EmoTracer: A Wearable Physiological and Psychological Monitoring System With Multi-modal Sensors

Abstract

This research report builds upon and improves the paper EmoTracer: A Wearable Physiological and Psychological Monitoring System With Multi-modal Sensors where published in the *Ubicomp Workshop* in 2022. We have refined the model presented in that paper by addressing three key areas: 1) reducing computation resources by proposing the prototype network for few-shot learning, and 2) improving the hardware design. The experimental outcomes demonstrate that the various enhanced models we propose have achieved varying degrees of performance improvement and evaluate the effectiveness and efficiency of our system.

Keywords: Affective Computing, Wearable Device, Few-shot Learning, Emotion Recognition, Prototype Network.

1 Introduction

Emotions served as human responses to humans or things, affecting decision-making, perception, and human interaction. As the standard of living develops and the level of medical care enhances, human health issues are gaining wider and wider attention. For the ageing and sub-healthy population, real-time collection of their physiological indicators can effectively monitor the health status and achieve the purpose of disease prevention and early treatment. Health issues are gaining wider and wider attention as living standards develop and as medical care improves. Targeting the ageing and sub-healthy population, real-time collection of their physiological indicators can effectively monitor the state of health and achieve the purpose of disease prevention and early treatment. At the same time, the focus is gradually shifting to mental health as a major issue. The impact of depression, anxiety and other mental disorders on people cannot be ignored. Currently, with the development of mobile computing technology, the attention of researchers has focus on emotion recognition based on mobile ubiquitous devices (ear-worn device or wrist worn device). Conpared with traditional professional medical devices such as electroencephalography and electrocardiograph, these devices can collect human physiological information such as galvanic skin response (GSR) and photoplethysmography (PPG) without the need of medical personnel operation. In order to make up for the shortcomings of the above research, this researcher propose a multimodal smart wearable-based human physiological data monitoring system called EmoTracer, which includes a wearable device and an intelligent terminal. The device is worn on the user's hand in a non-invasive manner to collect realtime physiological data, including heart rate signal, blood oxygen saturation signal, skin conductivity signal and skin temperature signal, and transmits the measured data to an intelligent terminal in real time via wireless transmission; In addition, the hardware development is based on commercial, low-cost sensors, and a set of portable wearable devices with integrated sensors and small size is designed on the STM32 development board independently. The device has the advantages of high sampling accuracy of each sensor, low packet loss and a rechargeable interface for multiple cycles. At the smart acquisition terminal, a mobile application is designed to pre-process the raw data from the transmitted sensors, storing, filtering and converting the signal. In turn, the human physiological data can be presented in real-time and accurately in the interface. As for our work, we improve the design of Emotracer system[1] in two aspects: algorithm and hardware. Specifically, we use the prototype network for few-shot learning to replace the ResNet18 in Emotracer, the motivation is the consideration of how to use limited data to significantly imrpove the cross-subject performance. Secondly, we diminish the size of the wrist-worn device for the wearability and confort of subjects.

2 Related works

2.1 Emotion Recognition with proprietary device

Specialized devices include EEG, ECG, MRI, functional near-infrared spectroscopy (fNIRS) and functional magnetic resonance imaging (fMRI), which are widely used in mental health research as mainstream data acquisition equipment for human physiological signals, such as emotion recognition [4,9,10,12,16,17,19,22,29,35,36], stress detection [3,24,37] and mental health detection [20,25,30,31,38]. In most of these studies, EEG signal is used to analyse the relationship between brain neuroscience and human cognition. The part of studies focused on analyzing dynamic, non-linear and complex temporal correlation patterns using advanced time series approaches. For example, Ding et al. [7] used a multi-scale convolutional neural network to analyse the temporal and channel characteristics of EEG signals at different granularities. Moreover, Ding et al. [8] introduced MASA-TCN, a novel model for EEG-based emotion recognition that addresses both continuous regression and discrete classification of emotional states. The model innovatively integrates a Space-Aware Temporal Convolutional Layer (SAT) to capture spatial relationships among EEG electrodes, enhancing the ability to discern nuanced emotional states. Additionally, a Multi-Anchor Attentive Fusion Block (MAAF) is designed to adaptively learn dynamic temporal dependencies within EEG signals. Due to the multi-channel nature of EEG signal acquisition (e.g., 32 channels on DEAP emotion dataset and 62 channels on SEED emotion dataset, etc.), many studies utilize graph neural networks to explicitly learn spatial dynamical features. Zhu et al. [39] developed a Graph Input layer attention Convolutional Network (GICN) for EEG-based depression recognition. The model utilizes a learnable weight matrix in the input layer, taking the brain function network as the adjacency matrix and linear EEG features as node features. It outperformed other methods, indicating its potential as an effective auxiliary tool for depression recognition. Apart from EEG signals, few studies separately conducted based on single source data such as ECG, MRI, etc. In other words, they often make use of multi modal physiological signals and then use the model to learn the independent properties of each modality and the potential correlations between the modalities, thus expanding the decision space that can be learn by the model. For instance, Zhang et al. [38] proposed a multi-modal Graph Neural Network (GNN) framework for early diagnosis of Alzheimer's Disease (AD) using structural magnetic resonance imaging (sMRI) and positron

[1]The manuscript of our work have submitted to ***IEEE Transactions on Affective Computing***.

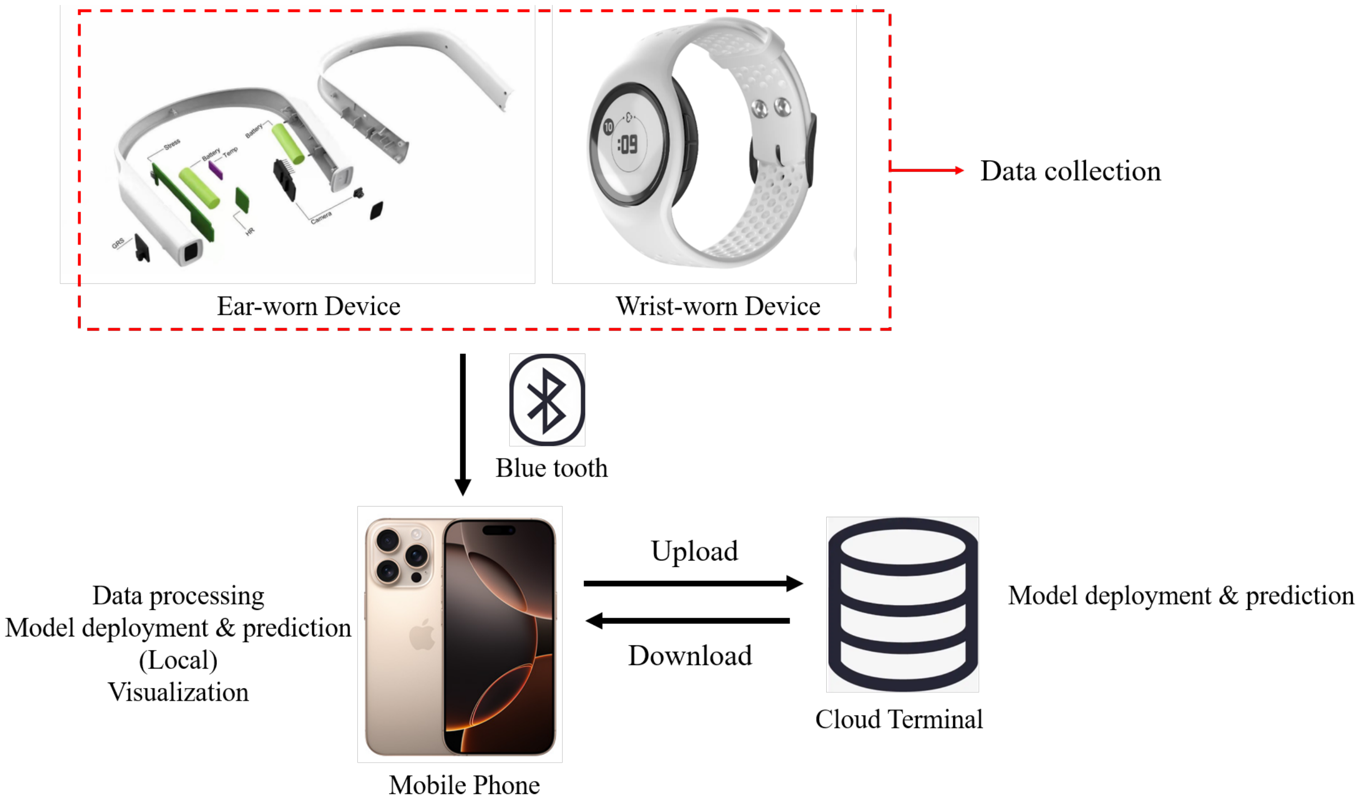

Figure 1. Classic Design process.

emission tomography (PET) scans. The study leverages GNNs to handle non-Euclidean data domains, creating brain networks from sMRI or PET images within a population graph framework that integrates phenotypic information with imaging features. The model, termed Multi-modal GNN, consists of separate GNN branches for each modality, with a fusion technique that combines node vectors and adjacency matrices at both levels. Late fusion is then applied to produce a final prediction. The method aims to capitalize on the complementary nature of structural and functional brain information provided by sMRI and PET, respectively, to enhance predictive models for AD diagnosis.

2.2 Emotion Recognition with ubiquitous device

In contrast with proprietary devices such as brain electrode caps, wearable devices for depression recognition technology have been attracting attention owing to their low cost and portability. These devices offer a more affordable and less intrusive approach to collecting user data, making them suitable for large-scale studies and daily monitoring of mental health parameters. They typically integrate some necessary sensors such as heart rate, blood oxygen saturation(infrared and red light signal), galvanic skin response, skin temperature, IMU signals to collect the physiological and behavior information of users.

Figure 1 illustrates a pipeline of such approaches, where the behavioural or physiological information of the subject is first captured using an ear-worn device or a wrist-worn device. The data is then transmitted to the user's mobile phone using the Bluetooth protocol. The researcher can deploy the model on the mobile phone to perform a series of processes and predictions on the processed data. Alternatively, the data can be uploaded to a cloud server with higher computing power resources for similar processing, and then the results can be downloaded to the mobile phone for visualisation of the results.

The use of pervasive wearable devices allows for regular daily monitoring of tasks such as emotion analysis [26, 33, 34, 40] and mental health diagnosis [1, 2, 18, 23]. Wang et al. [33] novelty designed a real-time mood analysis system, Emotracer, was designed; specifically, users wear a self-designed wristband device that collects earth rate, blood oxygen saturation, galvanic skin response, and skin temperature signals. The physiological information is transmitted via Bluetooth to a mobile phone for pre-processing, visualisation and prediction (Resnet18). Not only for emotion analysis, but the use of such devices for tasks such as monitoring a user's mental state has even greater applications. For example, Pedrelli et al. [23] conducted a study utilizing a combination of wearable and mobile sensors to monitor changes in depression severity. The experiment used E4 Empatica wearable mobile device to collect electrical skin, heart rate and peripheral epidermal temperature signals from subjects to monitor changes in depression severity. After necessary preprocessing of the data, the authors combined expert knowledge to generate dozens of manual features and finally use ensemble learning methods (logistic regression and XGBoost models).

3 Method

3.1 Prototype of Emotracer

The EmoTracer system consists of three parts: the physiological indicator collection device, the Android APP and the data processing.

3.1.1 Hardware Design

The wearable device includes a temperature module, a heart rate module, an skin conductivity module, a blood oxygen module, a wireless communication module and a microcontroller. These modules convert the collected data into digital signals via the ADC on the microcontroller. The microcontroller transmits the received data to the wireless communication module at set intervals using the UART protocol, which in turn sends the data to the intelligent device via the wireless communication module. In addition, the device is provided with a wrist strap which is secured to said encapsulated housing for securing the device to the user's wrist in a non-intrusive manner. The microcontroller is an STM32F103C8T6 Cortex-M3 32bit microcontroller. The main frequency of the microcontroller is up to 72 MHz, the power supply chip is SY8089 DC-DC buck chip with ultra-low quiescent current, the lithium battery voltage to 3.3V for the microcontroller and other sensors to use, the charging chip model TP4054 charging circuit, through the TypeC plug into the power supply can be set after the maximum 500ma current size to the lithium battery power supply.The type of Bluetooth module is the RF-Star EFR32BG22A1 ultra-low power Bluetooth module, which supports the BLE5.2 protocol.In addition, the MX1.25 sensor is used due to its features that allow the use of an optocoupler switch for individual control of the power supply voltage for each sensor to control power consumption; the provision of an MX1.25 sensor interface to receive physiological data signals from external sensor inputs; and the use of a low quiescent current DC-DC chip to reduce static power consumption to improve endurance. The sampling rate and data accuracy of the sensors of each acquisition module and the meaning of the acquired data are shown in Table 1. EmoTracer device uses the BLE communication protocol and the lower computer program sends data packets every 50ms. As each sensor has a different sampling rate, the oximetry module with the highest sampling rate

Table 1. Parameters of each sensor.

Signal type	Sensor model	Sampling rate	Data type	Price
Heart rate	Pulse sensor	400 Hz	Raw analog data	3.425 $
Blood oxygen saturation	Max30102	400 Hz	Red light and infrared light intensity data	0.959 $
Galvanic skin response	Grove GSR	200 Hz	Voltage simulation data	8.219 $
Skin temperature	LMT70	100 Hz	Voltage simulation data	3.425 $

is used for calibration and the data in the corresponding ADC data register of the sensor is read by taking the residual relationship. Once the blood oxygen module has collected 20 sets of data, the data from the different sensors are stored according to the protocol rules and sent to the Bluetooth module via the serial port.

3.1.2 Software Design

The software component of the EmoTracer system includes an Android application with the following features:

- Data Collection and Visualization Interface: This interface allows for Bluetooth connection to the wearable device and displays data from each sensor module (heart rate, oxygen saturation, skin temperature, and skin conductivity).
- User Information Interface: Collects personal information from users and conducts scale assessments for mental health status evaluation.
- Emotion Log: Enables users to record their daily emotions through log entries or pictures, tracking emotional changes over time.
- Special Disease Record Interface: Allows users to record episodes of psychological disorders, such as anxiety or panic attacks, including the timing, symptoms, and recovery, which can be shared with healthcare providers.
- The software processes the raw data from the sensors, stores it, and presents the physiological data in real-time and accurately on the user interface.

It also includes emotion recognition capabilities, utilizing machine learning models like ResNet18 for classifying emotions based on the collected physiological data.

3.2 Improvement

According to the temporal features of physiological signals, we design a few-shot cross-subject emotion recognition model inspired by the prototype network. This model not only achieves fast and accurate classification of few shots but also addresses the issues of over-fitting and limited information exchange between distant features that are common in the prototype network. The network design is shown in Fig. 2.

3.2.1 Task Generation

Task Generation: To create varied and authentic personal scenarios based on a given source dataset, we draw on the task generation strategy of MetaSense [14] to decompose the batch data into multiple sub-tasks, which are divided into support set and query set during each training process. The support set forms the prototypes, while the query set refines the position of the prototypes. In each sub-task, following the setting of n-way k-shot, N classes are selected from the source domain data, and k shots are chosen as the support set S from each selected class. For the support set S, a total of Nk shots are selected. From the remaining shots of each class in the N classes, b shots are chosen as the query set Q, resulting in a total of $N \times b$ shots. In this work, the shots come from a single user. Random sampling will be employed within the shots of user, ensuring an equal number of shots across all categories for this task. Specifically, we generate 300 tasks for each user data, and select 2 tasks from each batch. For the six basic emotions, the support set is constructed by the 6-way k-shot setting, with k shots selected for each class. The remaining shots for each class are chosen as the query set.

3.2.2 Embedded network design

In our design, the embedding network is optimized by the LMFN network structure, which can extract comprehensive and valuable information efficiently. Moreover, in order to prevent zigzagging gradient directions during back-propagation, we employ the LeakyReLU activation function as embedded network kernel. After the signals from each channel undergo source-specific feature extraction via LSTM, the data is concatenated along the feature dimension and fed into the LMFusion layer. The LMFusion layer first processes the multi-dimensional data through an LSTM network for multi-sourcr temporal feature extraction, and then the data is forwarded to two merged convolutional layers (Merged Conv1D1 and Merged Conv1D2) to further explore the inter-source correlations.

To mitigate over-fitting, the model is augmented with a global average pooling layer and a Dropout layer, facilitating propagation of effective information between modalities in the subsequent module for feature extraction. Furthermore, the Batch normalization is incorporated after each convolution layer to stabilize the network training and markedly reduce the computational cost of the model. Finally, the multi-source signals will be transformed into a feature vector of 64×1×1600 dimensions, effectively serving as the representative vector within the embedding space.

3.2.3 Training process

Prototypical Networks is an innovative machine learning algorithm that excels in few-shot learning tasks. The algorithm 1 by randomly sampling several classes from the training set, extracting a fixed number of support set S and query set Q for each class. Support set is used to construct class prototypes, while query set is employed to evaluate the performance of model. The core idea of the algorithm is to utilize support set to build prototypes for each class, with these prototypes representing the average of samples within the class. During the classification phase, the algorithm calculates the distance between samples in the query set and these prototypes to perform classification. The distance calculation employs the Inverse Multiquadric(IMQ) function to assign weights to samples, which helps mitigate the impact of outliers on prototype computation. This method not only enhances the robustness of the model but also simplifies the training process due to the minimal parameters of the IMQ function. The algorithm updates the class prototypes using the calculated weights, with the formula $p_k = \lambda_i x_i$. For each sample in the query set, the algorithm computes its distance from all class prototypes and uses the Softmax function to calculate the probabilities of belonging to each class. Ultimately, the sample is classified into the class with the highest probability. To address the issue of sparse gradient updates that may occur in few-shot data sets, the algorithm employs the LazyAdam optimizer, which can more effectively handle sparse gradient updates, thereby reducing the risk of overfitting. Additionally, by strategically reducing the learning rate at specific training epochs, the model's convergence speed and performance are further optimized.

In our work, the initial learning rate of the Prototypical Networks is set to 0.001, with a maximum of 300 training epochs, and the learning rate is reduced after the 100th and 200th training epochs to accommodate the dynamic needs of model training. Overall, Porototypical Networks demonstrates excellent performance and computational efficiency in emotion recognition, through their unique prototype construction and distance measurement methods.

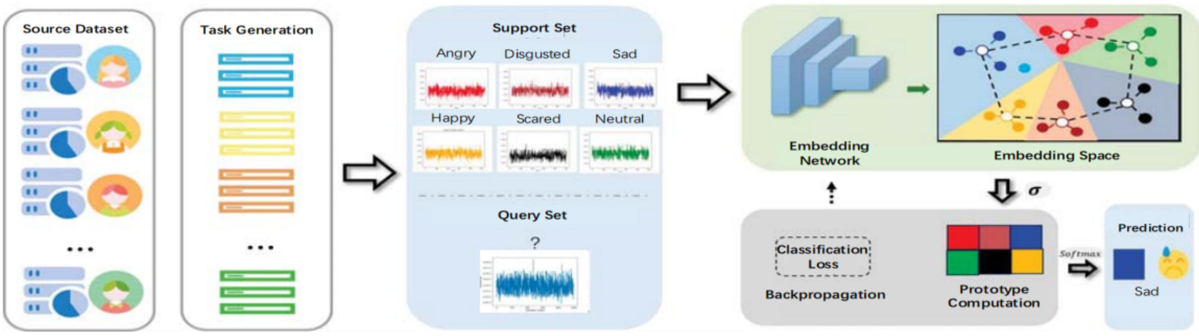

Figure 2. Prototype network design

4 Experiment

4.1 Data collection

Emotracer dataset: The description of dataset can be find at original paper.

Our dataset [2]: We recruit 30 participants aged between 17 and 56 years from our campus to collect real emotional shots by watching video clips. Before the experiment, all participants agree to sign the Institutional Review Board (IRB) agreement [15], and are given a 1-minute buffer to regulate their emotions to ensure reaching a calm state. Aiming to acquire diverse emotional inductions, we select audiovisual stimuli from multiple databases, including DEAP [21], FilmStim [27], IADS [5], SEED [11], and EMDB [6], complemented by internetsourced clips reflecting a range of emotions such as entertainment, horror, and anger. There are 173 audiovisual materials utilized across the experiments, which include 4, 54, 30, 15, and 50 clips respectively from aforementioned databases and 20 online video clips as experimental stimuli materials. The experiment takes place in a typical office environment with noise levels ranging from 40 dB to 55 dB and lasts for 2 to 3 hours for each participants. Participants are equipped with our designed wristband device on their right hands, and sit quietly to view the audiovisual stimuli. Simultaneously, they activate the device and press the 'Start

[2]Both two dataset are unavailable due to privacy or ethical restrictions. The code is available at Github in AdvTech of CSSE.

Table 2. The result of intra-subject emotion recognition

Test			LSTM	TSception [7]	DCN [28]	DANN [13]	Emotracer [32]	LMFN-PN(Ours)
Single Subject Test	Emotracer dataset	Accuracy	66.9%	94.1%	92.8%	86.6%	91.5%	**94.6%**
		F1 Score	68.3%	92.4%	90.6%	84.8%	89.8%	**92.9%**
	Our dataset	Accuracy	71.9%	89.2%	88.8%	74.1%	81.8%	**95.6%**
		F1 Score	68.5%	85.4%	85.7%	70.5%	77.1%	**88.8%**
Cross Subject Test	Emotracer dataset	Accuracy	57.2%	68.4%	66.7%	78.9%	59.2%	**85.1%**
		F1 Score	52.9%	66.1%	64.0%	75.7%	57.1%	82.6%
	Our dataset	Accuracy	39.4%	55.1%	54.1%	65.6%	39.1%	**80.7%**
		F1 Score	38.8%	53.6%	52.2%	63.2%	37.9%	**78.2%**

Test' button on main screen of app. As each clip concluded, participants cease the test, following which they rate their emotional experiences using the SAM scale from 1 to 9 for pleasure, arousal, and dominance and evaluate the familiarity and likability on a scale from 1 to 5 in their self-reports. Finally, we collect 11 GB of physiological data.

4.2 Comparison of Intra-subject and Cross-subject Emotion Classification

To test the validity of the improved model, both single-subject and cross-subject testing methods were carried out, specifically. We divided the training and test sets in the ratio of 8:2 as single-subject experimental data. On this basis, we further used the leave-one-subject-out cross-validation(LOSO-CV) method as the cross-subject experimental data. The experimental results are shown in the Table 2, and what can be clearly found is that all models under the cross-subject test are not as effective as the single-subject test. Therefore, it can be found that it is a challenge to deploy the models effectively under LOSO-CV tests that are more in line with practical application scenarios. However, whether it is single-subject test or cross-subject test, our model not only achieves effective improvement over the original Emotracer, but also exceeds the test performance of all baseline methods. This validates the superiority of small-sample learning strategies based on prototype networks.

4.3 Parameter Sensitivity

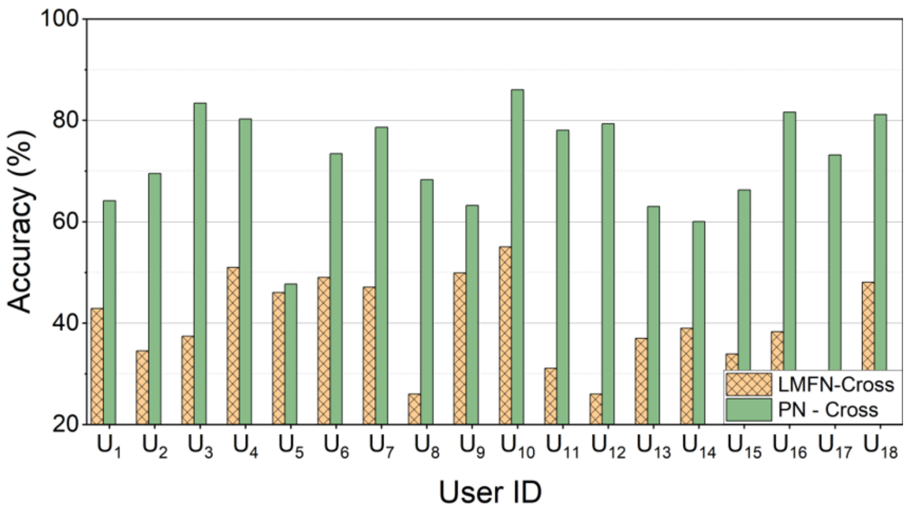

Figure 3. Prototype network design

To further demonstrate the advantages of the prototype network model, we tested the performance of the embedded network against the prototype network. As shown in the Figure 3, under the 1-shot setting condition, we can clearly find that the performance of the prototype network is substantially better than the embedded network, which reflects the advantage of meta-learning.

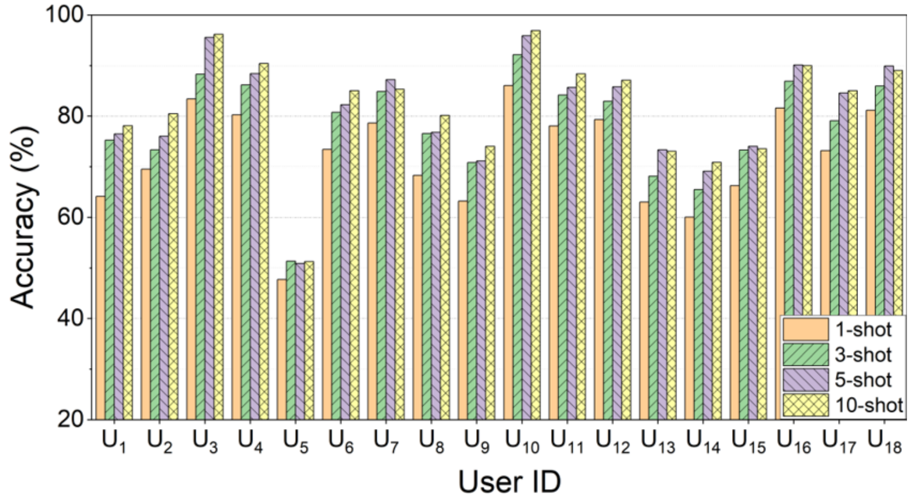

Figure 4. Impact of the number of samples provided in the support set on model performance

Furthermore, we test the effectiveness of our proposal under different number of shots in support set. We provide four types of setting, which is 1-shot, 3-shot, 5-shot, and 10-shot. The results is shown in Figure **??**. As the number of shots in the support set increases, the performance of emotion recognition gradually stabilizes. Considering the limitation of the few-shot physiological signals in our task and the essential goal of stable recognition performance for new users, we choose to utilize $k = 5$ shots for cross-subject emotion recognition in our research.

4.4 Confusion of Emotion Types

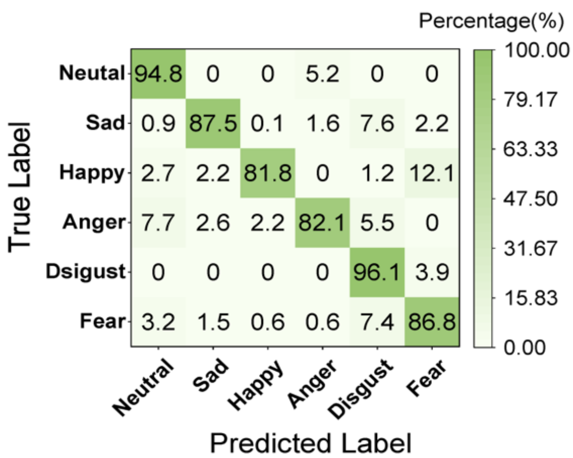

Figure 5. The results of single-subject.

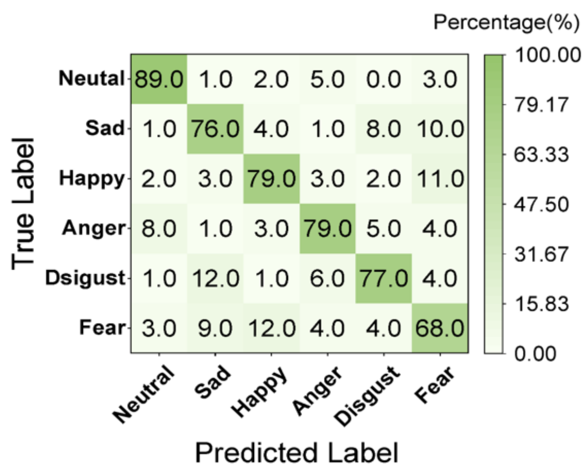

Figure 6. The results of cross-subject.

In our quest to identify emotions that are frequently misclassified in emotion recognition tasks, we have meticulously analyzed the performance of both intra-subject and crosssubject models across six distinct emotional categories, all under uniform experimental conditions. Fig. 5 reveals that while Neutral and Fear are accurately identified, there is significant confusion in distinguishing between Sadness, Disgust, and Anger. This overlap is likely attributed to the high arousal characteristics shared by these negative emotions, coupled with the similarities in the stimulus patterns they elicit. Additionally, Fig. 6 underscores that these classification challenges are consistent across different subjects. The differentiation between Happiness and Fear is particularly problematic, a difficulty that may stem from the high arousal levels intrinsic to both emotions, thus blurring the lines between them in the context of emotional recognition.

4.5 Hardware test

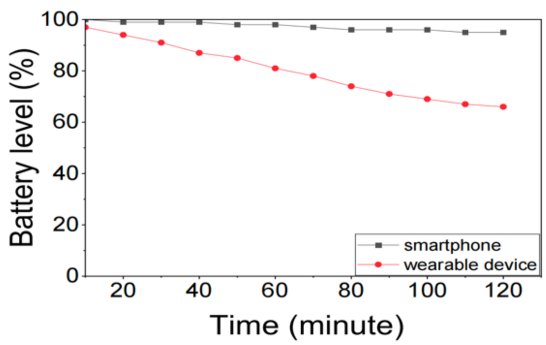

Figure 7. Battery consumption of mobile devices.

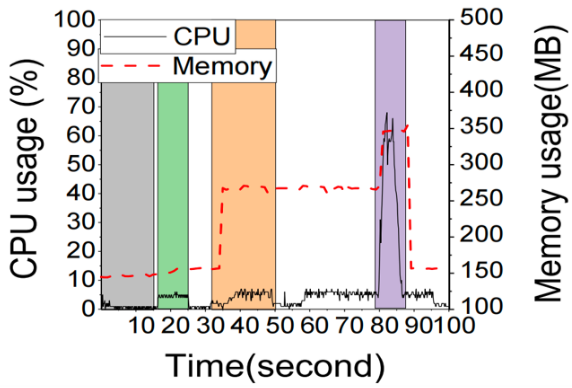

Figure 8. Load of CPU and Memory.

4.5.1 Battery Consumption

During the initial evaluation phase, we ensure that both the smartphone and the wearable device started with a full battery charge at 100%. The participants are instructed to wear the wearable device and have our software application running on their smartphones, with all other applications closed to minimize interference. This setup allows us to focus solely on the power consumption of our system. The evaluation lasted for two hours, providing us with valuable insights into the battery performance. During this time, we utilize the Android Debug Bridge, a versatile tool that enables us to obtain real-time battery information directly from the system. As illustrated in Fig. 7, after the two-hour evaluation period, the battery level of the smartphone decreased by 5%, while the wearable device experienced a more significant drop of 34%. Based on this trend, the entire system can achieve continuous data transmission for over 6 hours. To further enhance the usability of system and adapt it to daily usage habits, we introduce a low power mode specifically for the wearable device. In this mode, the data transmission rate is reduced, sending information only once every 5 minutes. Measurement results shows that, in low power mode, the hardware system can operate continuously for more than 13 hours, demonstrating a remarkable endurance capability. In conclusion, our system not only offers efficient data transmission but also boasts impressive battery life, making it a reliable and long-lasting solution for various applications that require continuous monitoring and data exchange.

4.5.2 Load of CPU and Memory

During the evaluation process of the CPU and memory usage for system operation, all other applications and services are closed to ensure accurate measurements. The system operation is divided into various stages, each with distinct resource demands in Fig. 8. Initially, in the Silent State, the CPU usage was minimal, fluctuating between 1% and 3%. As the system progressed to Data Visualization, the CPU and memory usage increased moderately, with the memory consumption jumping by 150 MB to 160 MB. During Data Preprocessing, the CPU usage remained relatively stable, while memory usage climbed further to 250 MB to 260 MB, indicating the system's engagement in data manipulation. Finally, during model inference, the system's resource consumption surged, with memory usage reaching 350 MB and CPU usage peaking at 65%, underscoring the computational intensity of the predictive modeling process. The system demonstrates low resource consumption in its idle state, indicating good energy efficiency when not in use. For the data visualization and preprocessing stages, there is an increase in resource demand, but it remains within reasonable limits, suggesting that the system can handle these tasks without excessive resource consumption. In the model prediction phase, resource consumption increases significantly, which is typical for deep learning models, but both CPU and memory usage remain within acceptable ranges, indicating that the system is capable of supporting complex emotion recognition tasks. Overall, the system performance is satisfactory.

5 conclusion

- This work was inspired by Emotracer to design a wearable emotion recognition system. Based on its design, both the hardware and algorithmic aspects of the system were improved.
- In terms of hardware, the wearable device designed in this work is smaller than the replicated work, which can improve the wearing comfort of the user.
- On the algorithmic side, the prediction performance under single-subject conditions and cross-subject conditions was improved by designing a small-sample learning method based on a prototype network and an improved LSTM network.
- Wearable device-based physiological fingerprint acquisition applied to emotion recognition has some specific deployment scenarios, such as recognising their stable emotions prior to high altitude work, but this research is not yet suitable for deployment in in-the-wild scenarios due to the design of the wearable device and the signal acquisition.

References

[1] Alaa Abd-Alrazaq, Rawan AlSaad, Farag Shuweihdi, Arfan Ahmed, Sarah Aziz, and Javaid Sheikh. Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. *NPJ Digital Medicine*, 6(1):84, 2023.

[2] Abdullah Ahmed, Jayroop Ramesh, Sandipan Ganguly, Raafat Aburukba, Assim Sagahyroon, and Fadi Aloul. Evaluating multimodal wearable sensors for quantifying affective states and depression with neural networks. *IEEE Sensors Journal*, 2023.

[3] Yara Badr, Usman Tariq, Fares Al-Shargie, Fabio Babiloni, Fadwa Al Mughairbi, and Hasan Al-Nashash. A review on evaluating mental stress by deep learning using eeg signals. *Neural Computing and Applications*, pages 1–26, 2024.

[4] Patrícia Bota, Tianyi Zhang, Abdallah El Ali, Ana Fred, Hugo Plácido da Silva, and Pablo Cesar. Group synchrony for emotion recognition using physiological signals. *IEEE Transactions on Affective Computing*, 2023.

[5] M Bradley and P Lang. The international affective digitized sounds (2-nd edition; iads-2): Affective ratings of sounds and instruction manual gainesville. *The Center for Research in Psychophysiology: Gainesville, FL, USA*, 2007.

[6] Sandra Carvalho, Jorge Leite, Santiago Galdo-Álvarez, and Oscar F Gonçalves. The emotional movie database (emdb): A self-report and psychophysiological study. *Applied psychophysiology and biofeedback*, 37:279–294, 2012.

[7] Yi Ding, Neethu Robinson, Su Zhang, Qiuhao Zeng, and Cuntai Guan. Tsception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition. *IEEE Transactions on Affective Computing*, 14(3):2238–2250, 2022.

[8] Yi Ding, Su Zhang, Chuangao Tang, and Cuntai Guan. Masa-tcn: Multi-anchor space-aware temporal convolutional neural networks for continuous and discrete eeg emotion recognition. *IEEE Journal of Biomedical and Health Informatics*, 2024.

[9] Guanglong Du, Wenpei Zhou, Chunquan Li, Di Li, and Peter X Liu. An emotion recognition method for game evaluation based on electroencephalogram. *IEEE Transactions on Affective Computing*, 14(1):591–602, 2020.

[10] Xiaobing Du, Cuixia Ma, Guanhua Zhang, Jinyao Li, Yu-Kun Lai, Guozhen Zhao, Xiaoming Deng, Yong-Jin Liu, and Hongan Wang. An efficient lstm network for emotion recognition from multichannel eeg signals. *IEEE Transactions on Affective Computing*, 13(3):1528–1540, 2020.

[11] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for eeg-based emotion classification. In *2013 6th international IEEE/EMBS conference on neural engineering (NER)*, pages 81–84. IEEE, 2013.

[12] Ziwang Fu, Feng Liu, Qing Xu, Xiangling Fu, and Jiayin Qi. Lmr-cbt: Learning modality-fused representations with cb-transformer for multimodal emotion recognition from unaligned multimodal sequences. *Frontiers of Computer Science*, 18(4):184314, 2024.

[13] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.

[14] Taesik Gong, Yeonsu Kim, Jinwoo Shin, and Sung-Ju Lee. Metasense: few-shot adaptation to untrained conditions in deep mobile sensing. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, pages 110–123, 2019.

[15] Christine Grady. Institutional review boards: Purpose and challenges. *Chest*, 148(5):1148–1155, 2015.

[16] Yunfei Guo, Tao Zhang, and Wu Huang. Emotion recognition based on physiological signals multi-head attention contrastive learning. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1929–1934. IEEE, 2023.

[17] Zhipeng He, Yongshi Zhong, and Jiahui Pan. An adversarial discriminative temporal convolutional network for eeg-based cross-domain emotion recognition. *Computers in biology and medicine*, 141:105048, 2022.

[18] Yuzhu Hu, Jian Chen, Junxin Chen, Wei Wang, Shen Zhao, and Xiping Hu. An ensemble classification model for depression based on wearable device sleep data. *IEEE Journal of Biomedical and Health Informatics*, 2023.

[19] Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE Journal of Biomedical and Health Informatics*, 22(1):98–107, 2017.

[20] Smith K Khare and U Rajendra Acharya. Adazd-net: Automated adaptive and explainable alzheimer's disease detection system using eeg signals. *Knowledge-Based Systems*, 278:110858, 2023.

[21] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

[22] Yong-Jin Liu, Minjing Yu, Guozhen Zhao, Jinjing Song, Yan Ge, and Yuanchun Shi. Real-time movie-induced discrete emotion recognition from eeg signals. *IEEE Transactions on Affective Computing*, 9(4):550–562, 2017.

[23] Paola Pedrelli, Szymon Fedor, Asma Ghandeharioun, Esther Howe, Dawn F Ionescu, Darian Bhathena, Lauren B Fisher, Cristina Cusin, Maren Nyer, Albert Yeung, et al. Monitoring changes in depression severity using wearable and mobile sensors. *Frontiers in psychiatry*, 11:584711, 2020.

[24] Sara Pourmohammadi and Ali Maleki. Stress detection using ecg and emg signals: A comprehensive study. *Computer methods and programs in biomedicine*, 193:105482, 2020.

[25] Shangran Qiu, Matthew I Miller, Prajakta S Joshi, Joyce C Lee, Chonghua Xue, Yunruo Ni, Yuwei Wang, Ileana De Anda-Duran, Phillip H Hwang, Justin A Cramer, et al. Multimodal deep learning for alzheimer's disease dementia assessment. *Nature communications*, 13(1):3404, 2022.

[26] Stanisław Saganowski, Bartosz Perz, Adam G Polak, and Przemysław Kazienko. Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review. *IEEE Transactions on Affective Computing*, 14(3):1876–1897, 2022.

[27] Alexandre Schaefer, Frédéric Nils, Xavier Sanchez, and Pierre Philippot. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and emotion*, 24(7):1153–1172, 2010.

[28] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.

[29] Tengfei Song, Wenming Zheng, Cheng Lu, Yuan Zong, Xilei Zhang, and Zhen Cui. Mped: A multi-modal physiological emotion database for discrete emotion recognition. *IEEE Access*, 7:12177–12191, 2019.

[30] Zhenxi Song, Bin Deng, Jiang Wang, and Guosheng Yi. An eeg-based systematic explainable detection framework for probing and localizing abnormal patterns in alzheimer's disease. *Journal of Neural Engineering*, 19(3):036007, 2022.

[31] Vimbi Viswan, Noushath Shaffi, Mufti Mahmud, Karthikeyan Subramanian, and Faizal Hajamohideen. Explainable artificial intelligence in alzheimer's disease classification: A systematic review. *Cognitive Computation*, 16(1):1–44, 2024.

[32] Danyang Wang, Jianhao Weng, Yongpan Zou, and Kaishun Wu. Emotracer: A wearable physiological and psychological monitoring system with multi-modal sensors. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, pages 444–449, 2022.

[33] Yixin Wang, Shuang Qiu, Dan Li, Changde Du, Bao-Liang Lu, and Huiguang He. Multi-modal domain adaptation variational autoencoder for eeg-based emotion recognition. *IEEE/CAA Journal of Automatica Sinica*, 9(9):1612–1626, 2022.

[34] Yujin Wu, Mohamed Daoudi, and Ali Amad. Transformer-based self-supervised multimodal representation learning for wearable emotion recognition. *IEEE Transactions on Affective Computing*, 15(1):157–172, 2023.

[35] Guowen Xiao, Meng Shi, Mengwen Ye, Bowen Xu, Zhendi Chen, and Quansheng Ren. 4d attention-based neural network for eeg emotion recognition. *Cognitive Neurodynamics*, pages 1–14, 2022.

[36] Tianyi Zhang, Abdallah El Ali, Chen Wang, Alan Hanjalic, and Pablo Cesar. Weakly-supervised learning for fine-grained emotion recognition using physiological signals. *IEEE Transactions on Affective Computing*, 2022.

[37] Xiaowei Zhang, Xiangyu Wei, Zhongyi Zhou, Qiqi Zhao, Sipo Zhang, Yikun Yang, Rui Li, and Bin Hu. Dynamic alignment and fusion of multimodal physiological patterns for stress recognition. *IEEE Transactions on Affective Computing*, 2023.

[38] Yanteng Zhang, Xiaohai He, Yi Hao Chan, Qizhi Teng, and Jagath C Rajapakse. Multi-modal graph neural network for early diagnosis of alzheimer's disease from smri and pet scans. *Computers in Biology and Medicine*, 164:107328, 2023.

[39] Jing Zhu, Changting Jiang, Junhao Chen, Xiangbin Lin, Ruilan Yu, Xiaowei Li, and Bin Hu. Eeg based depression recognition using improved graph convolutional neural network. *Computers in Biology and Medicine*, 148:105815, 2022.

[40] Yongpan Zou, Danyang Wang, Dan Wang, Canlin Zheng, Qifeng Song, Yuzheng Zhu, Changhe Fan, and Kaishun Wu. Research on wearable emotion recognition based on multi-source domain adversarial transfer learning (in chinese). *Chinese Journal of Computers*, 47(2):266–286, 2024.