NÖHÜ Müh. Bilim. Derg. / NOHU J. Eng. Sci., Erken Görünüm / InPress

Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi
Niğde Ömer Halisdemir University Journal of Engineering Sciences

Araştırma makalesi / Research article

www.dergipark.org.tr/tr/pub/ngumuh / www.dergipark.org.tr/en/pub/ngumuh

The effects of sensor and feature level fusion methods in multimodal emotion analysis

Çok modlu duygu analizinde sensör ve özellik seviyesi füzyon yöntemlerinin etkileri

Bahar Hatipoğlu Yılmaz[1,*], Cemal Köse[2]

[1] *Karadeniz Technical University, Department of Computer Engineering, 61000, Trabzon Türkiye*

Abstract

Fusion-based studies on multimodal emotion recognition (MER) are very popular nowadays. In this study, EEG signals and facial images are fused using Sensor Level Fusion (SLF) and Feature Level Fusion (FLF) methods for multimodal emotion recognition. The general procedure of the study is as follows. First, the EEG signals are converted into angle amplitude graph (AAG) images. Second, the most unique ones are automatically identified from all face images obtained from video recordings. Then, these modalities are fused separately using SLF and FLF methods. The fusion approaches were used to combine the obtained data and perform classification on the integrated data. The experiments were performed on the publicly available DEAP dataset. The highest accuracy was 82.14% with 5.26 standard deviations for SLF and 87.62% with 6.74 standard deviations for FLF. These results show that this study makes an important contribution to the field of emotion recognition by providing an effective method.

Keywords: Multimodal emotion recognition, Sensor level fusion, Feature level fusion, DEAP dataset

Öz

Füzyon tabanlı çok modlu duygu tanıma (MER) çalışmaları günümüzde oldukça popülerdir. Bu çalışmada, çok modlu duygu tanıma için EEG sinyalleri ve yüz görüntüleri sensör seviyesinde füzyon (SLF) ve öznitelik seviyesinde füzyon (FLF) yöntemleri ile birleştirilmiştir. Çalışmanın genel akışı şu şekildedir. İlk olarak EEG sinyalleri açı genlik grafiği (AAG) görüntülerine dönüştürülmektedir. İkinci olarak, video kayıtlarından elde edilen tüm yüz görüntülerinden en benzersiz olanlar otomatik olarak belirlenmektedir. Daha sonra, bu modaliteler SLF ve FLF yöntemleri kullanılarak ayrı ayrı birleştirilmektedir. Elde edilen verileri birleştirmek ve bütünleşik veriler üzerinde sınıflandırma yapmak için füzyon yaklaşımlar kullanılmıştır. Deneyler halka açık DEAP veri kümesi üzerinde gerçekleştirilmiştir. En yüksek doğruluk SLF için 5.26 standart sapma ile %82.14 ve FLF için 6.74 standart sapma ile %87.62 olarak elde edilmiştir. Bu sonuçlar, bu çalışmanın etkili bir yöntem sunarak duygu tanıma alanına önemli bir katkı sağladığını göstermektedir.

Anahtar kelimeler: Çok modlu duygu tanıma, Sensör seviyesinde füzyon, Öznitelik seviyesinde füzyon, DEAP veri seti

1 Introduction

Emotion can be defined as an important characteristic of humans associated with feelings, behaviors, etc. [1, 2]. Emotion recognition has become a significant research area for communication between people, reasoning, effective decision-making, etc. Many researchers have studied emotion recognition using various modalities such as speech, body gestures, facial images, electroencephalogram (EEG), etc. Studies that focus on only one modality, such as speech or EEG, are referred to as unimodal studies. With advancements in technology and research, researchers have increasingly turned their attention to multimodal studies that utilize two or more different modalities simultaneously. These multimodal studies provide researchers with effective and efficient fusion methods, combining different modalities. In this regard, multimodal emotion recognition (MER) systems play an important role in advancing emotion recognition research.

In the literature, numerous multimodal studies that integrate different modalities have been conducted for MER. Some remarkable studies utilizing different modalities for MER are outlined as follows. Verma and Tiwary [3] proposed an investigation into emotion representation models and the recognition of emotions from measured physiological signals. They employed the DEAP dataset and applied discrete wavelet transform for analyzing emotional signals. Additionally, they utilized four different classifiers Support Vector Machine (SVM), Multilayer Perceptron (MLP), *k*-Nearest Neighbor (*k*-NN), and Meta-multiclass (MMC). Their average results showed 81.45% accuracy with SVM, 74.37% with MLP, 57.74% with *k*-NN, and 75.94% with MMC, respectively. Luo et al. [4] proposed a MER system based on oil painting stimuli. They presented an emotional dataset with three emotional states named negative, neutral, and positive. In this dataset, they recorded EEG and eye-tracking data from 20 subjects with 114 oil painting stimuli. Additionally, they used accuracy and F1-

* Sorumlu yazar / Corresponding author, e-posta / e-mail: baharhatipoglu@ktu.edu.tr (B. Hatipoğlu Yılmaz)
Geliş / Recieved: 16.04.2024 Kabul / Accepted: 25.06.2024 Yayımlanma / Published: xx.xx.20xx
doi: 10.28948/ngmuh.1469289

score metrics to evaluate the performance of the study. They obtained 89.12±4.26 with SVM and 94.72±1.47 with their proposed method. Njoku et al. [5] proposed to compare the performance of deep learning models for MER. They applied early fusion, hybrid fusion, and multi-task learning. In the early fusion, hybrid fusion, and multi-task learning approaches, they achieved 78.41%, 68.33%, and 78.75%, respectively. Pan et al. [6] proposed a deep learning-based MER system. They integrated facial expressions, speech, and EEG modalities to improve performance, employing a Decision Level Fusion (DLF) technique. They used CK+, EMO-DB, and MAHNOB-HCI datasets to perform their proposed method. For the CK+ dataset, they achieved 98.27%, and for the EMO-DB dataset, they achieved 94.36%.

In this study, we utilized EEG signals and facial images together for MER, and similar studies that employed these modalities are listed as follows. Li et al. [7] utilized EEG signals and facial images for emotion recognition. They applied a DLF fusion method and tested their proposed approach on DEAP and MAHNOB-HCI datasets. Zhao and Chen [8] also employed EEG signals and facial expressions. They extracted facial expression features using the bilinear convolution network (BCN) and fused these features using a three-layer bidirectional LSTM. Additionally, they tested their method on DEAP and MAHNOB-HCI datasets. Similarly, Zhao and Chen employed a transfer learning model for facial expression detection and used an SVM classifier to detect EEG targets labeled as Arousal (ARO) or Valence (VAL). Finally, they adopted two DLF methods corresponding to the enumerate weight rule or an adaptive boosting technique to combine the modalities. They also evaluated their study using DEAP and MAHNOB-HCI datasets. Huang et al. [9] proposed two distinct DLF approaches and tested them on the DEAP and MAHNOB-HCI datasets. Yin et al. [10] used a Multiple-Fusion Layer-based Ensemble Classifier of Stacked Auto-Encoder (MESAE) to evaluate their MER system on the DEAP dataset.

In our work, we proposed SLF and FLF methods. In the first stage, EEG signals were transformed into AAG images. Subsequently, peak frames were automatically selected from all facial images using the maximum dissimilarity-based method (MAX-DIST). For SLF, these image modalities were merged before the feature extraction technique, and all subsequent stages, including feature extraction and classification, were performed on the merged images. For FLF, features were extracted separately from EEG images and facial images, followed by a fusion of the extracted features. After both fusion approaches, *k*-NN and SVM algorithms were applied for classifications. Furthermore, each modality was also classified separately to show the specific contribution of our study.

The rest of this paper is organized as follows: Section (2.1.1) presents a detailed overview of the DEAP dataset. Section (2.1.2) describes the angle-amplitude transformation method. Subsequently, Section (2.1.3) provides details on the selection and preparation of facial images. Sections (2.1.4) and (2.1.5) explain the methods used for feature extraction and classification. In Section (3), SLF and FLF methods are outlined. Finally, the experimental results and conclusions are presented in the last two sections.

2 Material and Methods

2.1 Dataset description

2.1.1 DEAP dataset

We evaluated the performance of a MER system using the well-known DEAP dataset [11]. The DEAP dataset was collected by a group of researchers at Queen Mary University. The dataset includes EEG and physiological signals recorded from 32 healthy subjects, with face videos recorded from the first 22 subjects. For all experiments in this study, we exclusively utilized data from the first 22 subjects, excluding s03, s05, s11, s14, and s20.

In the DEAP dataset, 32-channel EEG data were acquired following the international 10/20 electrode placement with a sampling frequency of 512~Hz. The data were preprocessed to remove outliers, and the recorded signals were down-sampled to 128~ Hz. Additionally, a bandpass filter with cutoff frequencies of 4.0-45.0~ Hz was applied.

In this paper, we focused on utilizing the EEG and face image modalities from the dataset. Our experiments were specifically conducted on the ARO and VAL dimensions. To categorize the trials, we set a threshold of 5, dividing them into two classes based on the rated levels of ARO and VAL. Specifically, for both ARO and VAL dimensions, ratings higher than 5 were labeled as positive classes, while ratings lower than 5 indicated negative classes.

2.1.2 Angle amplitude transformation

We implemented the angle-amplitude transformation (AAT) as previously suggested in our works [12, 13, 14]. But, in this work, the angle and amplitude values were computed using only the signals' maximum (max.) points, as illustrated in Figure 1. Specifically, all calculations were performed from max points to min points, and no calculations were conducted in the opposite direction (from min points to max points). Figure 1 provides a representation of a sample signal.

- Detect all of the local max and min points on a signal.
- Calculate the Euclidean distances between the right and left min points of each max point.
- Calculate the angle values (according to the tangent formula) between the left and right lines of a max point.
- Determine the amplitude value of the current max point belonging to the magnitude of left and right lines.
- Locate the angle and corresponding amplitude values to the quadrants of the graph.

Further mathematical details of the algorithm can be found in [12, 13]. Sample signal images obtained from the EEG signals are presented in Figure 2. In this figure, the upper side displays ARO class images, while the lower side shows VAL class images.

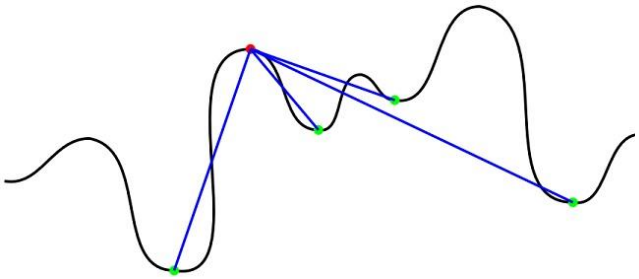

Figure 1. A representation of max. and min. points on a signal

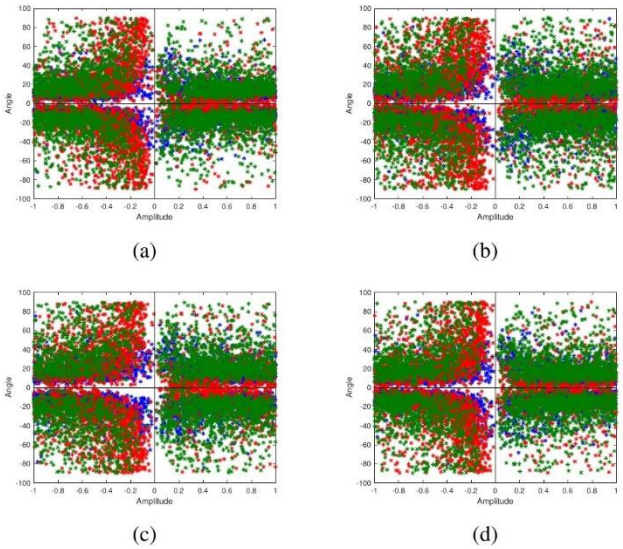

Figure 2. A representation of max. and min. points on a signal

2.1.3 *Peak frame selection from face images*

We automatically identified unique facial images, referred to as peak frames, from a face image sequence using the MAX-DIST method, as proposed in [15]. Notably, we incorporated Local Binary Patterns (LBP) for feature extraction, differing from their original algorithm. The algorithm is described as follows. Initially, we labeled a video sequence as V=$\{V_1, V_2, \dots, V_N\}$, where N is the number of frames. A sample video sequence belonging to a female subject, along with its peak frame, is illustrated in Figure 3. Subsequently, we created an $N \times N$-sized dissimilarity matrix $M(i,j)$, where $i,j\ in\ \{1,2,\dots,N\}$. The dissimilarity matrix was computed using the chi-square distance between LBP features extracted from all facial images. We calculated the average dissimilarity score d_j between the remaining $N-1$ frames and the j-th frame in the M matrices. Finally, we sorted the averages in descending order and identified the highest mean K values as peak frames. In this study, we chose $K = 1$, indicating that only the first image was used as the peak frame.

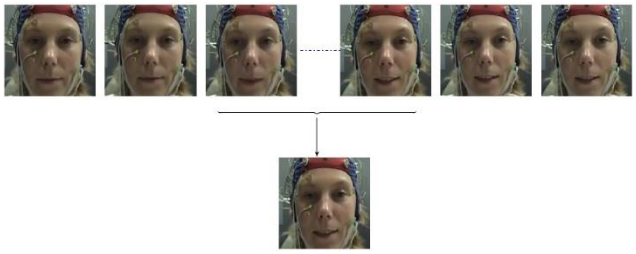

Figure 3. A female image sequence and its peak frame belongs DEAP dataset

2.1.4 *Feature extraction*

2.1.4.1 *Local binary pattern*

In this paper, we utilized the Local Binary Pattern (LBP) and Scale Invariant Feature Transform (SIFT) algorithms for feature extraction. LBP is an algorithm designed for identifying image textures [16]. The basic concept behind the LBP operator is that $2D$ surface textures can be characterized by two descriptive measures, namely local spatial patterns and gray-scale contrast [17]. The advantages of the LBP algorithm lie in its sensitivity to ambient lighting changes and its low computational cost. The algorithm operates by comparing the gray-level values of the eight pixels in the 3×3 neighborhood around the central pixel [18]. Consequently, the LBP operator can be conceptualized as an ordered set of pairwise comparisons between gray levels of the central pixel.

2.1.5 *Classification*

2.1.5.1 *K-nearest neighbor*

k-Nearest Neighbor (*k*-NN) is one of the well-known methods in machine learning utilizing supervised learning principles, which was initially introduced by Evelyn Fix and Joseph Hodges in 1951 [19]. The algorithm can be described as follows.

- *k* training set samples with known labels (neighbor points) are selected.
- The distance is calculated between *k* neighbor points and the test sample with Euclidean distance metric $d(x,y) = \sqrt{(\sum_{i=1}^{k}(x_i - y_i)^2)}$.
- According to the distance value, *k* nearest neighbor points are chosen.
- The number of training samples in each category of these *k* neighbors is determined.
- The category of the test sample is determined by looking at the majority of the categories.

2.1.5.2 *Support vector machine*

Support Vector Machine (SVM) was originally developed by Cortes and Vapnik [20]. Particularly in everyday problems, linearly separable data are quite rare. Hence, SVM, known for its high generalization ability, has been widely employed in the literature [20, 21]. SVM fundamentally aims to discover the optimal separating hyperplane for linearly separable data and conducts classification based on this hyperplane. In other words, the SVM mechanism is designed to identify the most suitable

classifying hyperplanes that meet the classification requirements [22]. In this study, we used the radial basis kernel function as the kernel function.

3 Fusion methods

3.1 Sensor level fusion

In this study, we applied a novel Sensor Level Fusion (SLF) using transformed $2D$ signal images and facial images. This involved merging signal images derived from recorded EEG signals with facial images, creating a single image where these two types of images were positioned side by side. The resulting combined image was then utilized to observe the contribution to the classification process. An illustrative representation of the merged image is presented in Figure 4. As depicted in Figure 4, the process begins by automatically detecting peak frame face images using an algorithm. Subsequently, the signal is transformed into the AAG image version. Finally, these two images are merged to form the final representation.

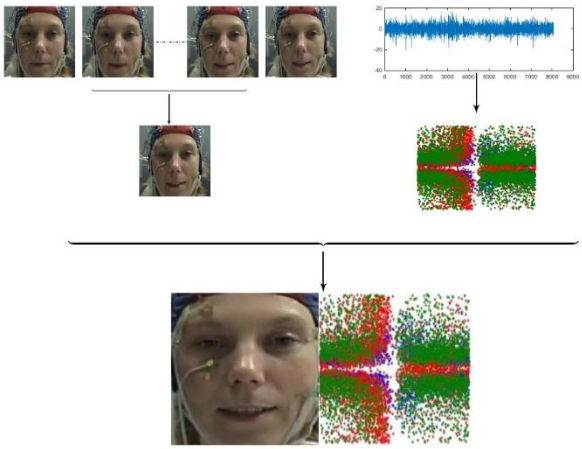

Figure 4. Merging a peak frame and signal image representation for SLF

3.2 Feature level fusion

In this study, we also applied a Feature Level Fusion (FLF) method that incorporated signal and face images. The fusion process took place after the feature extraction section, and the FLF method can be summarized as follows. To extract features from signals, the signals were transformed into $2D$ images, and the SIFT algorithm was applied to these images. Consequently, each image generated a varying number of descriptors ($Nx128$, where N is the number of descriptor vectors). Subsequently, the N descriptors were averaged to obtain a unique descriptor of size ($1x128$). The averaging of SIFT vectors was inspired by [23-26].

For extracting features from faces, frontal face videos were divided into individual frames, and their peak frames were automatically detected. Then, LBP was applied to these peak frames, resulting in a $1x256$-sized feature vector. After obtaining the feature vectors from both types of images, these vectors were fused. The final fusion vector, of size $1x384$, was calculated by concatenating them sequentially. An illustrative representation of the FLF algorithm is provided in Figure 5.

4 Experimental results

We aimed to propose novel sensor-level and feature-level fusion approaches. Additionally, we introduced a new AAT that utilizes only angle-amplitude values calculated from max points to min points. The procedure of the paper can be concisely summarized as follows. Firstly, we transformed signals into $2D$ images and automatically identified peak frames from frontal face videos. Then, for SLF, we merged the signal images and peak frames before feature extraction. Features were extracted from these merged images using LBP and finally, the features were classified using *k*-NN and SVM algorithms. For FLF, we extracted LBP features from peak frames and SIFT features from signal images. Consequently, these modalities were fused after feature extraction.

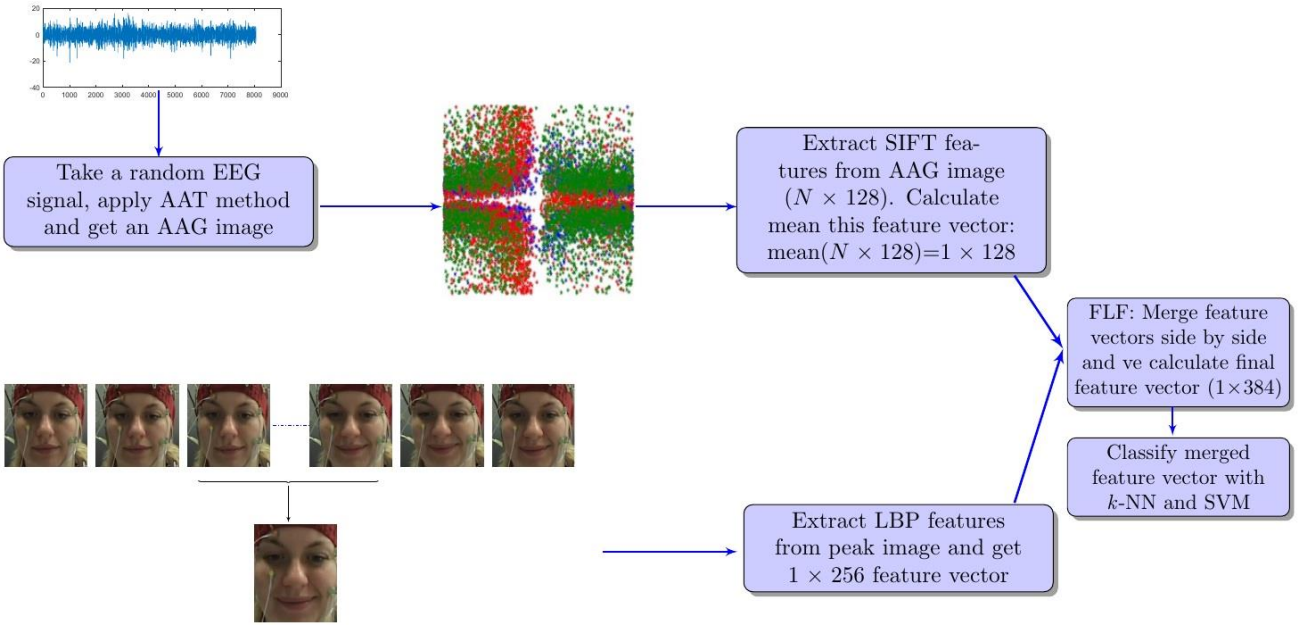

Figure 5. A representation of feature level fusion for a female subject in DEAP dataset

Finally, we classified the fused features with *k*-NN and SVM algorithms. We evaluated these new approaches on publicly available DEAP benchmark datasets. The performances were evaluated using the accuracy measure as given in Equation (1). In this notation, TP is the number of positive samples correctly identified, TN is the number of negative samples correctly identified, FP is the number of negative cases incorrectly identified, and FN is the number of positive cases incorrectly identified.

$$CA = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

We divided the DEAP dataset into 10 subsets. Accordingly, we used one of them as the test data and the others as the training data. Also, we repeated this procedure ten times and calculated the final accuracy by averaging the results. Because parameter selection plays an important role in classification, we identified the parameters associated with feature extraction and classification exclusively through the utilization of training data. For the feature extraction stage, the radius is an important parameter related to the LBP and SIFT algorithms. The optimal radius (r) values of these algorithms are searched in the range of $1 \leq r \leq 3$. In this work, the SIFT algorithm is used with Harris corner detection. Additionally, t and σ values are important parameters related to Harris. The optimal t and σ values of the algorithm are searched in the range of $1000 \leq t \leq 3000$ and $3 \leq \sigma \leq 6$, respectively. For the classification stage, k is an important parameter related to the *k*-NN algorithm. The optimal k value of the algorithm is searched in the range of $1 < k \leq (training\ set \div 2)$. Similarly, C and ɣ values are important parameters related to the SVM. The optimal C and ɣ values of the algorithm are searched in the range of $2^{-15} \leq C \leq 2^{+15}$ and $1 \leq \gamma \leq 300$ respectively.

Table 1. CA (%) results and standard deviations for ARO and VAL dimensions on DEAP dataset

Method	Face Images	AAG Images	SLF	FLF
k-NN-ARO	65.99 ±10.67	80.67±6.84	82.14±5.26	87.26±4.52
SVM-ARO	67.33 ±9.81	78.87±4.91	81.59±6.13	86.04±7.52
k-NN-VAL	58.39 ±7.22	77.38±5.47	80.69±4.39	85.24±4.27
SVM-VAL	61.38 ±7.58	78.61±3.71	78.87±5.56	87.62±6.74

Table 2. Comparison with Studies Conducted on the DEAP Dataset

Study	ARO	VAL
[3]	81.45	-
[8]	86.8	86.2
[10]	77.19	76.17
[7] (for Enumerator fusion)	58.75 ±12.26	71.00±7.00
[7] (for Adaboost fusion)	59.00 ±10.74	70.25 ±8.25
[9] (for First fusion)	74.23 ±10.34	80.30 ±11.37
[9](for Second fusion)	71.54 ±11.16	80.00 ±12.40
Proposed method (for SLF)	81.59±6.13	80.69±4.39
Proposed method (for FLF)	87.26±4.52	87.62±6.74

The average classification accuracies across all subjects are shown in Table 1. The first column shows method (classifier-dimension pair) names. The second and third columns show the only face and AAG image classification accuracies. The last two columns show the fusion results (the fourth column for SLF and the fifth one for FLF). For ARO dimension, *k*-NN shows 65.99% for face images, 80.67% for AAG images, 82.14% for SLF, and 87.26% for FLF. Similarly, the SVM classifier shows 67.33% for face images, 78.87% for AAG images, 81.59% for SLF, and 86.04% for FLF. For VAL dimension, *k*-NN achieves 58.39% accuracy for face images, 77.38% for AAG images, 80.69% for SLF, and 85.24% for FLF. Similarly, SVM shows 61.38% accuracy for face images, 78.61% for AAG images, 78.87% for SLF, and 87.62% for FLF.

Looking at all the results, AAG images have better classification accuracy than face images. Besides, the FLF approach achieves better accuracy than the SLF approach. Generally, better results are observed for ARO. Additionally, *k*-NN and SVM achieved better accuracy for ARO. For all results, achieved standard deviations are within acceptable ranges. A graphical comparison of all results is also given in Figure 6.

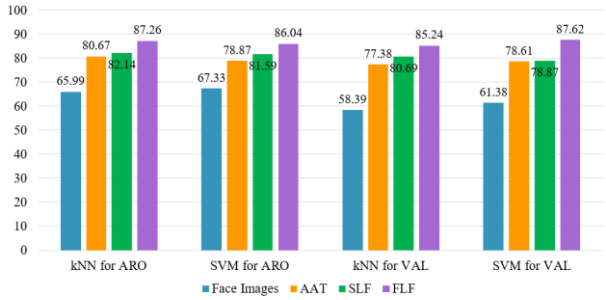

Figure 6. A graphical representation of DEAP dataset results

Additionally, there is a comparison table (Table 2) that includes studies in the literature. The first column of the table presents the related studies, while the second and third columns present the results for ARO and VAL. In order to provide a comprehensive understanding of the performance of the proposed methods, we have compared our results with several existing studies in the literature and listed a summary of the main findings from these studies in the following. [3] achieved an ARO accuracy of 81.45% but VAL was not reported. [7] got moderate accuracy levels, with Enumerator fusion yielding 58.75±12.26 for ARO and 71.00±7.00 for VAL, and Adaboost fusion resulting in 59.00±10.74 for ARO and 70.25±8.25 for VAL. [8] proposed two DLF methods based on the enumerate weight rule or an adaptive boosting technique to combine the face and EEG modalities. They achieved ARO and VAL accuracies of 86.8% and 86.2%, respectively. [9] yielded 74.23±10.34 for ARO and 80.30±11.37 for VAL with the first fusion method, while showed 71.54±11.16 for ARO and 80.00±12.40 for VAL with the second fusion method. Analyzing these studies, it is evident that the highest accuracy values for ARO and VAL

are achieved in [8]. In conclusion, the proposed FLF method demonstrates a significant improvement over existing studies, with an increase of 0.46% for ARO and 1.42% for VAL compared to the highest values reported in the literature. These results highlight the effectiveness of our approach in enhancing emotion recognition accuracy, offering a more robust and reliable solution.

5 Conclusion

We have evaluated the effect of sensor and feature-level fusion methods on MER using the famous DEAP dataset. We have also classified the modalities (signal images and face images) separately to better illustrate the contribution of fusion methods. In summary, (i) we first applied our original signal to image transformation method to the signals, (ii) we detected the peak frames between all facial images, (iii) then utilized our SLF and FLF methods, and (iv) finally we extracted and classified features according to the fusion methods. For a single trial classification, we conducted the experiments as a two-class classification experiment that ratings for ARO and VAL. For SLF, we got the best average classification accuracies with *k*-NN as 82.14% with 5.26 standard deviation for ARO. To the best of our knowledge, this is the first study to use SLF for MER, although the expected results were not obtained. Similarly, for FLF, we got the best average classification accuracies of 87.62% with 6.74 standard deviation for VAL. The achieved classification accuracies underscore the robustness and effectiveness of our proposed methodology, emphasizing its significant performance in MER tasks.

In future work, we will further aim to propose a novel multi-modal dataset to investigate the role of different modalities in emotion recognition process. Besides, we will combine the results on decision-level to improve the reliability and performance of our proposed approach.

Acknowledgment

This research was supported by the Turkish Scientific and Research Council (TUBITAK) through project 121E002 and 119E397.

Conflict of Interest

The authors declare that there is no conflict of interest.

Similarity Rate (iThenticate): %19

References

[1] A. F. M. N. H. Nahin, J. M. Alam, H. Mahmud and K. Hasan, Identifying emotion by keystroke dynamics and text pattern analysis. Behaviour & Information Technology, 33(9), 987–996, 2014. https://doi.org/10.1080/0144929X.2014.907343.

[2] A. Sapra, N. Panwar, and S. Panwar, Emotion recognition from speech. International journal of emerging technology and advanced engineering, 3(2), 341-345, 2013.

[3] G. K. Verma, and U. S. Tiwary, Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals. NeuroImage, 102, 162-172, 2014. https://doi.org/10.1016/j.neuroimage.2013.11.007.

[4] S. Luo, Y. T. Lan, D. Peng, Z. Li, W. L. Zheng, and B. L. Lu, Multimodal Emotion Recognition in Response to Oil Paintings. 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4167-4170, 2022. https://doi.org/10.1109/EMBC48229.2022.9871630.

[5] J. N. Njoku, A. C. Caliwag, W. Lim, S. Kim, H. Hwang, and J. Jung, Deep learning-based data fusion methods for multimodal emotion recognition. The Journal of Korean Institute of Communications and Information Sciences, 47(1), 79-87, 2022. https://doi.org/10.1109/10.7840/kics.2022.47.1.79.

[6] J. Pan, W. Fang, Z. Zhang, B. Chen, Z. Zhang, and S. Wang, Multimodal Emotion Recognition based on Facial Expressions, Speech, and EEG. IEEE Open Journal of Engineering in Medicine and Biology, 2023. https://doi.org/10.1109/10.7840/10.1109/OJEMB.2023.3240280.

[7] R. Li, Y. Liang, X. Liu, B. Wang, W. Huang, Z. Cai, and J. Pan, MindLink-eumpy: an open-source python toolbox for multimodal emotion recognition. Frontiers in Human Neuroscience, 15, 621493, 2021. https://doi.org/10.3389/fnhum.2021.621493.

[8] Y. Zhao and D. Chen, Expression eeg multimodal emotion recognition method based on the bidirectional lstm and attention mechanism. Computational and Mathematical Methods in Medicine, 1-12, 2021. https://doi.org/10.1155/2021/9967592.

[9] Y. Huang, J. Yang, S. Liu, J. Pan, Combining Facial Expressions and Electroencephalography to Enhance Emotion Recognition. Future Internet, 11(5):105, 2019. https://doi.org/10.3390/fi11050105.

[10] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. Computer methods and programs in biomedicine, 140, 93-110, 2017. https://doi.org/10.1016/j.cmpb.2016.12.005.

[11] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, and I. Patras, Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing, 3(1), 18-31, 2011. https://doi.org/10.1109/T-AFFC.2011.15.

[12] B. Hatipoglu, C. M. Yilmaz and C. Kose, A signal-to-image transformation approach for EEG and MEG signal classification. Signal Image and Video Processing, 13, 483–490, 2019. https://doi.org/10.1007/s11760-018-1373-y.

[13] B. Hatipoglu Yilmaz, C. M. Yilmaz and C. Kose, Diversity in a signal-to-image transformation approach for EEG-based motor imagery task classification. Medical Biological Engineering Computing, 58, 443–459, 2020. https://doi.org/10.1007/s11517-019-02075-x.

[14] B. Hatipoglu Yilmaz and C. Kose, A novel signal to image transformation and feature level fusion for multimodal emotion recognition. Biomedical

Engineering / Biomedizinische Technik, 66 (4), 353-362, 2021. https://doi.org/10.1515/bmt-2020-0229.

[15] S. Zhalehpour, Z. Akhtar, and C. Eroglu Erdem, Multimodal emotion recognition based on peak frame selection from video. Signal, Image and Video Processing, 10, 827-834, 2016. https://doi.org/10.1109/INISTA.2014.6873606.

[16] W. Yu, L. Gan, S. Yang, Y. Ding, P. Jiang, J. Wang and S. Li, An improved LBP algorithm for texture and face classification. Signal Image and Video Processing, 8, 155–161, 2014. https://doi.org/10.1007/s11760-014-0652-5.

[17] Matti Pietikäinen, Local Binary Patterns. http://www.scholarpedia.org/article/Local_Binary_Patterns, Accessed 13 March 2024.

[18] C. Turan, and K. M. Lam, Histogram-based local descriptors for facial expression recognition (FER): A comprehensive study. Journal of visual communication and image representation, 55, 331-341, 2018. https://doi.org/10.1016/j.jvcir.2018.05.024.

[19] Behzad Javaheri, KNN with Examples in Python, https://domino.ai/blog/knn-with-examples-in-python, Accessed 14 March 2024.

[20] T. Fletcher, Support vector machines explained. Tutorial paper, 1-19, 2009.

[21] C. J. Burges, A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery, 2(2), 121-167, 1998.

[22] A. Yuexuan, D. Shifei, S. Songhui and L. Jingcan, Discrete space reinforcement learning algorithm based on support vector machine classification. Pattern Recognition Letters, 111, 30-35, 2018. https://doi.org/10.1016/j.patrec.2018.04.012.

[23] T. Q. Anh, P. Bao, T. T. Khanh, B. N. D. Thao, T. A. Tuan and N. T. Nhut, Video retrieval using histogram and sift combined with graph-based image segmentation. Journal of Computer Science, 8(6), 853, 2012. https://doi.org/10.3844/jcssp.2012.853.858.

[24] N. D. Anh, P.T. Bao, B. N. Nam, N.H. Hoang, A New CBIR System Using SIFT Combined with Neural Network and Graph-Based Segmentation. In: Nguyen, N.T., Le, M.T., Świątek, J. (eds) Intelligent Information and Database Systems. Lecture Notes in Computer Science, 5990. Springer, Berlin, Heidelberg, 2010. https://doi.org/10.1007/978-3-642-12145-6_30.

[25] J. Xu, K. Lu, X. Shi, S. Qin, H. Wang and J. Ma, A DenseUnet generative adversarial network for near-infrared face image colorization. Signal Processing, 183, 108007, 2021. https://doi.org/10.1016/j.sigpro.2021.108007.

[26] H. Lacheheb, and S. Aouat, SIMIR: new mean SIFT color multi-clustering image retrieval. Multimedia Tools and Applications, 76, 6333-6354, 2017. https://doi.org/10.1007/s11042-015-3167-3.