Feasibility study of classification for music preference level based on galvanic skin response (GSR) and photoplethysmogram (PPG) sensor data with machine learning method

Li Yunheng, Eri Sato-Shimokawara

Tokyo Metropolitan University
6-6 Asahigaoka, Tokyo, Hino, 191-0065, Japan
{li-yunheng@ed., eri@}tmu.ac.jp

Abstract. In recent years, the increasing popularity of music streaming platforms has made it possible to listen a vast array of songs. Therefore, methods to recommend songs that match the preferences of various users have been attracting increasing attention. Traditional methods of music recommendation often rely on the user's playback history and explicit feedback, which may not accurately reflect users' actual preference. However, biological parameters such as photoplethysmogram (PPG) and galvanic skin response (GSR) reflect the electrical activity of the heart and the skin's conductance, providing more accurate preference recognition. In this study, we used short music clips to assess individual preferences, utilizing sensors to record the associated physiological signals and compiling datasets for the experiment. During the post-processing stage, we extracted features from the physiological signals and built a classification model using Random Forest Classifier (RF) and Support-Vector Machine (SVM).

Keywords: Music recommendation, physiological signals, machine learning.

1 Introduction

In recent years, the increasing popularity of music streaming platforms has enabled users to access a vast library of songs. This convenience has spurred interest in developing methods to recommend music that aligns with individual user preferences. Traditional music recommendation systems often rely on users' playback history and explicit feedback. However, these methods may not accurately capture users' actual music preferences.

Music profoundly influences emotions; therefore, the listener's emotions can reflect their preferences. Conventional approaches to emotion recognition are often based on facial expressions and speech signals, but these methods have significant limitations. Psychological researchers believe that emotions are not always expressed straightforwardly, and emotional state expressed in facial expressions or vocal tones may not always reflect a person's actual feelings [1], leading to inaccuracies in emotion detection.

Biological signals such as photoplethysmogram (PPG) and galvanic skin response (GSR) provide a more reliable means of evaluating music preferences, as they reflect the body's physiological reactions to stimuli. PPG measures the variations in blood volume within the micro-vascular bed of tissue. GSR indicates the skin's electrical conductance, which fluctuates with moisture levels and is particularly sensitive to stress responses.

This research proposes a novel music evaluation method that leverages these physiological signals to better understand and predict users' preferences during music listening. The practical application prospects of this music evaluation method are extensive. For instance, dynamic music streaming services could use these insights to adjust music playlists to better align with the listener's preference, potentially improving listener satisfaction.

In this study, we aim to compile datasets for analyzing individuals' responses to music by using short music clips to assess preferences and recording the corresponding physiological signals. We applied random forest classifier and SVM to classify the features extracted from the physiological signals. The experimental results were validated through K-fold cross-validation.

2 Related Work

2.1 Music recommendation method using interactive genetic algorithm

Saito's research [2] aimed to enable users to discover unexpected but enjoyable music. It focused on a music recommendation system using Interactive Genetic Algorithm (IGA), which can be optimized for users' diverse preferences. In this system, users evaluate a set of music data presented as a population, and the evolutionary calculation is repeated over several generations using the evaluation values as fitness values.

The original IGA system required high user effort during the evaluation process. To address this, the system's structure was modified to reduce the number of generations needed for optimization. A new individual selection method based on clustering, called "Cluster Ranking", was devised to streamline the process further (Figure 1).

This research contributes to the field by combining fusion-based recommendations and genetic algorithms to develop a more personalized and serendipitous music recommendation system. In this IGA method, users rate the music on a five-level scale from -2 to 2. This way of evaluating music has been applied in our work.

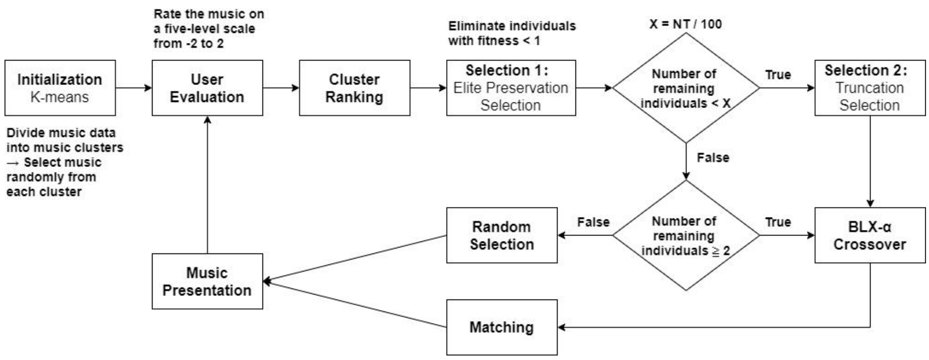

Fig. 1. Configuration diagram of the IGA method proposed by Saito.

2.2 The processing flow of physiological signals

Chen's research presented a novel approach to detecting mind-wandering (MW) using electroencephalography (EEG) signals [3]. The core of their proposed system involves collecting and processing of EEG data from the MM-SART (multi-modal Sustained Attention to Response Task) database. The EEG processing flow can be separated into five steps, as shown in Figure 2. In our study, the approach to analyzing the GSR and PPG data was inspired by Chen's method.

Fig. 2. The processing flow of EEG signals.

3 Methodology

Our research utilized GSR and PPG signals as primary source of physiological information. These signals are typically captured from the fingertips or palm areas [4]. GSR is a highly sensitive and reliable indicator of physiological arousal, which is influenced by sweat gland activity. Relaxing music may lead to an increase in GSR skin resistance due to the decrease in sweat production [5]. This is regulated by the body's sympathetic nervous system—a process not under direct conscious control.

PPG is an affordable, non-invasive optical methodology that uses infrared light to estimate skin blood flow [6]. Data collected from the fingertip or earlobe can be used to calculate heart rate (HR), which may be a valuable parameter in algorithms designed to identify individual preferences. A fluctuating heart rate might indicate a response to dynamic, high-tempo music, potentially correlating with a preference for such genres.

3.1 Signal acquisition

In our study, we used short music clips to assess individual preferences, utilizing the Shimmer3 GSR+ Unit and Consensys Software for capturing the associated physiological signals, thereby establishing a experimental database.

Shimmer3 GSR+ Unit has proven to be a dependable and precise wearable device for capturing bio-signals, making it a valuable tool for biomedical research [7]. This wireless sensor can transmit real-time GSR signal. Likewise, the optical pulse sensor attached to the GSR+ module can provide PPG signals from the fingertip or earlobe. This wearable sensor is characterized by its miniature size, weighing around 20 grams, suitable for collecting physiological data from individuals while maintaining their comfort and privacy. (Figure 3).

Fig. 3. Shimmer3 GSR+ Unit sensor.

3.2 Stimulus Material

The materials for assessing preferences consisted of 70 randomly selected pieces of music from the Music4All Database in various styles, each lasting 30 seconds. Music4All is a music database containing 109,269 songs, useful for music information retrieval (MIR) research [8], and provides key features such as metadata, tags, and genre information for every song.

3.3 Experimental setup

Two participants were invited for this experiment, both of them were students at Tokyo Metropolitan University, Japan. The experimental procedures took place in a quiet environment, free from external disturbances. A laptop was utilized for the experiment. Prior to the test, subjects received a verbal introduce of the experiment's objectives and the concept of fitness. They were also advised that the experiment would last approximately 35 minutes and involve listening to a total of 70 musical selections. After agreeing to participate in the experiment, the participants wore headphones and started listening.

To obtain GSR data, two dry electrodes secured with Velcro bands were strapped to the middle and index fingertips respectively. For PPG data, an optical pulse probe was attached to the ring fingertip (Figure 4). The Shimmer3 GSR+ Unit sensor was placed on the non-dominant arm.

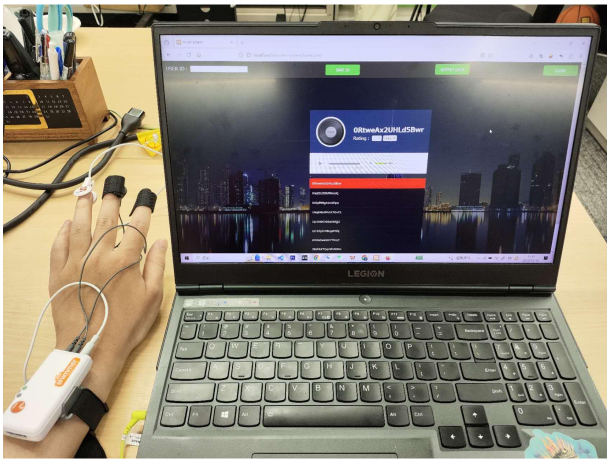

Fig. 4. Subject equipped with attached sensors.

We created an application and a database to store all participants' interactions: user ID, music ID, rating time, and fitness. The interface of this application as shown in Figure 5. During the data acquisition process, the real-time data was recorded at a 51.2 Hz sampling frequency and streamed to the laptop via Bluetooth.

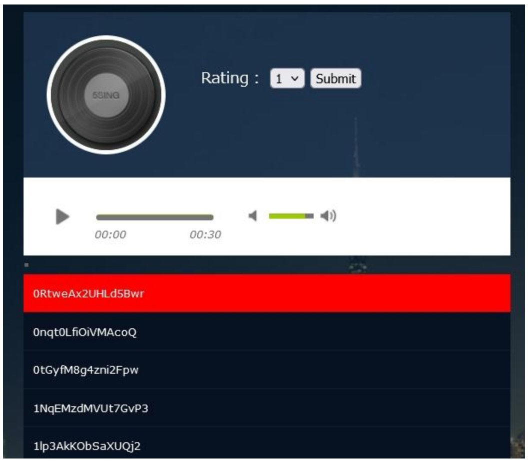

Fig. 5. Application interface.

During the evaluation phase, participants were asked to rate each song on a 5-level scale (1: I don't like this song at all. 2: I don't like this song. 3: Neither like nor dislike. 4: I like this song. 5: I like this song very much.), and then press the submit button. The application recorded their evaluation scores and timestamps into the database during the experiment. Figures 6 and 7 present the example of GSR and PPG signals during music listening period.

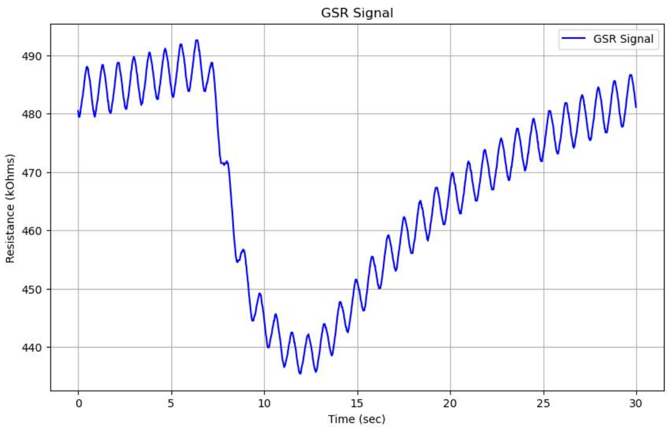

Fig. 6. Example of GSR signal during music listening (30 seconds).

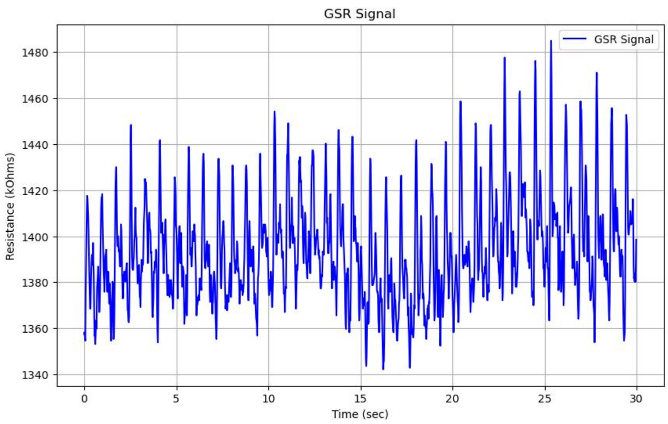

Fig. 7. Example of PPG signal during music listening (30 seconds).

3.4 Pre-Processing

Smoothing. The Shimmer3 GSR+ Unit device's output included the timestamps for each individual data point, allowing for the identification of when data collection began. Therefore, the raw PPG and GSR signals collected from experiments could be divided into several parts corresponding to these timestamps [9]. Next, moving-mean filters were used to remove the high-frequency noise. The most common sources of high-frequency noise are electrical noise at 50/60 Hz and the precision error of the sensor. High-frequency noise can be eliminated with post-processing tools that apply a noise reduction filter that removes the high-frequency components of the signal [10]. The comparison between the processed GSR signal and the raw GSR signal is shown in Figure 8. It can be seen that the signal curve after being processed by the filter is much smoother than the curve of the raw signal.

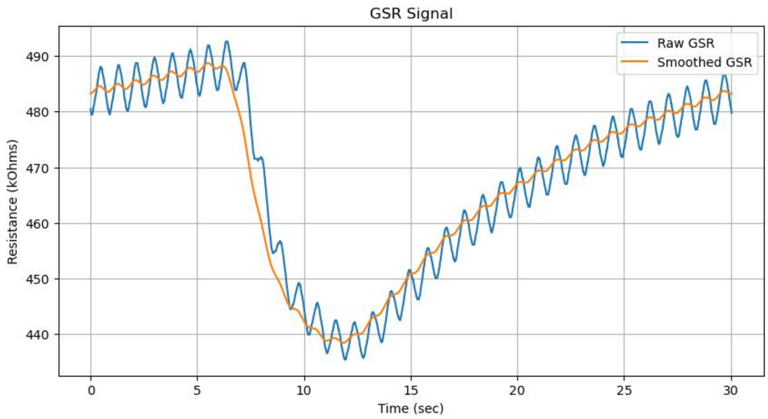

Fig. 8. A comparison between raw GSR and smoothed GSR signals.

Normalization. Due to individual variations in physiology, the value ranges of signals can differ significantly. Without normalization, larger-scale features may dominate and greatly impact the learning algorithm. Normalization aims to scale and standardize the values of different features to be a similar scale. The normalized formula can be represented as:

$$X_i' = \frac{X_i - X_{imin}}{X_{imax} - X_{imin}} \tag{1}$$

In this formula, X_i represents the i-th initial characteristic data. X_{imin} and X_{imax} are the minimum and maximum values respectively of this dimension characteristic data. The resulting normalized value is expressed as X_i' [11]. This formula can transform a data set into a fixed range of [1,0].

Latency Removal. When an emotional event occurs, the electrical conductance of the skin does not change immediately. Instead, it shows a specific delay or latency. According to related research, this latency exhibits variability among individuals and typically spans 1 to 5 seconds following the initiation of the stimulus event [12]. So, the first 2.5 seconds of data from each GSR signal were removed to eliminate the latency.

Finally, all raw signals were adjusted to a consistent length of 27.5 seconds. The preprocessed physiological signals can be plotted as curves for convenient and intuitive observation. Example of preprocessed GSR signals and PPG signals are presented on Figure 9.

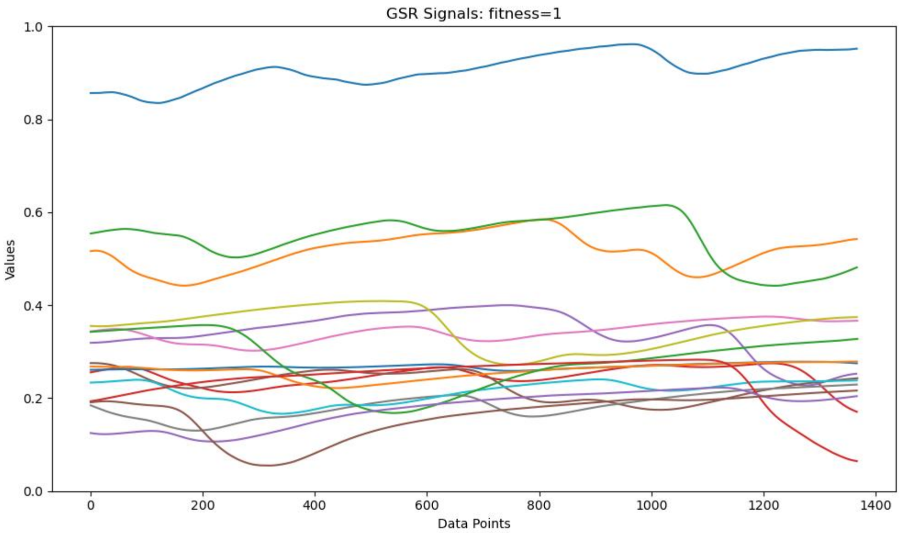

Fig. 9. Example of preprocessed GSR signals.

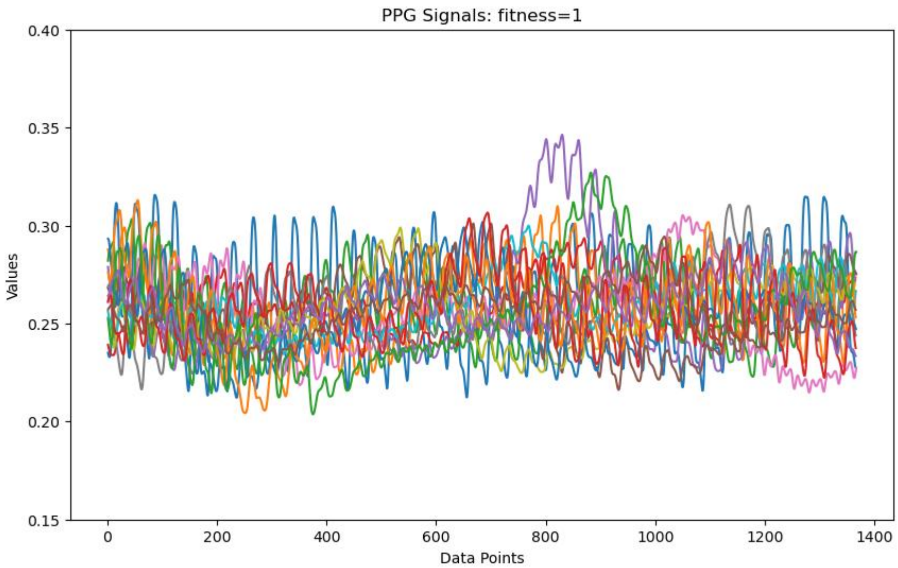

Fig. 10. Example of preprocessed PPG signals.

3.5 Features Extraction

As shown in Table 1, nine features represented by typical statistical values were extracted from the GSR data.

Table 1. Features extracted from the GSR data.

Feature ID	Feature Code	Feature Meaning
1	mean	Mean
2	std	Standard Deviation
3	min_val	Minimum Value
4	max_val	Maximum Value
5	median	Median
6	q1	First Quartile
7	q3	Third Quartile
8	iqr	Interquartile Range
9	cv	Coefficient of Variation

For PPG data, thirteen measures were extracted from the time-domain by HeartPy, which is a Python module for heart rate analysis [13], [14]. Features extracted from the PPG signal are shown in Table 2.

Table 2. Features extracted from the PPG signal.

Feature ID	Feature Code	Feature Meaning
1	hr	Heart Rate
2	ibi	Inter-Beat Interval
3	sdnn	Standard Deviation of NN intervals
4	sdsd	Standard Deviation of Successive Differences
5	rmssd	Root Mean Square of Successive Differences
6	pnn20	Proportion of NN intervals greater than 20 ms
7	pnn50	Proportion of NN intervals greater than 50 ms
8	hr_mad	Heart Rate Mean Absolute Deviation
9	sd1	Standard Deviation of points perpendicular to the line of identity in a Poincaré plot
10	sd2	Standard Deviation of points along the line of identity in a Poincaré plot
11	s	Area of the Poincaré plot
12	sd1/sd2	The ratio of SD1 to SD2
13	breathing rate	The breathing rate

3.6 Classification

We classified the participants' physiological signals into 5 classes based on their fitness values, ranging from 1 to 5. The classification process utilized Random Forest techniques, according to the participants' personal assessment ratings.

Data Fusion. To leverage the complementary information from multiple physiological signals, we fused the extracted features into a single data structure. This fusion facilitated a holistic representation of an individual's physiological state during physical activities.

After the feature extraction, we concatenated these feature arrays horizontally, resulting in a single 2D array 'X', where each row represents a sample and columns are features from GSR signals, PPG signals, and music tempo.

Classifier. This study employed Random Forest classifier and SVM for feature classification. The Random Forest classifier operates by creating multiple decision trees and aggregates their predictions to reach a final classification. The SVM operates by finding the hyperplane that best separates different classes in the feature space.

We increased the model's prediction accuracy, by optimizing its hyperparameters through a grid search methodology [15]. Grid search exhaustively considers all parameter combinations, selecting the one that maximizes accuracy.

3.7 Cross-Validation

Cross-validation is crucial in machine learning as it offers an accurate assessment of the model's performance on unseen data.

To evaluate our classification model's effectiveness and ensure its generalizability, we used a 5-fold cross-validation [16]. This technique divided the feature datasets into 5 equally sized folds. There were 70 pieces of music, so each fold had 14 samples. In each iteration, one fold was designated as the test set, while the other 4 folds served as the training set. This procedure was carried out 5 times, ensuring that every fold had a chance to be the test set exactly once.

The K-fold cross-validation ensured robustness against overfitting and provided reliable estimates of model performance across different subsets of the data. For each fold, we applied three evaluation metrics for each class: precision, recall, and F1-score.

The formulas for these computations are as follows:

$$\text{TP} = \text{true positive rate}, \text{FP} = \text{false positive rate}, \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{4}$$

$$\text{F1} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \tag{5}$$

4 Results and discussion

70 sets of physiological signal data were obtained from each participant, classified into 5 classes according to their ratings for each piece of music. Due to differences in participants' preferences, the quantity of datasets in each class varied. Table 3 shows the distribution of the data sets for each participant across different classes.

Table 3. Distribution of the data sets for each participant across different classes.

	Fitness 1	Fitness 2	Fitness 3	Fitness 4	Fitness 5
ID1	16	15	15	12	12
ID2	1	22	28	18	1

For Participant ID2, Fitness 2,3 and 4 are more frequent than Fitness 1 and 5. This may cause the models to be biased towards the majority classes. So the parameter "class_weight='balanced'" is set to give higher weights to minority classes, reducing the effects of class imbalance.

Table 4 presents the accuracy for the classification results of two participants. The observations from the results can be summarized as follows:

- For both participants, the Random Forest model generally performs better than the SVM model.
- Participant ID2's data is more accurately classified by both models compared to Participant ID1's data.
- The model generally exhibits variability across different folds.

Table 4. Accuracy for the classification results of Participants ID1 and ID2.

	ID1		ID2	
Fold	Random Forest	SVM	Random Forest	SVM
1	0.14285714	0.21428571	0.5	0.28571429
2	0.14285714	0.14285714	0.57142857	0.57142857
3	0.21428571	0.14285714	0.42857143	0.42857143
4	0.28571429	0.14285714	0.28571429	0.5
5	0.21428571	0.28571429	0.42857143	0.28571429
Mean	0.2	0.18571429	0.44285714	0.41428571

Figures 10 and 11 show the confusion matrix for the classification performance across 5 classes. There are noticeable diagonal patterns, indicating some correct classifications for each class. However, the performance is poor for fitness 1 and 5 in Participant ID2's results. This participant rates 1 and 5 to only one song each. This inconsistency highlights the need for a more balanced data distribution or additional samples.

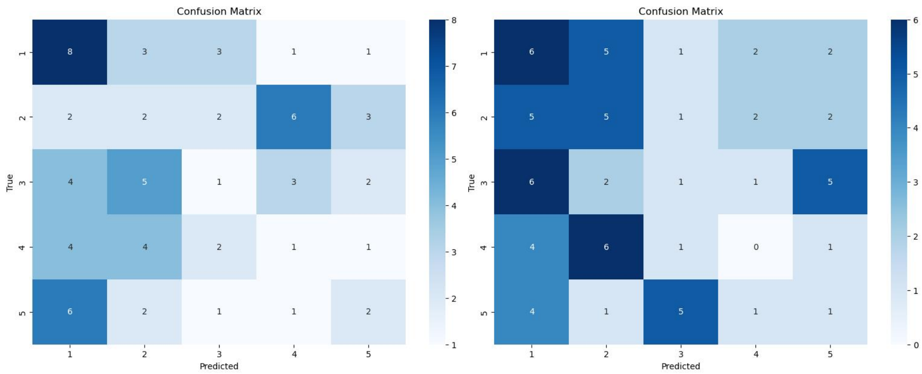

Fig. 11. Confusion matrix for RF (left) and SVM (right) classification of Participant ID1.

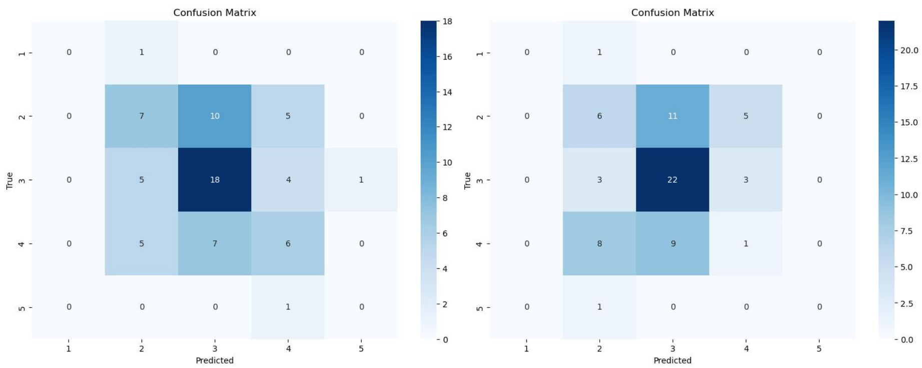

Fig. 12. Confusion matrix for RF (left) and SVM (right) classification of Participant ID2.

5 Conclusion

In this paper, we present a music evaluation method using physiological sensors. An experiment was designed to assess individual preferences. Participants were asked to listen to music and rate each piece from 1 to 5. GSR and PPG signals were recorded to train Random Forest model and SVM model. Nine features were extracted from GSR signal, while thirteen features were extracted from PPG signal.

The results indicate some success in classifying fitness levels utilizing these physiological measurements. The Random Forest model performs marginally better than SVM for both participants. However, the mean accuracy is relatively low, and the performance across folds is variable.

This feasibility study has the issue of relatively low accuracy. This may be due to GSR and PPG signals being more sensitive to changes in emotion and stress, while classifying the degree of music preference is relatively difficult. Furthermore, current machine learning algorithms may not adequately handle the complexity and variability of human responses to music, highlighting the need for further research. Certain fitness levels had fewer samples, impacting the model's capacity to generalize well across these classes. Larger datasets would likely yield more stable and reliable performance metrics.

Future research will likely explore advanced feature extraction techniques and machine learning models to improve accuracy, while also accommodating more participants to collect a larger and more diverse datasets, thereby improving the model's generalization ability.

Additionally, personal variations in age, gender, and other characteristics can lead to generalized models that might not be universally applicable [17]. Consequently, future work may incorporate personalized models to account for individual differences. We also plan to deploy the proposed method to Saito's music recommendation systems and enhance the model's accuracy for practical applications in the future.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Number JP22H03698.

Disclosure of Interests. The authors affirm that they have no competing interests that are pertinent to the subject matter of this article.

References

1. Ikeda, Y., Horie, R., & Sugaya, M.: Estimating emotion with biological information for robot interaction. Procedia computer science, 112, 1589-1600 (2017)
2. Saito, T., & Sato-Shimokawara, E.: Music Recommender System Considering the Variations in Music Selection Criterion Using an Interactive Genetic Algorithm. In International Conference on Computer Information Systems and Industrial Management, pp. 382-393. Cham: Springer Nature Switzerland (2023). doi: 10.1007/978-3-031-42823-4_28
3. Chen, Y. T., Lee, H. H., Shih, C. Y., Chen, Z. L., Beh, W. K., Yeh, S. L., & Wu, A. Y.: An effective entropy-assisted mind-wandering detection system using EEG signals of MM-SART database. IEEE Journal of Biomedical and Health Informatics, 26(8), 3649-3660 (2022)
4. van Dooren, M., & Janssen, J. H.: Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. Physiology & behavior, 106(2), 298-304 (2012)
5. Pagar, V., Bhadane, P., & Shaligram, A.: Stress Detection Using Galvanic Skin Response With Affective Impact Of Different Climate Conditions. Journal of Pharmaceutical Negative Results, 3961-3969 (2022)
6. Elgendi, M.: On the analysis of fingertip photoplethysmogram signals. Current cardiology reviews, 8(1), 14-25 (2012)
7. Santana, I. A. P., Pinhelli, F., Donini, J., Catharin, L., Mangolin, R. B., Feltrim, V. D., & Domingues, M. A.: Music4all: A new music database and its applications. In 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 399-404. IEEE (2020)

8. Burns, A., Doheny, E. P., Greene, B. R., Foran, T., Leahy, D., O'Donovan, K., & McGrath, M. J.: SHIMMER™: An extensible platform for physiological signal capture. In 2010 annual international conference of the IEEE engineering in medicine and biology, pp. 3759-3762. IEEE (2010)
9. Li, Z., Xing, Y., Pi, Y., Jiang, M., & Zhang, L.: A novel physiological feature selection method for emotional stress assessment based on emotional state transition. Frontiers in Neuroscience, 17, 1138091 (2023)
10. Galvanic skin response (GSR). https://connect.tobii.com/s/article/galvanic-skin-response-gsr?language=en_US#GSR_artifacts
11. Udovičić, G., Đerek, J., Russo, M., & Sikora, M.: Wearable emotion recognition system based on GSR and PPG signals. In Proceedings of the 2nd international workshop on multimedia for personal health and health care, pp. 53-59 (2017)
12. Galvanic skin response (GSR). https://connect.tobii.com/s/article/galvanic-skin-response-gsr?language=en_US#ResponseTimeGSR
13. Van Gent, P., Farah, H., Van Nes, N., & Van Arem, B.: Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the taking the fast lane project. Journal of Open Research Software, 7(1), 32 (2019). doi: 10.5334/jors.241
14. Van Gent, P., Farah, H., Van Nes, N., & Van Arem, B.: HeartPy: A novel heart rate algorithm for the analysis of noisy signals. Transportation research part F: traffic psychology and behaviour, 66, 368-378 (2019). doi: 10.1016/j.trf.2019.09.015
15. Saranya, G., & Pravin, A.: Grid search based Optimum feature selection by tuning hyperparameters for heart disease diagnosis in machine learning. The Open Biomedical Engineering Journal, 17(1) (2023)
16. Petrescu, L., Petrescu, C., Oprea, A., Mitruț, O., Moise, G., Moldoveanu, A., & Moldoveanu, F.: Machine learning methods for fear classification based on physiological features. Sensors, 21(13), 4519 (2021)
17. Liu, Y., Li, J., Cui, D., & Sato-Shimokawara, E.: Multi-modal emotion classification in virtual reality using reinforced self-training. Journal of Advanced Computational Intelligence and Intelligent Informatics, 27(5), 967-975 (2023)