Coursework 2	**Emotion Recognition: Benefits and Challenges – Mini Project and Individual Report**
Module	**Affective Computing and Human-Robot Interaction (COMP0053)**

Group 7

Individual Report

Student's code:

ZYWB7

Instructor:

Prof. Nadia Berthouze:

University College London

November 11, 2023

1 Aim of the designed emotion recognition systems for the selected application

This study aims to develop a video emotion recognition system that analyzes physiological data to identify users' emotions while watching short videos [1]. Existing algorithms primarily rely on static preferences and cannot dynamically respond to users' emotional changes. Our proposed system incorporates emotion recognition, offering a dynamic and responsive user experience.

Author's LjUs startup focuses on interactive light therapy lamps to alleviate seasonal affective disorder (SAD) in high-latitude regions [2]. Integrating emotion tracking and recognition can enhance SAD interventions by allowing personalized adjustments to light parameters and therapeutic strategies. Our emotion recognition project for short videos lays the foundation for effective multimodal interaction designs based on physiological signals.

Emotion recognition has potential in gaming, mental health, and learning design technology [3]. Our study contributes to research on ubiquitous computing, emphasizing human-centered computing's importance for enhancing user experience [4]. Although wearable devices face challenges, advances in sensor technology and signal processing techniques address these issues [5, 6].

During system deployment, privacy protection and data security are ethical concerns [2]. Algorithmic imperfections may result in biases, requiring diverse and representative datasets [7]. Collecting emotion-related physiological data raises privacy and data protection issues; adhering to GDPR and relevant laws is essential [8]. Users must be informed about data usage for advertising and provided opt-out options. Explicit consent mechanisms should be employed [9]. Ethical guidelines should be followed to prevent potential issues for users, society, and the environment.

2 Affective states and labelling approach

In our project, we employed two methods to label emotions: self-report by volunteers (weight 0.6) and manual annotation by two fixed observers (weight 0.4). $E = 0.6 \cdot S_{\boldsymbol{self}} + 0.2 \cdot S_{\boldsymbol{obs1}} + 0.2 \cdot S_{\boldsymbol{obs2}}$. We used the Pleasure and Arousal dimensions from the SAM questionnaire and selected three scales for Pleasure and Arousal each (-1, 0, 1) based on prior literature supporting this choice for emotion analysis. However, our preliminary analysis revealed some limitations in the selected affective states and labeling approach.

The chosen affective states, though grounded in research, may not capture the full range of human emotions. Consequently, alternative models, such as the circumplex model or the discrete emotion model, should be explored to ensure a more comprehensive representation of emotions [10, 11].

Regarding the labeling approach, self-reporting can be influenced by personal subjective awareness and emotional context, leading to errors and biases in emotion recognition [12]. Our mini-project results indicated that self-reported labeling was sparse, a challenge for subsequent machine learning analysis [13]. Observers understanding and standards may differ, and relying on visual observation for annotation can result in inaccuracies and biases [14].

To improve the reliability and validity of emotion annotation, we propose employing semi-supervised learning methods, which can use unlabeled data to recognize emotions based on limited labeled datasets. Active learning can be applied to select the most informative unlabeled data for annotation [15]. Researchers have successfully employed semi-supervised learning and active learning to address labeling challenges in EEG-based emotion recognition [16]. Deep learning models with semi-supervised learning can leverage unlabeled data representations to enhance emotion recognition accuracy. Domain adaptation learning techniques can be used to handle different emotional data domains and improve model generalization [17].

To eliminate biases, we recommend aggregating multiple independent annotators' labeling data using techniques such as the Dawid-Skene estimator, which reduces biases and improves consistency [18]. In real-world applications, integrating speech recognition technology, natural language processing, and other physiological signal data (e.g., electromyographic, electroencephalographic, thermal sensing data) can extract higher-dimensional physiological features [19]. This approach allows for a more comprehensive and accurate capture of emotional experiences while considering the dynamic aspects of facial expressions [20].

To improve emotion annotation, use alternative models, semi-supervised learning, multiple annotators, and combine physiological signals and natural language processing.

3 Modality and sensors selection

In our experiment, we primarily utilized respiratory, cardiac-related activity, and facial data for emotion recognition and analysis. To facilitate emotion classification, we averaged the data. However, as emotions are continuous and context-dependent, this method might not be completely accurate in practical applications [21].

3.1 Strengths and limitations of current modalities:

In our experiments, we focused on two main categories of data: physiological and facial. The physiological data includes breathing and heartbeat, while the facial data captures expressions.

Facial expressions: Facial data depends on video quality and Open-Face software's ability to recognize and track facial landmarks. Our 720p front-facing camera could be improved, and lighting control could mitigate potential effects on data. Micro-expressions and cultural variations might not be captured, as individuals may mask their emotions [22].

Physiological data: Supplementing facial data with physiological data reduces bias but can be influenced by factors unrelated to emotions (e.g., physical activity, stress) [23]. Differentiating subtle emotional differences may be challenging, and device selection should consider individual factors (e.g., height, weight, gender) [24]. Accuracy may be affected by textile thickness in respiratory detection equipment.

3.2 Proposed additional modality: Thermal imaging for facial analysis

Prior research indicates that sympathetic activation results in decreased temperature of the nose, attributed to vasoconstriction and blood flow restriction to the skin surface. The upper lip or maxillary region exhibits a similar decrease due to sweat gland activation. However, an increase in temperature was observed in the forehead and the area between the eyes and the nose [25].

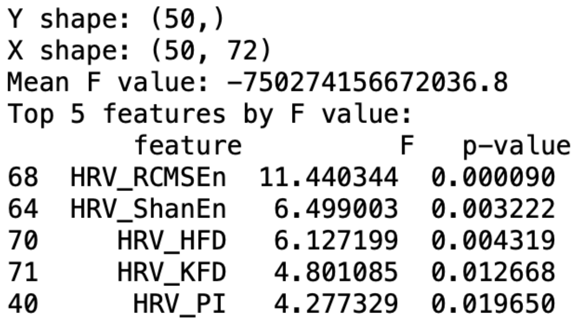
Benefits: Thermal imaging captures facial temperature distribution and detects subtle changes related to blood flow and autonomic nervous system activity, this has been shown to have a strong relationship with human emotions, and the technology requires no contact [26]. It complements video sensors' disadvantages by overcoming limitations such as lighting conditions and cultural variations in emotional expression ability to establish keen data acquisition for some micro-expressions.

Sensor selection: A high-resolution thermal imaging camera is a potential choice for capturing facial temperature changes.

Integrating thermal imaging with physiological data and facial expression data can provide a comprehensive, multi-modal approach to emotion recognition. This combination can overcome individual modality limitations and create a more robust and accurate emotion recognition system.

4 Feature extraction

In data preprocessing, we removed feature columns with excessive zeros and NaN values, resulting in 72 PPG heart rate, 19 RESP respiration, and 30 facial data features. ANOVA showed that breathing features were important in emotion recognition. HRV is closely linked to emotions, with HRV-related heart rate fluctuations associated with emotional responses, and the parasympathetic nervous system playing a role [24]. HRV is also related to respiration rate, with respiratory sinus arrhythmia used as an indicator of cardiac vagal function. Even in our data, the average F value of the preliminary calculated heart rate-related features is very low, we still choose PPG-related features for further analysis, in Figure 1. We selected significant HRV-HFD and HRV-KFD features with p-values less 0.05, especially according to related research they provide important information in nonlinear sequence data [27].



```
Y shape: (50,)
X shape: (50, 72)
Mean F value: -750274156672036.8
Top 5 features by F value:
        feature          F   p-value
68  HRV_RCMSEn  11.440344  0.000090
64  HRV_ShanEn   6.499003  0.003222
70     HRV_HFD   6.127199  0.004319
71     HRV_KFD   4.801085  0.012668
40      HRV_PI   4.277329  0.019650
```

Figure 1: Top five largest F-values of PPG

4.1 Feature 1: HRV-HFD (Heart Rate Variability - Higuchi's Fractal Dimension)

Rationale: HFD is a nonlinear analysis technique that measures the complexity of time series data, such as heart rate variability (HRV) and neuronal bioelectricity. HFD quantifies the fractal properties of data, offering insight into underlying physiological processes [28]. In emotion recognition, HFD's ability to capture subtle changes in heart rate-induced sequential data associated with different emotional states is particularly valuable. Several studies have successfully used HFD to analyze HRV data, demonstrating its potential for assessing autonomic nervous system function and detecting certain medical conditions [29].

Limitations: HRV-HFD's primary limitations include sensitivity to noise and artifacts in HRV data, which may affect the accuracy of the analysis, and the relatively narrow interval of HFD values, making it challenging to differentiate similar emotions among different individuals [28]. Additionally, external factors like physical activity can influence HRV measurements, potentially impacting the reliability of HFD-based emotion recognition.

Computation: [30]

1. Given a time series $x_1, x_2, \dots, x_N$ with N data points, first create k new time series by selecting every k-th data point, starting from the m-th data point:

$$X^{(k,m)} = \{x_m, x_{m+k}, x_{m+2k}, \dots, x_{m+\lfloor \frac{N-m}{k} \rfloor k}\} \tag{1}$$

   where $m = 1, 2, \dots, k$ and $\lfloor \cdot \rfloor$ denotes the floor function.

2. Calculate the length of each new time series $X^{(k,m)}$:

$$L_m(k) = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{\lfloor \frac{N-m}{k} \rfloor} \sum_{j=1}^{\lfloor \frac{N-m}{k} \rfloor} \left| x_{m+(j-1)k} - x_{m+jk} \right| \tag{2}$$

3. Compute the average length for each k:

$$L(k) = \frac{1}{k} \sum_{m=1}^{k} L_m(k) \tag{3}$$

4. Estimate the fractal dimension by analyzing the relationship between $L(k)$ and k. Theoretically, we expect that:

$$L(k) \propto k^{-D} \tag{4}$$

   where D is the Higuchi Fractal Dimension.

5. To estimate the value of D, we can take the logarithm of both sides and perform a linear regression:

$$\log L(k) = -D \log k + C \tag{5}$$

   where C is a constant.

6. The slope of the linear regression line represents the Higuchi Fractal Dimension D.

4.2 Feature 2: HRV-KFD (Heart Rate Variability - Katz's Fractal Dimension)

Rationale: The use of KFD as a feature is justified by its relevance to non-linear analysis methods and its potential value in evaluating autonomic nervous system function, identifying medical conditions, and differentiating emotional states. This is supported by relevant citations [31, 32].

Limitations: Limitations of KFD are discussed, including its sensitivity to noise and exponential transformation compared to HFD, supported by relevant sources [33]. Numerical accuracy issues due to high floating-point operations are also addressed.

Computation: [31]

1. Given a time series $x_1, x_2, \ldots, x_N$ with N data points, first calculate the cumulative sum of the time series:

$$y_i = \sum_{j=1}^{i} x_j \quad \text{for} \quad i = 1, 2, \ldots, N \tag{6}$$

2. Determine the total length of the cumulative sum time series, L. This is the sum of the distances between consecutive points:

$$L = \sum_{i=1}^{N-1} \sqrt{1 + (y_{i+1} - y_i)^2} \tag{7}$$

3. Calculate the diameter of the cumulative sum time series, d. This is the largest distance between any two points in the time series:

$$d = \max_{1 \leq i,j \leq N} |y_j - y_i| \tag{8}$$

4. Determine the length of the straight line connecting the first and last points of the cumulative sum time series, a. This can be calculated using the Euclidean distance formula:

$$a = \sqrt{(N-1)^2 + (y_N - y_1)^2} \tag{9}$$

5. Calculate Katz's Fractal Dimension, D, using the following formula:

$$D = \frac{\log_{10}(L/a)}{\log_{10}(d)} \tag{10}$$

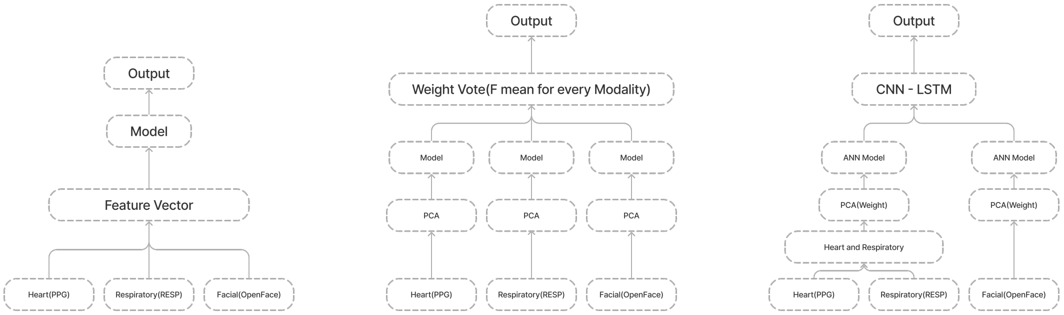
5 Modeling, Fusion and Evaluation

5.1 Modelling:

PCA can reduce dimensionality while retaining 0.99 of variance for each modality, and alternative feature extraction techniques can be explored. Grayscale information, edge maps, and area image enhancement can detect emotionally relevant areas in facial recognition [34, 35]. PPG reflects blood volume changes driven by the cardiac cycle, with peak detection using the Pan-Tompkins algorithm [36, 37]. For respiratory data, bandpass filters can eliminate non-respiratory frequencies, followed by complex filtering and sampling techniques [38]. These methods capture a comprehensive range of emotional patterns.

Various machine learning models can classify emotions, including traditional models like Random Forest, Support Vector Machines (SVM), and k-Nearest Neighbors (KNN) [39]. Deep learning models, such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks can also be explored to capture more complex patterns and adapt better to the challenges posed by the dataset, in particular, deeper CNNs are significantly more accurate than other models in related studies[40, 41, 42].

5.2 Fusion:



(a) Early Fusion

(b) Mid Fusion

(c) Late Fusion

Figure 2: Multimodal Fusion Strategies

To address the challenges posed by the dataset, we propose a robust early-mid-late fusion strategy in Figure 2, combining PPG, RESP, and OpenFace data. The **early fusion** strategy merges features from different sensors into a single feature vector and trains them using machine learning algorithms such as SVM and KNN [43]. The **mid-fusion** strategy applies PCA for dimensionality reduction and trains individual classifiers based on features from different data sources, using majority or weighted voting for output prediction results [44, 45]. The **late fusion** strategy, utilizing ANOVA testing for feature significance and F-value-based sensor weighting, reduces dimensionality and employs fully connected networks for preprocessing before deep learning prediction. Importantly, recent studies highlight improved affective computing model performance by combining heart and respiratory systems during emotion elicitation[46]. This robust fusion approach effectively manages multimodal data complexities and uncertainties, leveraging integrative physiological features for enhanced emotion recognition[45].

5.3 Evaluation:

Performance Metrics:

Accuracy, F1-score, and confusion matrix are essential metrics for evaluating the classifier's ability in multimodal emotion recognition. In addition, we could consider incorporating metrics such as precision, recall, and area under the receiver operating characteristic curve (AUROC) to gain a more comprehensive understanding of the model's performance across different levels of sensitivity and specificity [47, 48].

Cross-validation:

We propose using LOSO CV, a k-fold cross-validation approach that enhances generalization to unseen data by accounting for individual differences [49]. Training on data from different subjects ensures the model can recognize entirely new data in the validation set, and has proven effective in some emotion recognition models [49]. Additionally, LOSO CV can prevent overfitting to some extent by avoiding identical data in the training and testing sets [50]. This approach results in a more robust and reliable model that can handle each individual's unique emotional patterns.

Benchmarking:

We identified relevant literature on PPG, RESP, and OpenFace for benchmarking our multimodal emotion recognition system [51], confirming the accuracy of our random forest and CNN models. We will continue to use performance metrics such as accuracy, F1 score, confusion matrix, precision, recall, and AUROC to evaluate and optimize our system.

References

[1] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston *et al.*, “The youtube video recommendation system,” in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 293–296.

[2] R. A. Calvo, S. D’Mello, J. M. Gratch, and A. Kappas, *The Oxford handbook of affective computing*. Oxford Library of Psychology, 2015.

[3] R. A. Calvo and S. D’Mello, “Affect detection: An interdisciplinary review of models, methods, and their applications,” *IEEE Transactions on affective computing*, vol. 1, no. 1, pp. 18–37, 2010.

[4] G. D. Abowd and E. D. Mynatt, “Charting past, present, and future research in ubiquitous computing,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 7, no. 1, pp. 29–58, 2000.

[5] J. A. Healey and R. W. Picard, “Detecting stress during real-world driving tasks using physiological sensors,” *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.

[6] S. Park and S. Jayaraman, “Wearables: Fundamentals, advancements, and a roadmap for the future,” in *Wearable sensors*. Elsevier, 2021, pp. 3–27.

[7] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, “Emotion recognition in human-computer interaction,” *IEEE Signal processing magazine*, vol. 18, no. 1, pp. 32–80, 2001.

[8] B.-J. Koops, B. C. Newell, T. Timan, I. Skorvanek, T. Chokrevski, and M. Galic, “A typology of privacy,” *U. Pa. J. Int’l L.*, vol. 38, p. 483, 2016.

[9] C. Fiesler and N. Proferes, ““participant” perceptions of twitter research ethics,” *Social Media+ Society*, vol. 4, no. 1, p. 2056305118763366, 2018.

[10] J. A. Russell, “A circumplex model of affect.” *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[11] P. Ekman, “An argument for basic emotions,” *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[12] L. F. Barrett, B. Mesquita, and M. Gendron, "Context in emotion perception," *Current directions in psychological science*, vol. 20, no. 5, pp. 286–290, 2011.

[13] L. Romeo, A. Cavallo, L. Pepa, N. Bianchi-Berthouze, and M. Pontil, "Multiple instance learning for emotion recognition using physiological signals," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 389–407, 2019.

[14] T. Bänziger, M. Mortillaro, and K. R. Scherer, "Introducing the geneva multimodal expression corpus for experimental research on emotion perception." *Emotion*, vol. 12, no. 5, p. 1161, 2012.

[15] B. Settles, "Active learning literature survey," 2009.

[16] X. Jia, K. Li, X. Li, and A. Zhang, "A novel semi-supervised deep learning framework for affective state recognition on eeg signals," in *2014 ieee international conference on bioinformatics and bioengineering*. IEEE, 2014, pp. 30–37.

[17] J. Deng, Z. Zhang, F. Eyben, and B. Schuller, "Autoencoder-based unsupervised domain adaptation for speech emotion recognition," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1068–1072, 2014.

[18] A. P. Dawid and A. M. Skene, "Maximum likelihood estimation of observer error-rates using the em algorithm," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 28, no. 1, pp. 20–28, 1979.

[19] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[20] E. G. Krumhuber, A. Kappas, and A. S. Manstead, "Effects of dynamic aspects of facial expressions: A review," *Emotion Review*, vol. 5, no. 1, pp. 41–46, 2013.

[21] A. Metallinou and S. Narayanan, "Annotation and processing of continuous emotional attributes: Challenges and opportunities," in *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*. IEEE, 2013, pp. 1–8.

[22] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." *Journal of personality and social psychology*, vol. 17, no. 2, p. 124, 1971.

[23] K. R. Scherer, "What are emotions? and how can they be measured?" *Social science information*, vol. 44, no. 4, pp. 695–729, 2005.

[24] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological psychology*, vol. 84, no. 3, pp. 394–421, 2010.

[25] S. Ioannou, V. Gallese, and A. Merla, "Thermal infrared imaging in psychophysiology: potentialities and limits," *Psychophysiology*, vol. 51, no. 10, pp. 951–963, 2014.

[26] I. Pavlidis, N. L. Eberhardt, and J. A. Levine, "Seeing through the face of deception," *Nature*, vol. 415, no. 6867, pp. 35–35, 2002.

[27] M. T. V. Yamuza, J. Bolea, M. Orini, P. Laguna, C. Orrite, M. Vallverdu, and R. Bailon, "Human emotion characterization by heart rate variability analysis guided by respiration," *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2446–2454, 2019.

[28] S. Kesić and S. Z. Spasić, "Application of higuchi's fractal dimension from basic to clinical neurophysiology: A review," *Computer methods and programs in biomedicine*, vol. 133, pp. 55–70, 2016.

[29] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, pp. 1031–1051, 2006.

[30] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.

[31] M. J. Katz, "Fractals and the analysis of waveforms," *Computers in biology and medicine*, vol. 18, no. 3, pp. 145–156, 1988.

[32] G. Tamulevičius, R. Karbauskaitė, and G. Dzemyda, "Speech emotion classification using fractal dimension-based features," *Nonlinear Analysis: Modelling and Control*, vol. 24, no. 5, pp. 679–695, 2019.

[33] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, "A comparison of waveform fractal dimension algorithms," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 2, pp. 177–183, 2001.

[34] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, vol. 30, no. 4, pp. 1334–1345, 2007.

[35] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," *Signal processing: Image communication*, vol. 12, no. 3, pp. 263–281, 1998.

[36] M. Elgendi, R. Fletcher, Y. Liang, N. Howard, N. H. Lovell, D. Abbott, K. Lim, and R. Ward, "The use of photoplethysmography for assessing hypertension," *NPJ digital medicine*, vol. 2, no. 1, p. 60, 2019.

[37] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.

[38] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, R. Beale, and P. J. Watkinson, "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiological measurement*, vol. 37, no. 4, p. 610, 2016.

[39] R. A. Nugrahaeni and K. Mutijarsa, "Comparative analysis of machine learning knn, svm, and random forests algorithm for facial expression classification," in *2016 International Seminar on Application for Technology of Information and Communication (ISemantic)*. IEEE, 2016, pp. 163–168.

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[41] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[42] Y. Cho, N. Bianchi-Berthouze, and S. J. Julier, “Deepbreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings,” in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2017, pp. 456–463.

[43] S. M. Alarcao and M. J. Fonseca, “Emotions recognition using eeg signals: A survey,” *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 374–393, 2017.

[44] R. J. Barry and F. M. De Blasio, “Eeg frequency pca in eeg-erp dynamics,” *Psychophysiology*, vol. 55, no. 5, p. e13042, 2018.

[45] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, “Multimodal fusion for multimedia analysis: a survey,” *Multimedia systems*, vol. 16, pp. 345–379, 2010.

[46] G. Valenza, A. Lanata, and E. P. Scilingo, “Oscillations of heart rate and respiration synchronize during affective visual stimulation,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 683–690, 2012.

[47] M. Soleymani, M. Pantic, and T. Pun, “Multimodal emotion recognition in response to videos,” *IEEE transactions on affective computing*, vol. 3, no. 2, pp. 211–223, 2011.

[48] J. Wei, E. Pei, D. Jiang, H. Sahli, L. Xie, and Z. Fu, “Multimodal continuous affect recognition based on lstm and multiple kernel learning,” in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*. IEEE, 2014, pp. 1–4.

[49] F. Hou, Q. Gao, Y. Song, Z. Wang, Z. Bai, Y. Yang, and Z. Tian, “Deep feature pyramid network for eeg emotion recognition,” *Measurement*, vol. 201, p. 111724, 2022.

[50] P. DeSouza, R. Kahn, T. Stockman, W. Obermann, B. Crawford, A. Wang, J. Crooks, J. Li, and P. Kinney, “Calibrating networks of low-cost air quality sensors,” *Atmospheric Measurement Techniques*, vol. 15, no. 21, pp. 6309–6328, 2022.

[51] C. Filippini, A. Di Crosta, R. Palumbo, D. Perpetuini, D. Cardone, I. Ceccato, A. Di Domenico, and A. Merla, “Automated affective computing based on bio-signals analysis and deep learning approach,” *Sensors*, vol. 22, no. 5, p. 1789, 2022.

Coursework 1	**Emotion Recognition: Benefits and Challenges – Mini Project and Group Report**
Module	**Affective Computing and Human-Robot Interaction (COMP0053)**

Group 7

Group Report for Affective Computing and Human-Robot Interaction Mini-Project

List of students' codes:

XKKP1
ZYWB7
WZNR2
ZYJY7
XQMP3
BTBC6

University College London

November 11, 2023

1 Part 1

1.1 *Aim of the emotion recognition system*

Our Affect Recognition System (ARS) aims to recognize emotions while individuals watch different kinds of videos. This technology can provide valuable insights into how individuals respond to different stimuli and has various applications in industries such as healthcare, marketing, and entertainment.

In our mini-project, we recruited volunteers to collect physiological and self-reported emotional data while they were watching short videos labeled with affective states. Physiological data and facial expression, while participants self-reported their emotions during video watching. The system is trained to recognize three affective states - fear, joy, and no emotion - and is capable of accurately identifying an individual's emotional state as one of these three categories.

1.2 *Data collection*

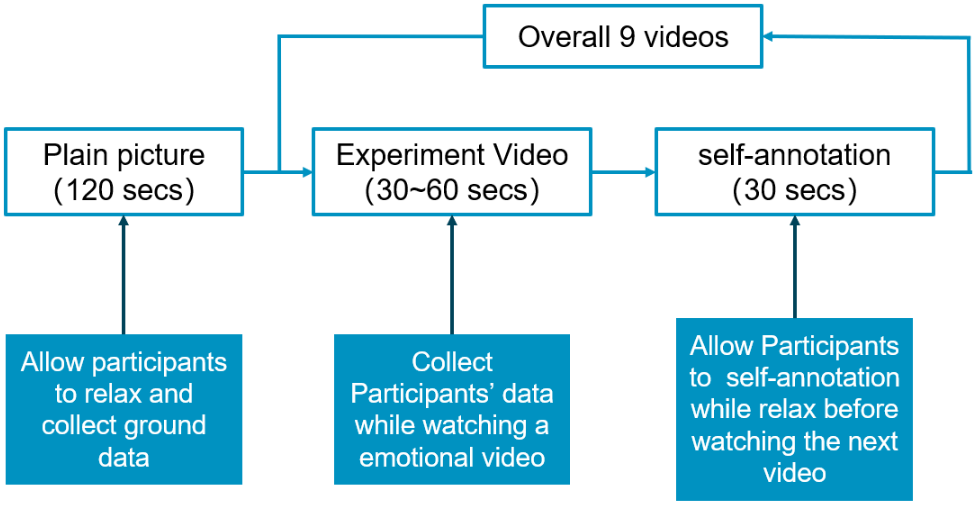
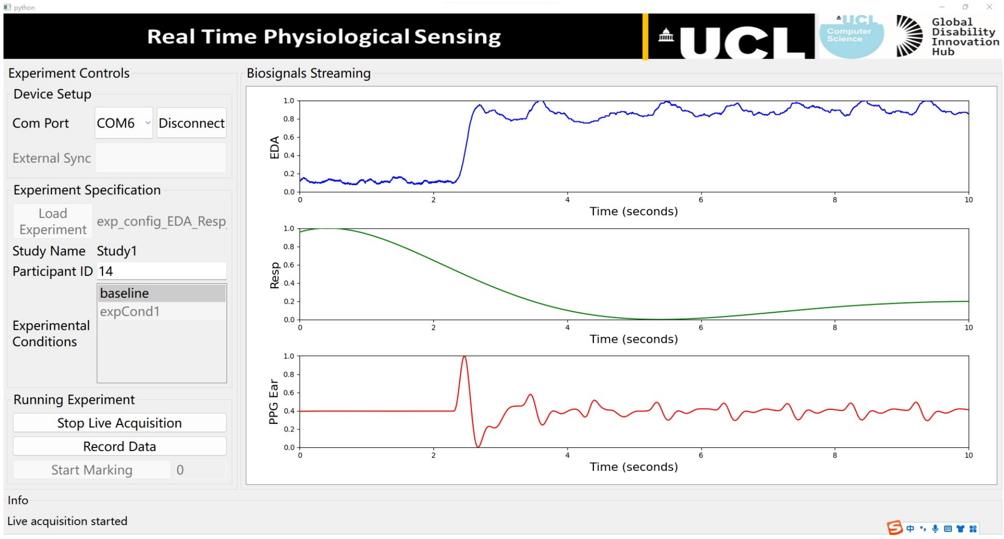

Overall 9 videos
Plain picture (120 secs)
Experiment Video (30~60 secs)
self-annotation (30 secs)
Allow participants to relax and collect ground data
Collect Participants' data while watching a emotional video
Allow Participants to self-annotation while relax before watching the next video

Figure 1: Experiment design flow chart

The project's data is divided into psychological and facial data. Physiological data is collected by `physiologicAlab`[1] using multiple sensors such as the Respiration belt, Galvanic Skin Response (GSR), and Photo-plethysmography (PPG), as depicted in the accompanying figure. Meanwhile, facial data is recorded by a camera and analyzed by `OpenFace`[2]. Our supervisor provided the equipment, including the mentioned sensors, with our primary focus on Respiration and PPG as the physiological data sources. GSR's influence is negligible since participants did not move their hands and fingers while resting them on the table during the experiment.



Real Time Physiological Sensing

Experiment Controls

Device Setup

Com Port COM6

External Sync

Experiment Specification

Load Experiment exp_config_EDA_Resp

Study Name Study1

Participant ID 14

Experimental Conditions baseline
expCond1

Running Experiment

Info

Live acquisition started

Figure 2: Interface to collect physiological data

The experimental design, illustrated in Figure 1, was carefully crafted to ensure a robust and rigorous approach. The design features nine videos and one static image, specifically a calming forest picture[3]. To enhance the experiment's continuity, these media components were meticulously edited into a single, seamless video. The labeled videos were obtained from a reputable online database[4]. To maintain ethical standards, our video selection excluded content featuring violence, crime, or gore that might make participants feel uncomfortable. Instead, we focused on emotions that could be classified according to the V-A dimension[5]. Consequently, the final video set featured three emotions: joy, fear, and neutral, with three videos per emotion category.

Prior to commencing the experiment, participants were required to submit an Ethic certification[6] and group members received GDPR[7] training to ensure compliance with data protection regulations. The participants were informed about the distressing nature of some video stimuli and were given the option to decline participation.

To establish a baseline, participants were presented with a plain image for 120 seconds to allow them to relax while the sensors collected data. The nine experiment videos were then played in a predetermined order, with each video lasting 30 to 60 seconds and followed by a 30-second period for self-annotation. The first six videos were designed to elicit fear and joy emotions, while the last three were no-emotion videos.

All resulting data and experiment videos were securely stored in UCL One Drive, including experiment videos, raw and processed data, and labeling.

1.3 *Labelling*

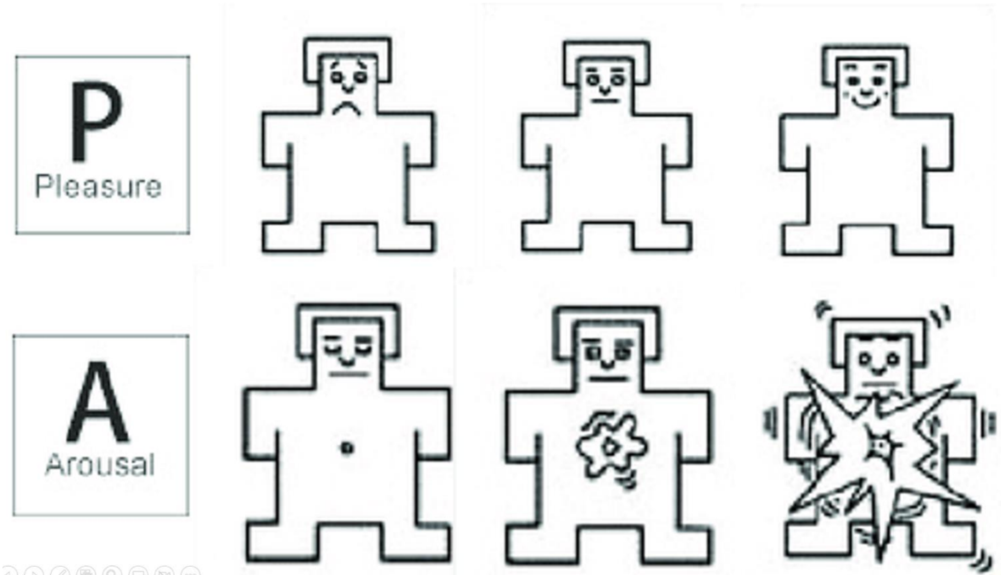


Figure 3: Annotation Model

The annotation process of this experiment relied heavily on Figure 3, which depicts two distinct label types: valence (pleasure) and arousal. A higher valence score indicates greater pleasure derived from the video, while a higher arousal score suggests a more significant impact from the video. The original annotation model, shown in Figure 4, comprised three label types and ten levels[5]. However, we simplified the model to only two labels with three levels for this experiment, as the original model was deemed overly complex and unnecessary for our data requirements. This approach allowed for greater accuracy and ease of understanding for both participants and observers, with scores of '-1,' '0,' and '1' used to represent the states (left to right: -1, 0, 1).

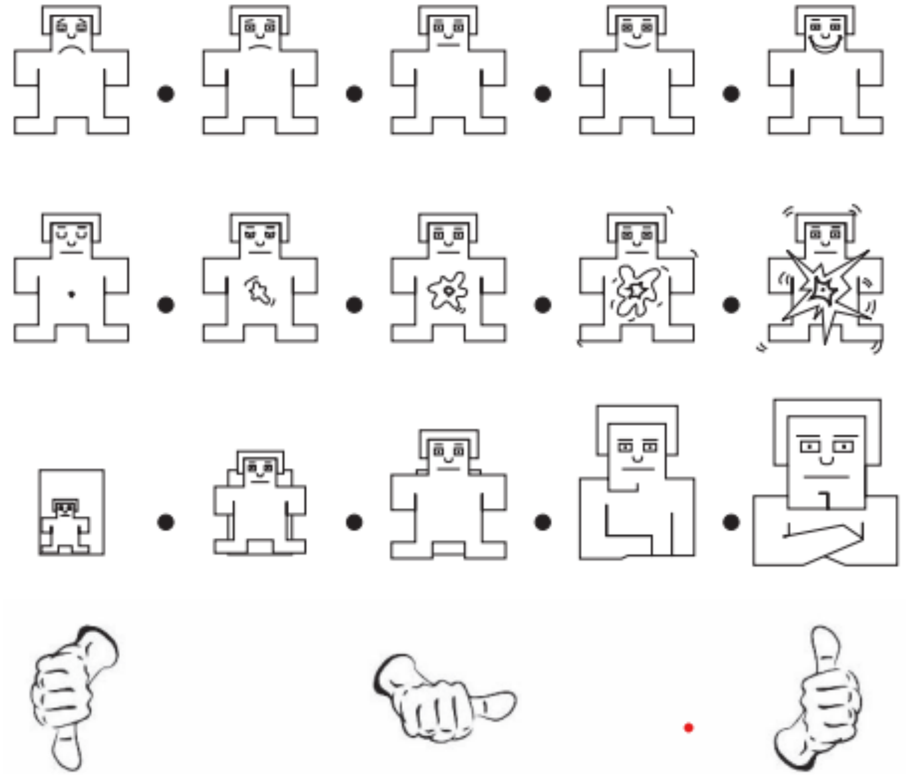


Figure 4: Image for self assessment[5]

During the labeling process, both self-annotation and observer annotation were utilized to generate the final labeled results. Initially, labeling data was collected from two participants, and it was found that self-annotation was more consistent than observer annotation. This was due to the fact that observers primarily based their labeling on the participants' facial expressions, which can sometimes be unreliable. Therefore, two observers were employed, each contributing 0.2 weight to the final labeled result, while self-annotation contributed 0.6 weight.

1.4 *Analysis*

In total, data was collected from 13 participants, and the data analysis will be presented based on modalities. The data was stored in excel tables and UCL one drive, with two strategies employed to manage the data: one for individual videos and the other for data organized by emotion types, as explained in Section 1.4.2 on physiological data.

1.4.1 *Labeling*

The labeling data was obtained from each video and also by averaging the labeling for each emotion. The former method is depicted in Figure5 while the latter is shown in Figure6.

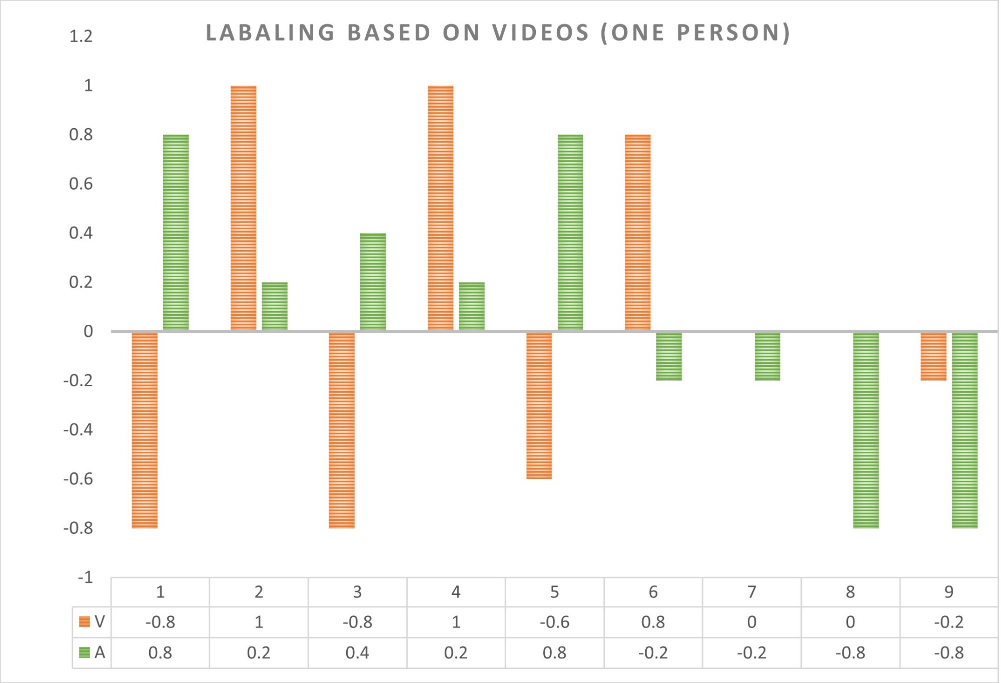


Figure 5: The labeling data for one participant

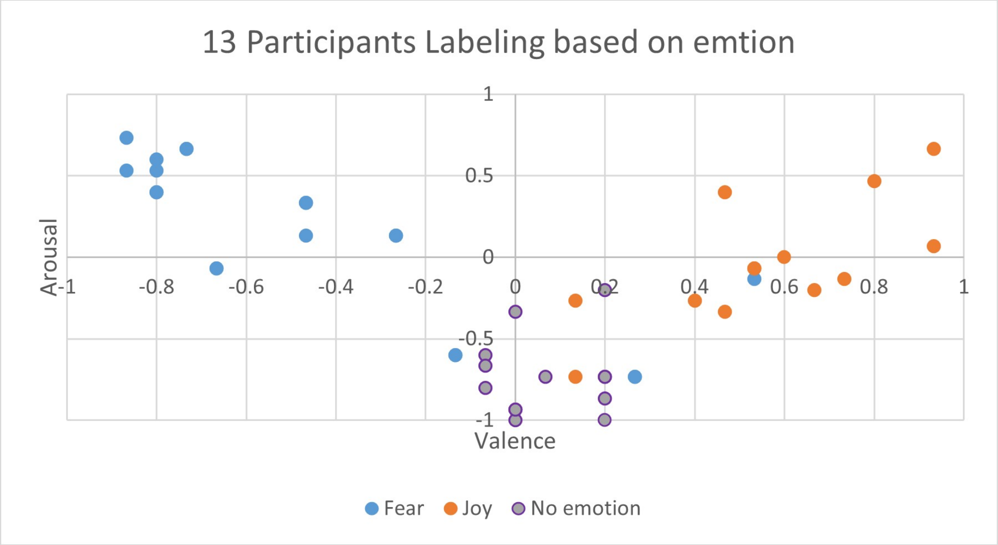


Figure 6: The labeling data for all participants

1.4.2 *Physiological data*

The physiological data is analyzed by the build-in scripts in `PhysiologicAlab`[1]. The baseline data and experiment data was gathered separately. As a result there are two output excel files for each participants.

Cardiovascular signature Physiological data was analyzed using the built-in scripts in `PhysiologicAlab`[1]. Baseline data and experiment data were analyzed separately, and each participant generated two output excel files. After loading the data, which is shown in Figure 7, the sample rate of 230 was determined, and the PPG data was filtered, which is shown in Figure 8 and processed using the Neurokit library. The peak of PPG and the PPG rate were computed, as shown in Figure 9 and Figure 10. Finally, the HRV features, which is shown in Figure 11 were extracted and stored in an excel table, and some features with NaN values were deleted.

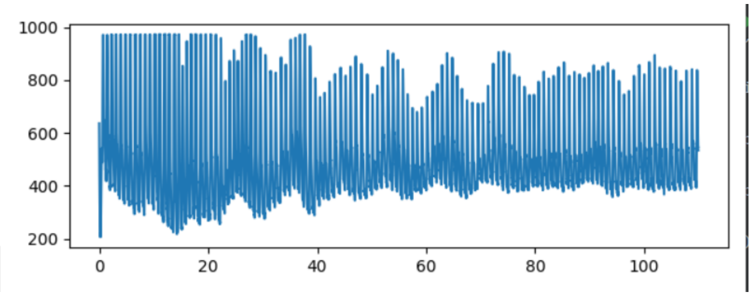


Figure 7: PPG data for baseline

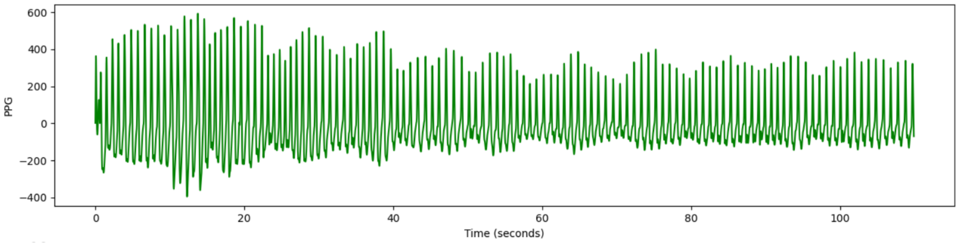


Figure 8: PPG data for baseline after filtering

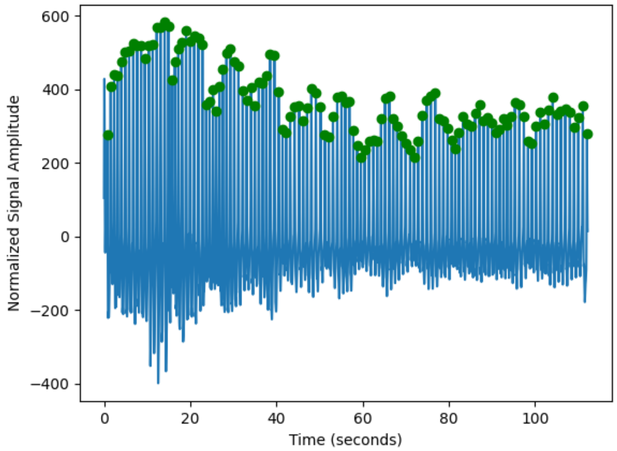


Figure 9: PPG peak for baseline after processing

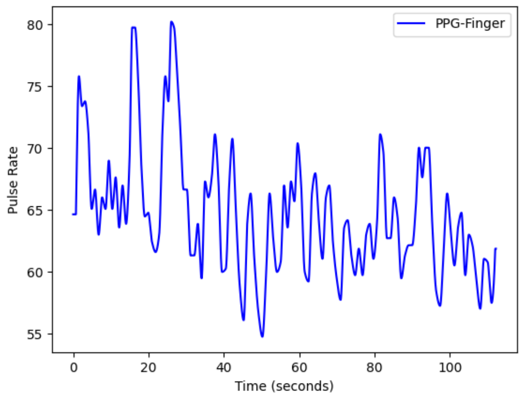


Figure 10: PPG rate for baseline after processing

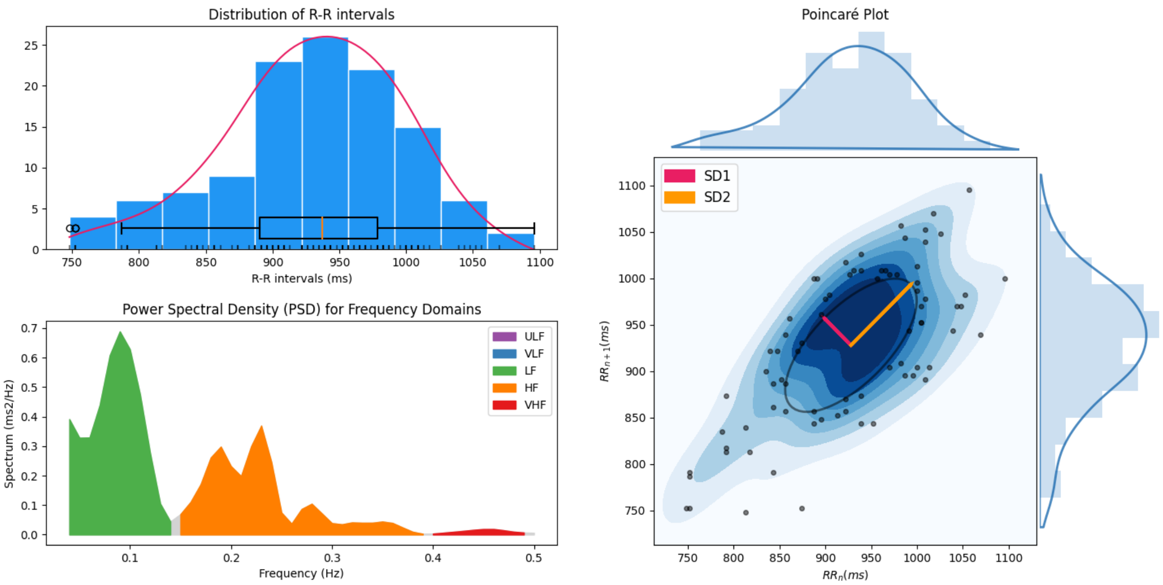


Figure 11: HRV feature

In the experimental phase, the PPG data is extracted using a similar approach. Due to continuous data collection, we extract relevant clips based on the timeline of the entire video. For instance, for the first video from 2:00 to 3:00, we uniformly separate the data and extract relevant data for a specific period. The remaining steps are the same as the baseline, and the HRV features are also stored with the baseline data.

Respiratory signature The respiratory data was analyzed using a similar approach as the cardiovascular data. We utilized built-in scripts and followed the same data collection process as the PPG. For instance, we loaded the baseline data and sampled it at a rate of 230. The raw data is shown in Figure 12. After filtering, we used the Neurokit library to extract respiratory features. Figure 13 displays the respiratory peak, and Figure 14 displays the respiration rate. Additionally, Figure 15 and Figure 16 show other respiratory distribution features.

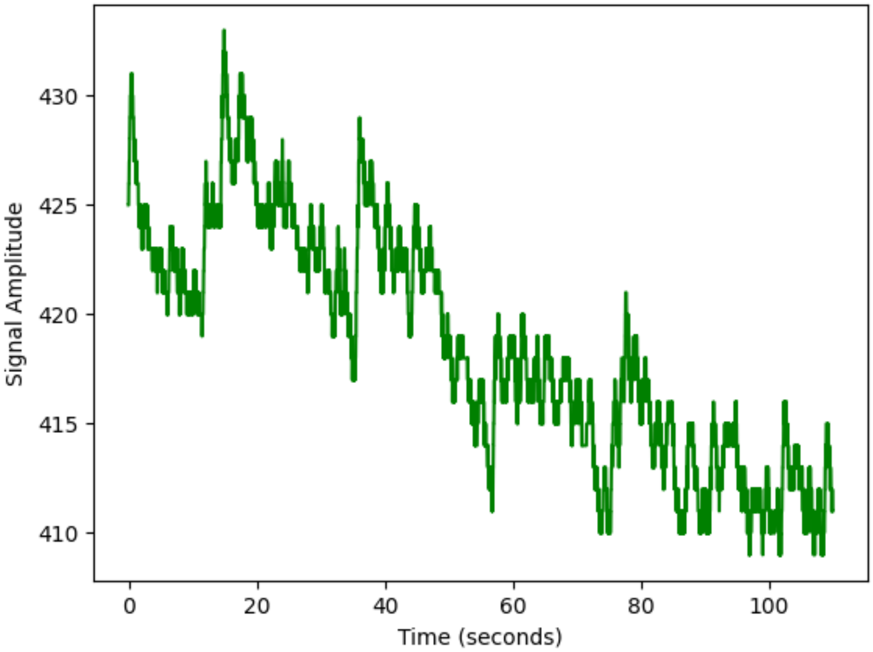


Figure 12: Raw respiratory for baseline

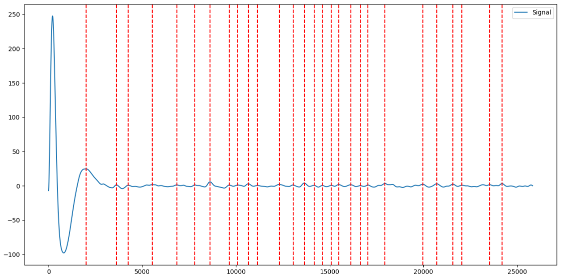


Figure 13: Respiratory peak for baseline

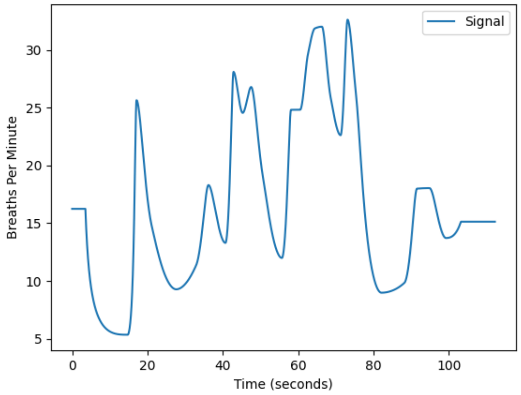


Figure 14: Respiratory rate for baseline

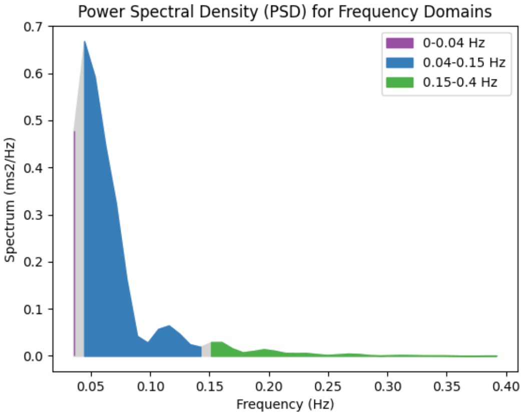


Figure 15: Respiratory rate for baseline in frequency domain

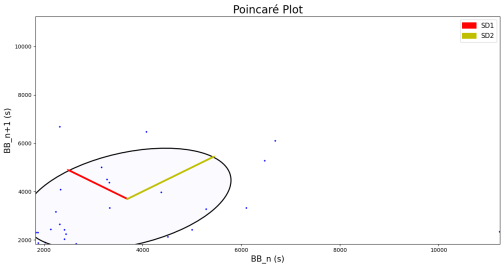
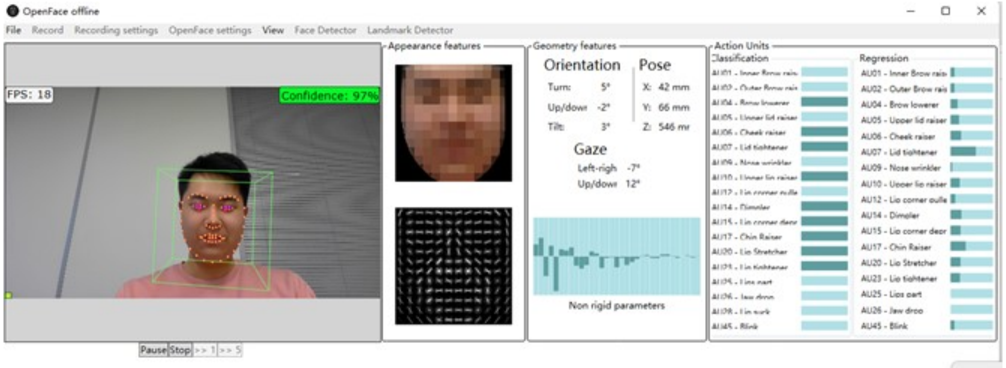


Figure 16: Respiratory rate standard deviation for baseline

In order to analyze the experimental data, we initially attempted to extract features per video. However, due to some videos having a short timeline, the collected data was insufficient for analysis and resulted in outliers and errors, including some cases with no values. To address this, we explored an alternative approach based on emotions. We split the data for each video and integrated the data for videos with the same emotion. Then, we analyzed them simultaneously to obtain results for a participant when watching a particular type of video, such as fear. The processed data for both methods were stored, and features with NaN values were removed.

1.4.3 Facial Expression

The facial expression data was extracted using Openface, and the data for each participant was stored in an Excel file frame by frame. The data was down-sampled to one row per second, and the data was extracted for when participants were watching videos. Using the Action Unit (AU) descriptions, the data for each video and each emotion was extracted and stored in an Excel file. Figure 17 shows the Openface interface, and Figure 18 displays the AU descriptions.



Turn: 5°
Up/down -2°
Tilt: 3°
X: 42 mm
Y: 66 mm
Z: 546 mr
Left-righ -7°
Up/down 12°

Figure 17: Open face interface for extracting facial data

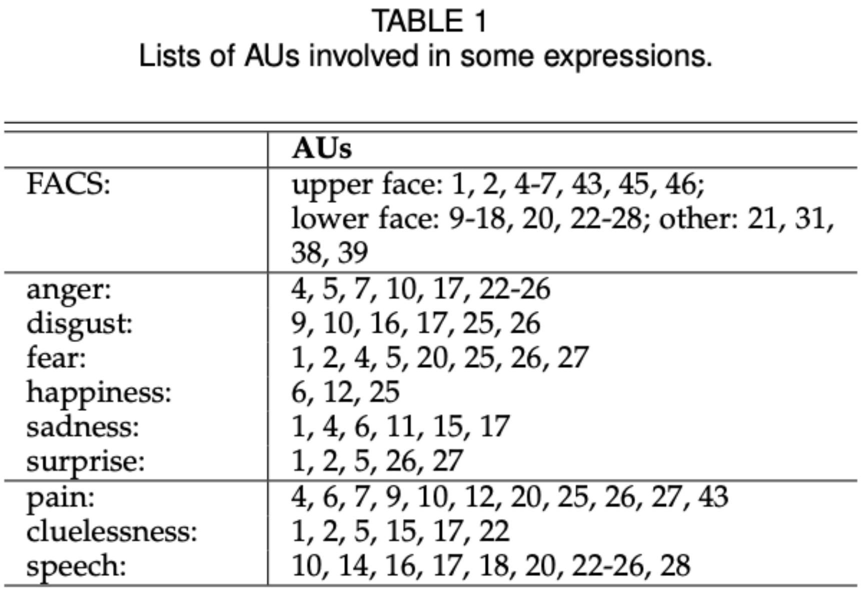


TABLE 1
Lists of AUs involved in some expressions.

	AUs
FACS:	upper face: 1, 2, 4-7, 43, 45, 46; lower face: 9-18, 20, 22-28; other: 21, 31, 38, 39
anger:	4, 5, 7, 10, 17, 22-26
disgust:	9, 10, 16, 17, 25, 26
fear:	1, 2, 4, 5, 20, 25, 26, 27
happiness:	6, 12, 25
sadness:	1, 4, 6, 11, 15, 17
surprise:	1, 2, 5, 26, 27
pain:	4, 6, 7, 9, 10, 12, 20, 25, 26, 27, 43
cluelessness:	1, 2, 5, 15, 17, 22
speech:	10, 14, 16, 17, 18, 20, 22-26, 28

Figure 18: AUs for emotion[8]

1.4.4 Discussion

Compared the Figure 6 and affective dimensions[9, 10], Figure 19, our labeling for fear, joy and no emotion is approximately aliened on the same dimension as the affective space. According to the aims of the ARS system, we could use our data to identify the valence and arousal value. This could help us identify the emotion of participant through the quadrant.

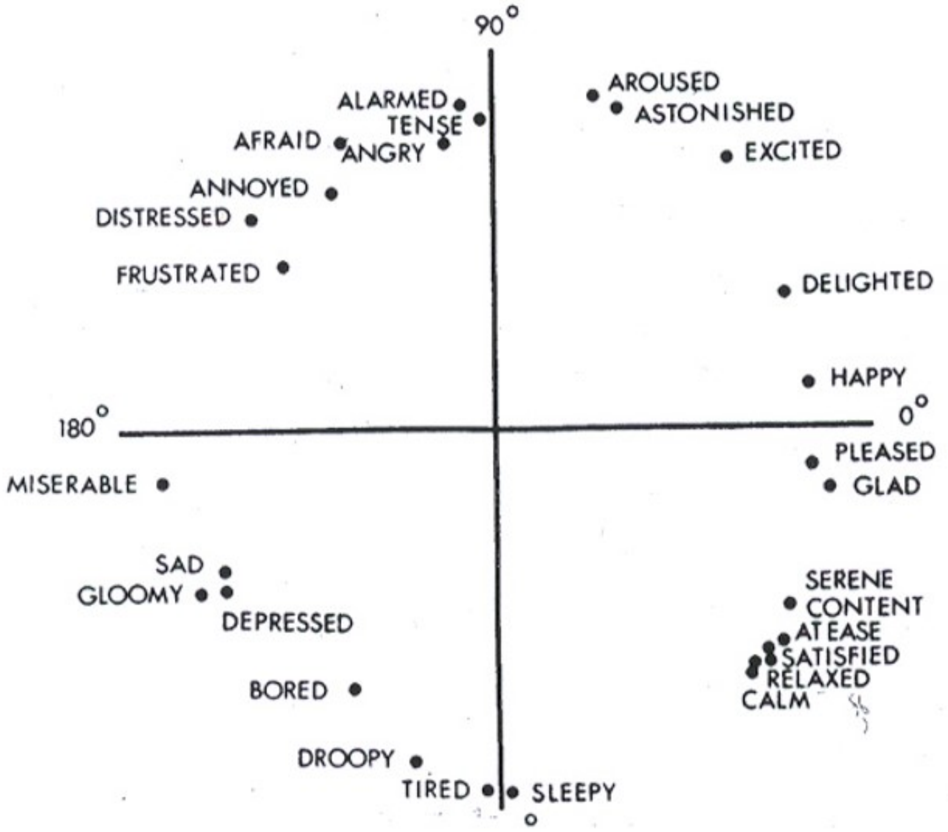


Figure 19: Two dimensions affective space[8]

In Ferdinando's research[11], HRV and EEG were used to evaluate valence and arousal, while in Qiang's research[12], deep learning methods were applied to RSP and PPG data to compute these values. Our data can be processed similarly to identify emotions. Additionally, these researches have utilized the Facial Action Coding System (FACS) [13] and action units (AUs) to identify emotions, with frequencies of AUs being evaluated to predict the emotion. Facial data can directly estimate emotion state, offering an alternative to physiological data analysis. AUs and their corresponding facial actions are listed in Figure 20.

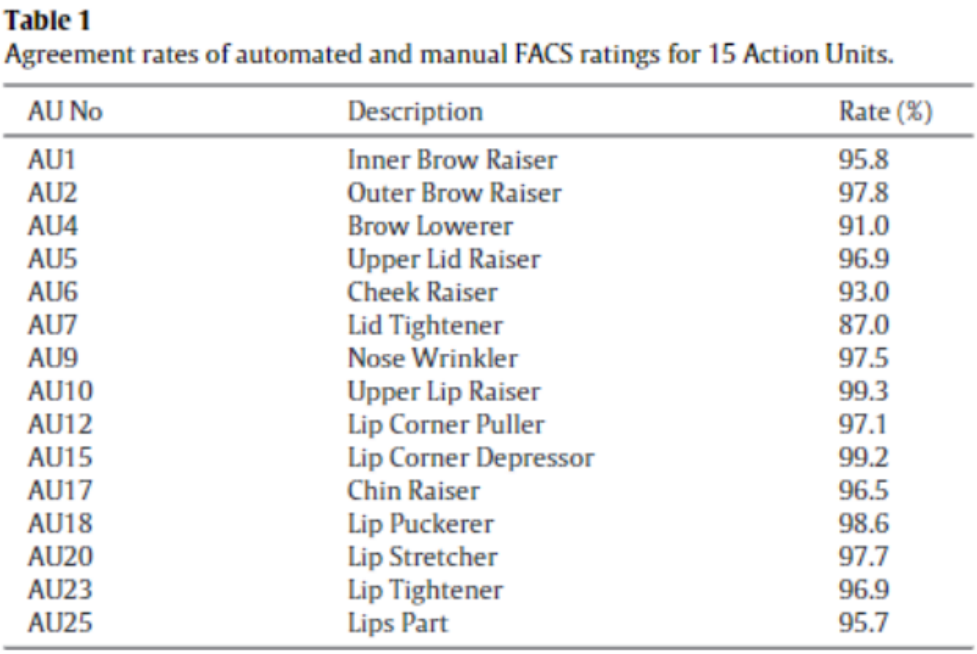


Table 1
Agreement rates of automated and manual FACS ratings for 15 Action Units.

AU No	Description	Rate (%)
AU1	Inner Brow Raiser	95.8
AU2	Outer Brow Raiser	97.8
AU4	Brow Lowerer	91.0
AU5	Upper Lid Raiser	96.9
AU6	Cheek Raiser	93.0
AU7	Lid Tightener	87.0
AU9	Nose Wrinkler	97.5
AU10	Upper Lip Raiser	99.3
AU12	Lip Corner Puller	97.1
AU15	Lip Corner Depressor	99.2
AU17	Chin Raiser	96.5
AU18	Lip Puckerer	98.6
AU20	Lip Stretcher	97.7
AU23	Lip Tightener	96.9
AU25	Lips Part	95.7

Figure 20: The description of AUs[14]

There are error in No.2 participants and missing baseline in NO.8 participants, we decided to ignore them and noted them with yellow.

2 Part 2

2.1 *Aim of the emotion recognition system*

The EmoPain dataset[15] is designed to explore the correlation between emotional states and physical behaviors, with a particular emphasis on detecting protective movements. Emotions are known to impact human behavior and physical motion, and this dataset aims to classify and recognize motion data and protective behaviors. Machine learning models can be trained on the dataset to autonomously detect whether physical behaviors are protective movements. The dataset comprises 66 features collected from various body locations, and additional electromyography data from lumbar paraspinal and upper trapezius muscles, along with a feature indicating the presence of protective behavior. By analyzing this dataset, researchers can gain deeper insights into the relationship between emotions and behavior, thus enhancing our understanding of human emotions and behaviors.

2.2 *Modelling approach*

Data preprocessing: ANOVA To build an accurate model, we must preprocess the data by calculating ANOVA F-values and p-values for each feature in relation to the target variable and generating a correlation matrix. This analysis helps select the most relevant features for modelling.

Modelling After data preprocessing, the construction of models is necessary. The models listed below will be used to train and test in our dataset and then compared to discover the best model:

- CNN

Our developed CNN(Convolutional Neural Network) model comprises of two convolutional layers with 32 feature maps each, a 3 kernel size, and Rectified Linear Unit (ReLU) activation function. The model also includes two max-pooling layers with size 2, a flattening layer, and a fully connected layer containing 32 neurons. To enhance the model's generalization ability, we introduced a 20% dropout rate. For binary classification, the output layer employs the Sigmoid activation function with one neuron.

For training configuration, we selected critical parameters, including binary cross-entropy loss, Adam optimizer with 0.0001 learning rate, and accuracy as the evaluation metric. These settings were chosen to optimize the model's performance for binary classification.

We chose a CNN model for its effectiveness in image classification. This architecture can learn relevant features from input data and classify images with high accuracy by using convolutional layers to extract relevant features.

- Random forest

Random forest is an ensemble learning algorithm that builds multiple decision trees to improve prediction accuracy. Parameters, including the number of trees, maximum depth, and minimum samples per leaf, can affect its performance. We employ grid search to identify the optimal set of hyperparameters for the best performance on the validation set. By fitting the model with the grid of hyperparameters, we evaluate each set of hyperparameters performance on the validation set and choose the best one.

- Logistic Regression

Logistic Regression estimates the probability of the dependent variable based on independent features. Regularization is used to avoid overfitting by adding a penalty term to the cost function. L1 regularization (Lasso) helps control model complexity and prevent overfitting to training data.

- KNN

KNN compares new input data to existing data, selecting the k-nearest neighbors based on Euclidean distance. The output is the majority class of the neighbors. To optimize the model, we normalize input features and use grid search to find the optimal value of k.

- SVM

SVM is effective in handling high-dimensional data and can produce accurate and robust results even with limited training data. When dealing with non-linearly separable data, SVM employs kernel functions. In this case RBF(Radial Basis Function) kernel and a pipe model to combine a non-linear SVM with StandardScaler is used.

2.3 *Fusion of modalities*

Neural network is one of the most famous machine learning algorithm these years. The ability to train in high accuracy is playing a significant role in large amount of areas. Neural networks excel in deep learning with single-domain datasets, but current research focuses on multimodal systems that use multiple sensors to extract and combine essential information for improved performance. These systems are expected to outperform single-modal approaches.

This report covers 3 methods to improve performance using the Emopain Dataset, which includes data from 22 healthy individuals and 18 with lower-back pain. The dataset includes IMU and sEMG data for 5 motions, with 22 sets of XYZ coordinates and surface electromyographic data for body joints.

2.3.1 Early fusion

Early fusion, or data-level fusion, is a traditional method for integrating data before experiments by merging data in a lower-dimensional common space. We must assume conditional dependence between multiple sources, although this may not always be true due to highly correlated features. To visualize the output, correlation matrices, and P and F values will be compared.

In figure.21, features are extracted pre-fusion, then all 70 features are combined and trained. The trained model serves as a binary classifier, categorizing new observations into one of two classes. Four methods include random forest (RF), support vector machines (SVM), K-Nearest Neighbors (KNN), and convolutional neural network (CNN) for classification.

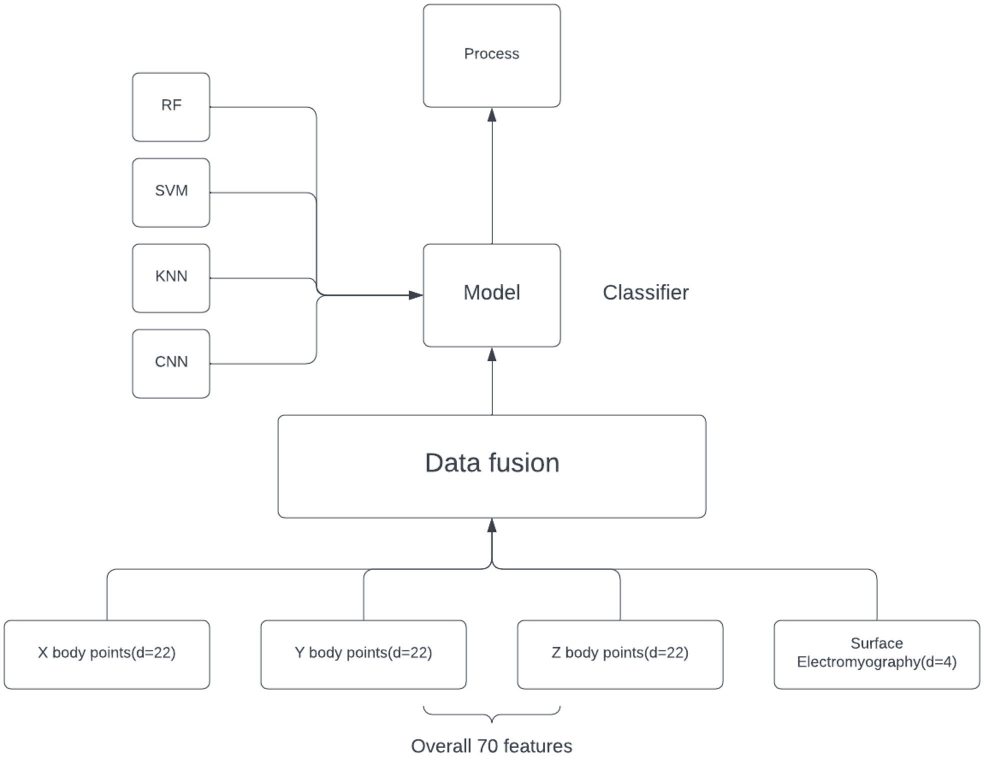


Figure 21: Flowchart:Early Fusion

2.3.2 Mid fusion or Intermediate fusion

The architecture of mid fusion is built on several methods of machine learning and it is the most flexible method in all three fusion methods. The Figure 22 shows one choice of all the options which is suitable to extract features. It can hold different structure and change easily. The machine learning algorithms have improved the performance significantly.

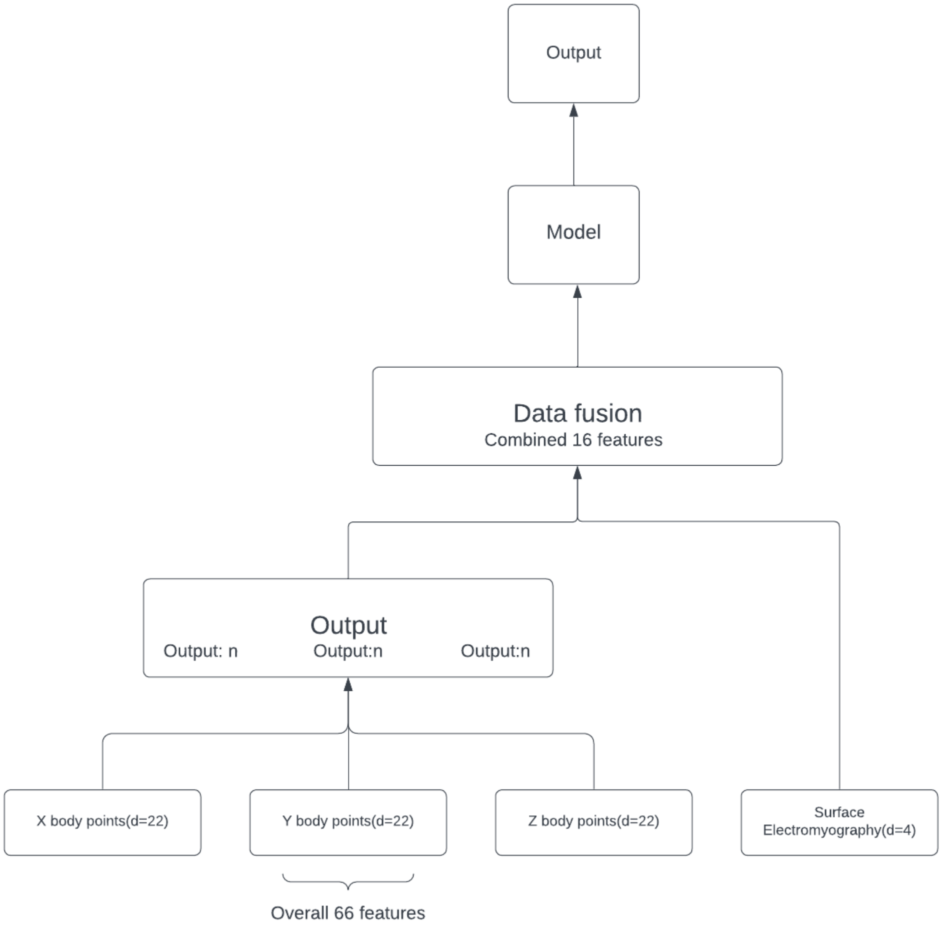


Figure 22: Flowchart:Mid Fusion

In Figure 22, the XYZ points of body joints have 66 original features. The XYZ part matches the dimensions of the surface electromyography data features, which have 4. After the first extraction, 12 features proceed to the next stage. Surface electromyography then contributes 4 features to the data fusion step, followed by processing with the designated model.

Dimension reduction can be achieved using statistical solutions like Principal Component Analysis (PCA), which simplifies large variable sets while retaining crucial information analyzed by feature importance.

2.3.3 Late fusion

In [16], Late fusion employs independent data points at the decision-making stage. Inspired by ensemble classifiers, this method is simpler than early and mid fusion techniques, providing improved performance due to uncorrelated errors from multiple model processing. Late fusion is advantageous when data sources have significant differences in sampling rate and dimensionality. Various optimal methods can combine all features.

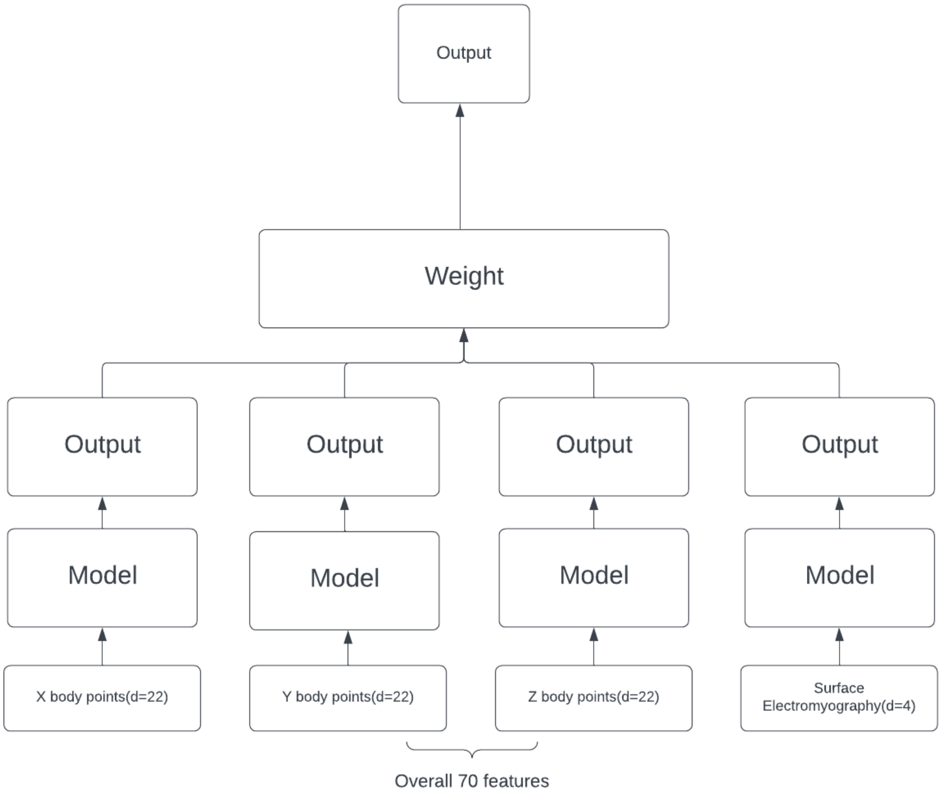


Figure 23: Flowchart:Late Fusion

In Figure 23, each data feature is sent to the model. The PCA algorithm is then used to reduce data dimensionality, retaining 95 percent of the information. Data is processed separately using layered networks. Finally, all retained features are combined and the algorithm is conducted again.

2.4 *Evaluation*

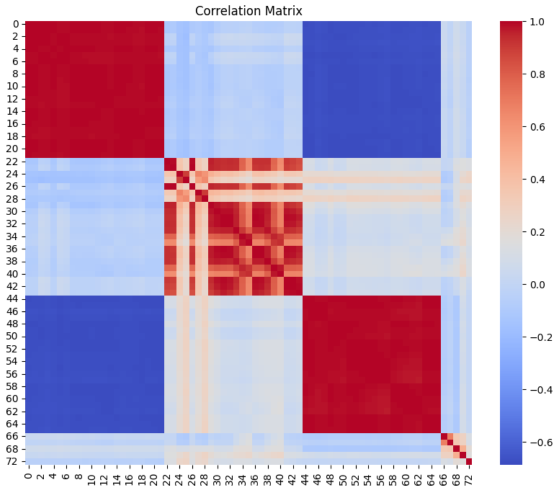


Figure 24: ANOVA Correlation matrix

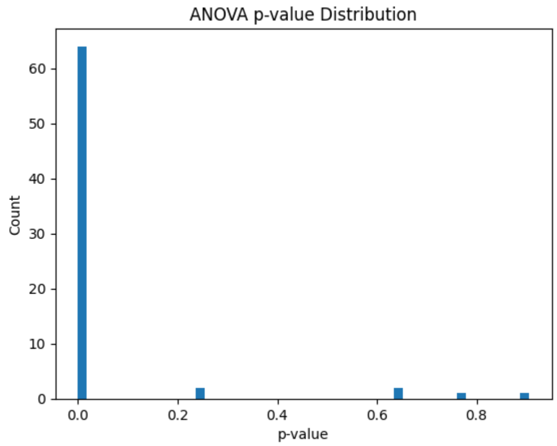


Figure 25: ANOVA P-value histogram

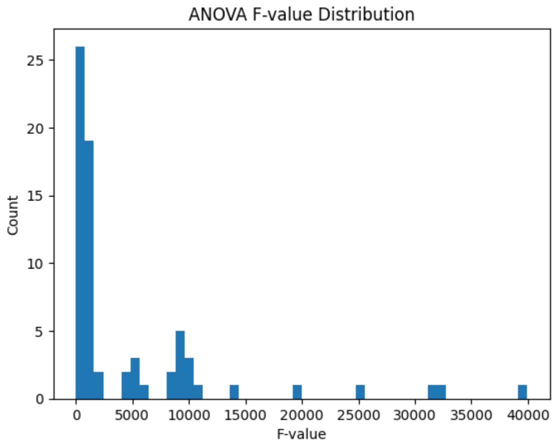


Figure 26: ANOVA F-value histogram

Figure 24 shows the correlation matrix; Figures 25 and 26 depict the histograms for p-value and f-value, respectively. During the ANOVA analysis for data preprocessing, it is found six features had p-values greater than 0.05, indicating a lack of significant discriminative power for the target variable. As a result, two different feature vector configurations were attempted but did not show significant improvement in training results. Further analysis suggested the 46th, 47th, 49th, 50th, 56th and 57th features, located at the Y and Z modalities, is highly correlated with other Y and Z features. Despite its lack of discriminative power, we decided to retain this feature in our model based on the correlation matrix, which showed its strong correlation with other significant features.

The model results shown below were tested in the test dataset, which ensures the fitting capability of generalization of models, although the results were almost the same as those in the training dataset. To evaluate the model results, confusion matrices are used to compute the four metrics including accuracy, recall, precision and F1 score.

The confusion matrix in early fusion of CNN, random forest, logistic regression, KNN and SVM models are shown respectively in Figure 27, Figure 28, Figure 29, Figure 30, and Figure 31

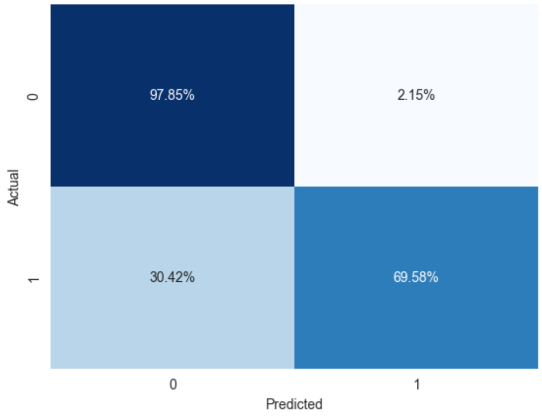


Figure 27: CNN Confusion Matrix in Early Fusion

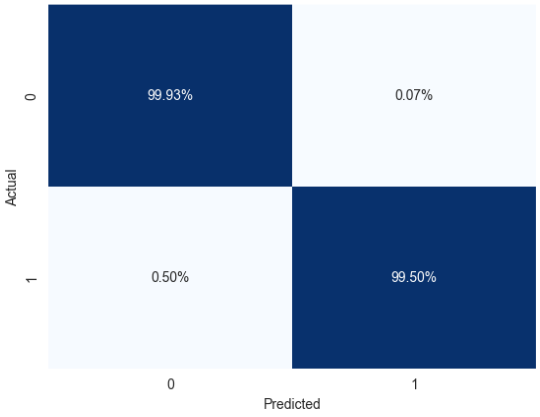


Figure 28: Random Forest Confusion Matrix in Early Fusion

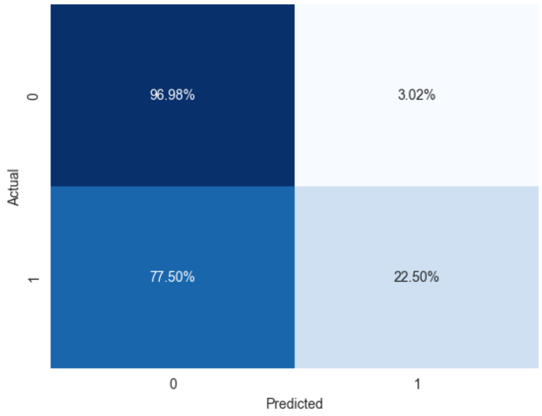


Figure 29: Logistic Regression Confusion Matrix in Early Fusion

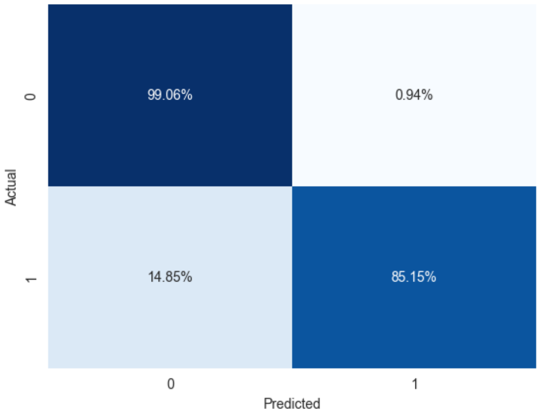


Figure 30: KNN Confusion Matrix in Early Fusion

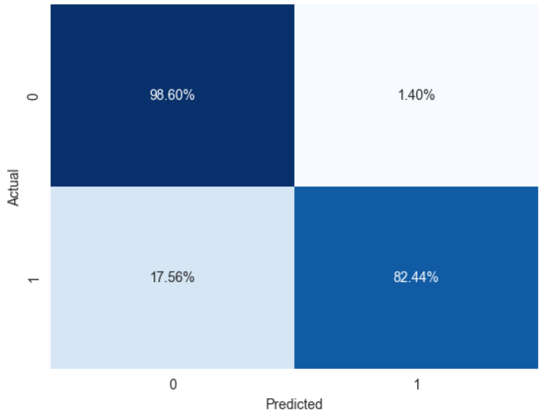


Figure 31: SVM Confusion Matrix in Early Fusion

Model	**Accuracy**	**Recall**	**Precision**	**F1 score**
CNN	93.6%	70.0%	85.1%	76.6%
Random Forest	99.9%	99.5%	99.6%	99.6%
Logistic Regression	85.8%	22.5%	56.8%	32.2%
KNN	97.0%	85.1%	94.1%	89.4%
SVM	96.2%	82.4%	91.2%	86.6%

Table 1: Comparison of Models in Early Fusion

The evaluation of each model within the early fusion is presented in Table 1, illustrating that the random forest model achieves the highest performance across all four metrics, followed by KNN. SVM and CNN exhibit moderate results, while logistic regression yields the least favorable outcomes, particularly in terms of recall, precision, and F1 score. It can be deduced that random forest and KNN possess advantages in affective recognition tasks over CNN, as the latter is designed for high-dimensional data like images, while random forest and KNN can handle low-dimensional data more efficiently, which means for tabular data or time series data, random forest and KNN may produce better results. The reason for logistic regression the worst performance may be it is more prone to outliers than random forest and KNN, and therefore can result in the overfitting and misclassification problem which significantly influences its performance. And this can be seen in the left bottom side of the result in the logistic regression confusion matrix. Moreover, it can be caused by the difficulty of converging of the trained model, which may not give a reasonable result.

Furthermore, the way of modalities fusion may probably also affect the results. Figure 32 demonstrated the mid fusion confusion matrix, where the random forest model is chosen as it performs best. Figure 33 shows the confusion matrix of late fusion. Apart from the confusion matrices, the comparison is illustrated in Table 2

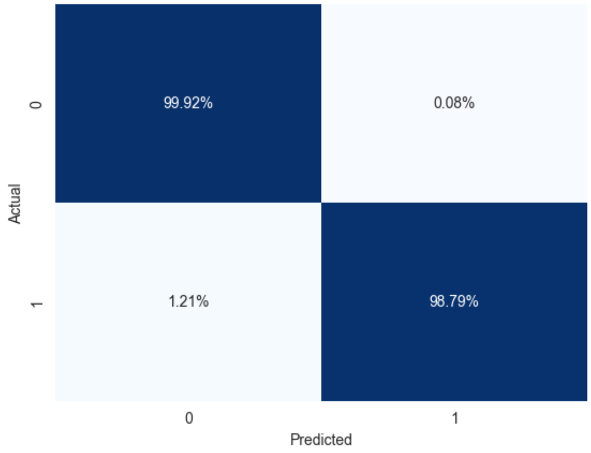


Figure 32: Confusion Matrix in Mid Fusion

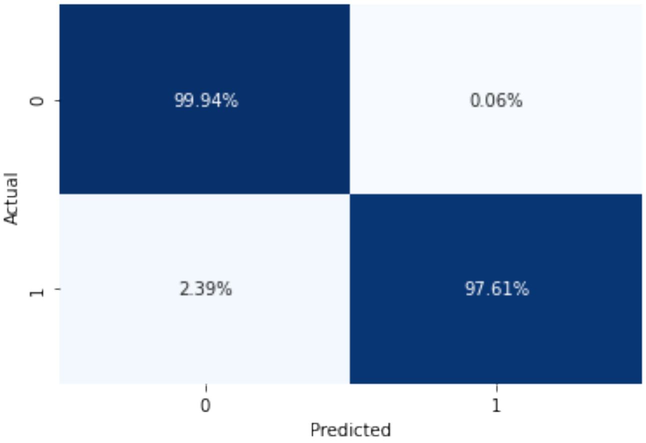


Figure 33: Confusion Matrix in Late Fusion

Model	**Accuracy**	**Recall**	**Precision**	**F1 score**
Early Fusion	99.9%	99.5%	99.6%	99.6%
Mid Fusion	99.7%	98.8%	99.5%	99.2%
Late Fusion	99.6%	97.6%	99.7%	98.6%

Table 2: Comparison of Models in Early Fusion

The comparative analysis of the three fusion architectures reveals that the overall performance of all three architectures are acceptable with F1 scores nearing 99% and accuracies exceeding 99.5%. Despite these overall results, early fusion outperforms both mid fusion and late fusion. The presence of relatively strong correlations among input features may contribute to this outcome, as evidenced by the positive and negative associations depicted in the red and blue blocks respectively in the Figure 24. Nevertheless, it is essential to acknowledge the potential benefits and adaptability of late fusion in future applications where late fusion may exhibit increased robustness when handling independent, complex and asynchronous modalities.

In conclusion, we preprocessed the data by analyzing and selecting the most relevant features for our model using ANOVA F-values, p-values, and a correlation matrix. Our analysis revealed that 6 features lacked significance on its own but was highly correlated with other significant features, leading us to retain it in our model. After the data preprocssing, different models including CNN, random forest, logistic regression and KNN models are compared in the early fusion, and the different fusion architectures are compared also. The results demonstrated that the random forest model excelled among other machine learning models, and the early fusion outperformed other fusion architectures.

References

[1] J. Jitesh, W. Katherine, and C. Youngjun, "Physiokit: Open-source, accessible physiological computing toolkit." https://github.com/PhysiologicAILab/PhysioKit.

[2] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 2018, pp. 59–66.

[3] "The best calming images of all time you can't miss looking at," https://www.calmsage.com/calming-images.

[4] L. Israel, P. Paukner, L. Schiestel, K. Diepold, and F. Schönbrodt, "Open library for affective videos (openlav)," 2021.

[5] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.

[6] B. Nadia and H. Adam, "Information sheet for student participants in affective computing and human-computer interaction (achri) module, comp0053," https://moodle.ucl.ac.uk/mod/url/view.php?id=4076736.

[7] "Ucl data protection for undergraduate and masters level students," https://www.ucl.ac.uk/data-protection/ucl-data-protection-undergraduate-masters-level-students.

[8] B. Tadas, "Emotion recognition using openface," https://github.com/TadasBaltrusaitis/OpenFace/issues/424.

[9] J. A. Russell, "Affective space is bipolar." *Journal of personality and social psychology*, vol. 37, no. 3, p. 345, 1979.

[10] P. Kuppens, F. Tuerlinckx, J. A. Russell, and L. F. Barrett, "The relation between valence and arousal in subjective experience." *Psychological bulletin*, vol. 139, no. 4, p. 917, 2013.

[11] H. Ferdinando, T. Seppänen, and E. Alasaarela, "Comparing features from ecg pattern and hrv analysis for emotion recognition system," in *2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, 2016, pp. 1–6.

[12] Q. Zhang, X. Chen, Q. Zhan, T. Yang, and S. Xia, "Respiration-based emotion recognition with deep learning," *Computers in Industry*, vol. 92, pp. 84–90, 2017.

[13] P. Ekman and W. V. Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978.

[14] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, "Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders," *Journal of Neuroscience Methods*, vol. 200, no. 2, pp. 237–256, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016502701100358X

[15] M. S. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, M. Valstar, H. Meng, A. Kemp, M. Shafizadeh *et al.*, "The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal emopain dataset," *IEEE transactions on affective computing*, vol. 7, no. 4, pp. 435–451, 2015.

[16] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms: Second Edition*, 01 2014, vol. 47.