IEEE Transactions on Affective Computing*, vol.13, no.3, pp.1195-1215, 2020.
- [11] L. F. Barret, R. Adolphs, S. Marsella, A. M. Martinez and S. D. Pollak, Emotion expressions reconsidered: Challengers to inferring emotion from human facial movements, *Psychological Science in the Public Interest*, vol.20, no.1, pp.1-68, 2019.
- [12] M. Yamada and Y. Kageyama, Temperature analysis of face regions based on degree of emotion of joy, *International Journal of Innovative Computing, Information and Control*, vol.18, no.5, pp.1383-1394, 2022.
- [13] A. Miura, Y. Kageyama, H. Shirai, C. Ishizawa, K. Suehiro, N. Takahashi, H. Saito, T. Kobayashi, F. Watanabe, H. Satake and N. Sato, Estimation of emotional arousal using facial skin temperature and body movement information for older people playing esports, *ICIC Express Letters, Part B: Applications*, vol.15, no.12, pp.1235-1241, 2024.
- [14] R. Tanabe, R. Kikuchi, M. Zou, K. Suehiro, N. Takahashi, H. Saito, T. Kobayashi, H. Satake, N. Sato and Y. Kageyama, Emotion recognition using temporal facial skin temperature and eye-opening degree during digital content viewing for Japanese older adults, *Sensors*, vol.25, no.21, 6545, 2025.

- [15] J. Almeida, L. Vilaca, I. N. Teixeira and P. Viana, Emotion identification in movies through facial expression recognition, *Applied Sciences*, vol.11, no.15, 2021.
- [16] Nippon Avionics Co., Ltd., *R500EX-S Japanese Version Catalog*, <http://www.avio.co.jp/products/infrared/lineup/pdf/catalog-r500exs-jp.pdf>, Accessed on March 4, 2026.
- [17] Panasonic Co., Ltd., *Digital 4K Video Camera HC-VX2M*, <https://panasonic.jp/dvc/c-db/products/HC-VX2M.html>, Accessed on March 4, 2026.
- [18] R. Reisenzein and M. Junge, Measuring the intensity of emotions, *Quantitative Psychology and Measurement*, vol.15, 2024.
- [19] Z. Amvadar, J. F. Cohn and L. L. Reed, All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous, *Journal of Nonverbal Behavior*, vol.33, no.1, pp.17-34, 2009.
- [20] Github, *google-ai-edge/mediapipe*, <https://github.com/google-ai-edge/mediapipe>, Accessed on January 9, 2026.
- [21] E. Salazar-Lopez, E. Dominguez, V. J. Ramos, J. de la Fuente, A. Meins, O. Iborra, G. Galvez, M. A. Roderiguez-Artacho and E. Gomez-Milian, The mental and subjective skin: Emotion, empathy, feelings and thermography, *Consciousness and Cognition*, vol.34, pp.149-162, 2015.
- [22] H. C. Abercrombie, A. S. Chambers, L. Greischar and R. M. Monticelli, Orienting, emotion, and memory; Phasic and tonic variation in heart rate predicts memory for emotional pictures in men, *Neurobiology of Learning and Memory*, vol.90, pp.644-650, 2008.
- [23] V. Kosonogov, L. De Zorzi, J. Honore, E. S Martinez-Velazquez, J. Nandrino, J. M Martinez-Selva and H. Sequeira, Facial thermal variations: A new marker of emotional arousal, *PLoS ONE*, vol.12, no.9, Article no.sE0183592, 2017.
- [24] C. A. Smith and R. S. Lazarus, Emotion and adaptation, *Handbook of Personality Theory and Research*, no.21, pp.609-637, 1990.
- [25] G. Ke, Q. Meng, T. Fineley, T. Wang, W. Chen, W. Ma, Q. Yei and T. Liu, LightGBM: A highly efficient gradient boosting decision tree, *The 31st Conference on Neural Information Processing Systems*, vol.30, 2017.
- [26] S. Kamada and Y. Kageyama, A study on classification of laughter types focusing on temperature changes on the facial surface, *Proc. of the 87th National Convention of the Information Processing Society of Japan*, 7ZK-02, 2025.
- [27] M. T. Nguyen, Q. K. Nguyen, K. Kazunori and P. Siritanawan, Towards recognizing facial expressions as deeper level: Discriminating genuine and fake smiles from a sequence of image, *2019 6th NAFOSTED Conference on Information and Computer Science*, pp.387-392, 2019.
- [28] K. Inoue, D. Lala and T. Kawahara, Can a robot laugh with you? Shared laughter generation for empathetic spoken dialogue, *Frontiers in Robotics and AI*, vol.9, pp.1-11, 2022.

Author Biography



Sou Kamada obtained the B.E. degree in Electrical and Electronic Engineering from the Faculty of Engineering, Tohoku Gakuin University, Japan, in 2024, and the M.E. degree in Computer Science and Engineering from Akita University, Japan, in 2026. He is currently with Ricoh IT Solutions Co., Ltd. His research interests include human-centered sensing and signal processing.



Yoichi Kageyama obtained the B.E. and M.E. degrees in Computer Science and Engineering, followed by a Dr. Eng. degree from Akita University, Japan, in 1995, 1997, and 2001, respectively. He joined Akita University as a Research Associate in 1997, later advancing to Assistant Professor in 2001 and Associate Professor in 2004. He is currently a Professor at the Graduate School of Engineering Science and the Faculty of Informatics and Data Science. His research focuses on human-centered sensing, remote sensing technologies, and computational image processing.