Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 17 No. 16 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijim.v17i16.42735

PAPER

Heterogeneous Convolutional Neural Networks for Emotion Recognition Combined with Multimodal Factorised Bilinear Pooling and Mobile Application Recommendation

D. Saisanthiya(✉), P. Supraja

Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India

saisantd@srmist.edu.in

ABSTRACT

The field of emotion recognition has garnered considerable interest due to its diverse applications in mental health, personalised advertising and enhancing user experiences. This research paper introduces a unique and innovative method for emotion recognition by integrating heterogeneous convolutional neural networks (CNNs) with multimodal factorised bilinear pooling. Furthermore, the paper also incorporates the integration of mobile application recommendations as part of the overall approach. The proposed method leverages the power of CNNs to extract high-level features from different modalities, including facial expressions, speech signals and physiological signals. By using heterogeneous CNNs, each modality is processed independently to capture modality-specific emotional cues effectively. To fuse the extracted features, multimodal factorised bilinear pooling is employed, which captures the complex interactions between different modalities while reducing the computational complexity. This pooling technique efficiently combines the modality-specific features, resulting in a compact and discriminative representation of the emotional state. In addition to emotion recognition, this paper also introduces the integration of mobile app recommendations. By leveraging the recognised emotion, the system recommends relevant mobile applications that are tailored to the user's emotional state. This integration enhances user experience and facilitates emotion regulation through the utilisation of appropriate mobile apps. Experimental evaluations are conducted on benchmark emotion recognition datasets, including the DEAP and MAHNOB_HCI datasets. The findings of the study highlight the effectiveness of the proposed methodology in terms of accuracy and robustness, surpassing existing approaches in the field. Additionally, the integration of the mobile app recommendation system showcases encouraging outcomes by offering personalised recommendations tailored to the user's emotional state.

Saisanthiya, D., Supraja, P. (2023). Heterogeneous Convolutional Neural Networks for Emotion Recognition Combined with Multimodal Factorised Bilinear Pooling and Mobile Application Recommendation. *International Journal of Interactive Mobile Technologies (iJIM)*, 17(16), pp. 129–142. https://doi.org/10.3991/ijim.v17i16.42735

Article submitted 2023-05-01. Resubmitted 2023-06-09. Final acceptance 2023-06-17. Final version published as submitted by the authors.

KEYWORDS
heterogeneous CNN, bilinear pooling, mobile application, recommendation system, multimodal data

1 INTRODUCTION

Emotion, being an integral part of human experience, plays a crucial role in our daily lives, influencing our behaviour, decision-making processes and overall well-being. Emotion recognition has emerged as a fascinating field of study within the broader domain of affective computing, aiming to develop smart systems capable of recognising and responding to human perceptions. In contrast to conventional methods that primarily rely on visual cues like facial expressions, there is an increasing focus on integrating physiological signals to improve the precision and reliability of emotion recognition systems. This shift in approach reflects the growing interest in leveraging physiological data to enhance the accuracy and robustness of such systems.

Among the various physiological signals that hold promise in the domain of emotion recognition, electroencephalography (EEG) has emerged as a powerful modality. EEG records the electrical activity of the brain and provides valuable insights into cognitive and affective states. Its non-invasiveness, high temporal resolution and direct measurement of neural activity make it an ideal candidate for capturing emotional responses in real-time. However, emotion recognition using EEG signals poses significant challenges due to the complex and dynamic nature of emotions, the inherent variability across individuals and the presence of noise in the recordings. To overcome these challenges, recent research efforts have focused on adopting a multimodal approach that combines multiple sources of information, such as facial expressions, physiological signals and behavioural cues, to increase the accuracy and strength of emotion credit systems.

Traditional emotion recognition approaches have often relied on unimodal data, such as facial expressions or speech signals, leading to limited accuracy and robustness. Recognising that emotions are complex and multi-dimensional phenomena, we leverage multiple modalities, including facial expressions, speech signals and physiological data, to capture a comprehensive representation of human affective states. The proposed framework adopts heterogeneous CNNs that are tailor-made for each modality to effectively extract discriminative features from diverse input sources.

To fuse the information from different modalities, we employ multimodal factorised bilinear pooling (MFB) technique. It is a powerful technique that captures cross-modal interactions and exploits the complementary nature of the modalities. By modelling the interactions between features, this pooling method effectively integrates multimodal information while preserving the unique characteristics of each modality, thereby improving the discriminative power of the model. Additionally, we extend the scope of our framework beyond emotion recognition by incorporating mobile application recommendation. Leveraging the insights gained from emotion recognition, we propose a recommendation mechanism that suggests mobile applications tailored to the user's emotional state. By bridging the gap between affective computing and mobile application recommendation, we aim to enhance the user experience and promote personalised interaction between users and their mobile devices.

The contributions of this paper are threefold: First, we propose a novel framework that combines heterogeneous CNNs, multimodal factorised bilinear pooling

and mobile application recommendation to advance the field of emotion recognition. Second, we conduct extensive experiments on benchmark datasets to demonstrate the effectiveness of our approach in accurately recognising emotions across multiple modalities. Third, we present a comprehensive evaluation of the mobile application recommendation component, showcasing the potential of our framework to provide personalised recommendations based on the user's emotional state.

2 LITERATURE REVIEW

Emotion recognition has emerged as a significant research area due to its potential applications in various domains, including healthcare, human-computer interaction and entertainment. Traditional approaches to emotion recognition have primarily focused on unimodal data sources such as facial expressions or speech signals. However, the complex and multi-dimensional nature of emotions necessitates the integration of multiple modalities to achieve a comprehensive understanding of human affective states. In this literature review, we explore existing research in emotion recognition, highlight the limitations of unimodal approaches and emphasise the need for multimodal frameworks.

Unimodal approaches to emotion recognition have demonstrated promising results in specific contexts. Facial expression analysis, for instance, has been extensively studied, leveraging techniques such as facial action coding systems, geometric features, or deep learning-based methods [1]. These methods excel at capturing visual cues but often struggle with the high inter- and intra-subject variability in facial expressions, as well as the influence of external factors such as lighting conditions or occlusions. Similarly, speech-based emotion recognition has been widely explored, utilising features like prosody, pitch and spectral content [2]. While speech provides valuable information about emotional states, it is susceptible to noise, variations in speech patterns and individual differences in vocal expression, making it challenging to achieve robust and accurate emotion recognition solely based on speech signals.

To overcome these constraints, scholars have progressively shifted their focus towards multimodal strategies that combine various data sources, including facial expressions, speech and physiological signals. One popular modality is electroencephalography (EEG), which directly measures brain activity and offers insights into cognitive and affective states. EEG-based emotion recognition has shown promise due to its high temporal resolution and non-invasiveness [3]. However, it suffers from challenges related to signal noise, individual differences and the need for advanced processing techniques. In recent years, CNNs have had a transformative impact on the domain of computer vision, showcasing exceptional capabilities in tasks related to image recognition. Researchers have extended CNNs to handle multimodal data, allowing for the fusion of information from different modalities. One effective technique for multimodal fusion is factorised bilinear pooling, which models interactions between features across modalities and captures their joint representations [4]. By combining the strengths of CNNs and factorised bilinear pooling, researchers have achieved significant improvements in multimodal emotion recognition, thereby enhancing the discriminative power of the models.

One area where heterogeneous CNNs have shown promise is in multimodal learning, where information from different modalities, such as images, text and audio, is combined to enhance the learning process. For example, in image captioning, heterogeneous CNNs can integrate visual features extracted from images with

textual features to generate more accurate and meaningful captions [5]. Similarly, in video analysis, combining visual and audio information using heterogeneous CNNs has led to improved action recognition and event detection [6]. Another application domain where heterogeneous CNNs have gained attention is medical image analysis. Medical images often come in different modalities, such as MRI, CT, or PET scans, and require specialized processing techniques. Heterogeneous CNNs have been employed to integrate information from multiple modalities to improve disease classification, tumour segmentation and diagnosis accuracy [7].

In a recent study, a novel multimodal fusion approach was introduced, which integrated audio and visual information using a linear latent-space mapping. The researchers utilised a Dempster-Shafer theory-based evidence fusion technique to project features into a cross-modal space and combine them with the textual modality. The evaluation conducted on the DEAP [8] dataset demonstrated the superiority of this approach compared to other comparative methods. Over the past few years, significant advancements have been made in multimodal emotion recognition, with CNNs playing a pivotal role. For example, an ensemble CNN (ECNN) was proposed to extract features from different modalities and utilise a voting strategy to create an ensemble model for fusing and classifying multimodal signals. Likewise, a hierarchal fusion CNN (HFCNN) [9] was developed to extract and combine emotion-related convolutional features from multimodal signals in an end-to-end manner. Additionally, a CNN architecture was applied to extract facial and EEG features, followed by a voting strategy for emotion classification. The continuous progress in integrating CNNs and other deep learning techniques has resulted in remarkable achievements in multimodal emotion recognition, particularly in improving the feature extraction and fusion processes. These advancements open up exciting opportunities for further exploration and advancement in the field of multimodal deep learning for emotion recognition [10].

Bilinear pooling has emerged as a prominent technique in various fields in recent years. For example, in the domain of acoustic scene classification, Kek et al. proposed a method that employed a dual-flow CNN structure incorporating both time and frequency information. By leveraging bilinear pooling, they successfully fused acoustic features extracted from two CNNs, showcasing the potential of this approach [11]. To effectively recognise driver's emotions, Du et al. introduced CBLNN (Foldable Bidirectional Neural Network with Long-term and Short-term Memory), a novel deep learning framework. This framework utilised multimodal factorised bilinear pooling (MFB) to combine emotion information derived from geometric facial features and heart rate features. Their results demonstrated real-time and efficient emotion detection capabilities [12]. Moreover, Nguyen et al. presented a fusion model based on bilinear pooling that integrated feature vectors encompassing facial expression, posture, physical action and voice. Through their proposed fusion strategy, they achieved effective interaction among the elements of each component vector, capturing the complex and intrinsic relationships among the different modalities. Numerous studies have affirmed the efficacy of bilinear pooling in integrating multimodal signal characteristics, thereby enhancing the performance of multimodal emotion recognition systems. These findings underscore the advantages of bilinear pooling in diverse multimodal emotion recognition tasks.

Furthermore, integrating emotion recognition with mobile application recommendation systems opens exciting possibilities for personalised user experiences. By leveraging the detected emotional states, these systems can adapt recommendations to match the user's current affective context [13]. This integration not only enhances user engagement but also provides opportunities for context-aware and emotionally intelligent mobile applications.

In summary, the literature review highlights the limitations of unimodal approaches to emotion recognition and emphasises the need for multimodal frameworks. The integration of various modalities, such as facial expressions, speech and EEG signals, enables a more comprehensive understanding of human emotions. Leveraging the power of CNNs and factorised bilinear pooling further enhances the discriminative capabilities of the models. Additionally, the combination of emotion recognition with mobile application recommendation systems allows for personalised and context-aware user experiences. The proposed framework in this paper aims to address these challenges by employing heterogeneous CNNs, multimodal factorised bilinear pooling and mobile application recommendation, contributing to the advancement of emotion recognition and its practical applications.

3 METHODOLOGY

The HC-MFB multimodal emotion recognition model, as depicted in Figure 1, consists of four sequential tasks: EEG signal channel selection, heterogeneous feature extraction, multimodal fusion and classification. One of the key steps in this model is the utilization of the Normalized Mutual Information (NMI) method to identify the most relevant channel from the available EEG signal channels. The Hierarchical Convolutional Neural Networks (HCNNs) are employed for extracting heterogeneous features from each modality. These features are then combined using the Modality Fusion Block (MFB) to capture the complementary information across different modalities. To classify emotions, an ensemble strategy is applied to leverage the distinctive characteristics present in various EEG signal bands. By incorporating these components, the HC-MFB model aims to effectively recognise emotions using multimodal inputs, enhancing the overall performance and accuracy of the emotion recognition system. From this classification, mobile applications will recieve recommendation for their wellness.

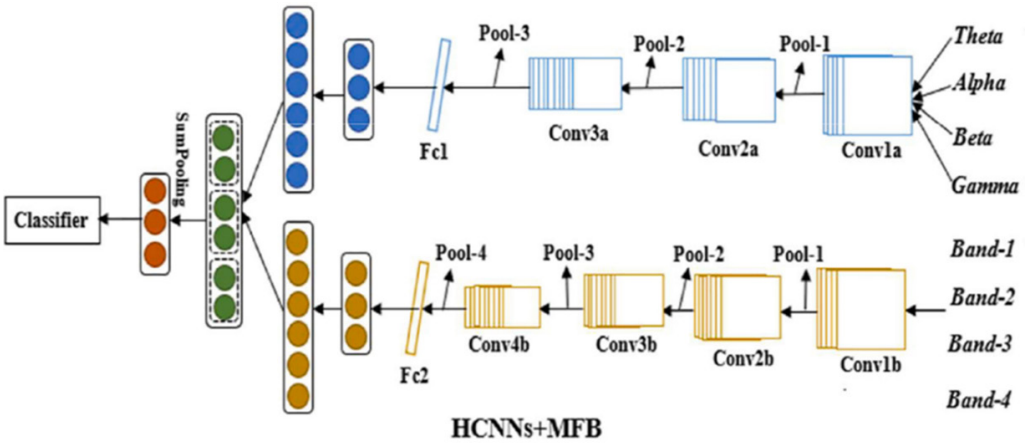


Fig. 1. The proposed HC_MFB model

3.1 EEG selection of channels

The selection of EEG channels is an important step in emotion recognition, as EEG signals provide valuable insights into cerebral cortex activity. However, utilising

the full set of EEG channels may result in redundant data, potentially leading to decreased accuracy in emotion recognition. To address this, we employ the *NMI* method proposed in reference [14] to identify a subset of EEG signal channels.

$$MI(X, Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X, Y) \tag{1}$$

The *NMI* measures the interdependence between two variables and is calculated based on mutual information. Specifically, the mutual information between two channels, *X* and *Y*, can be expressed as the difference between their combined entropy and the sum of their individual entropies (1)

$$NMI(X, Y) = MI(X, Y) / \mathrm{H}(X) + \mathrm{H}(Y) \tag{2}$$

To normalize the mutual information and obtain a value between 0 and 1, the *NMI* formula (2) is utilised.

$$Gn = \int_{i=1}^{N} NMI\, i(X, Y) \tag{3}$$

To generate the connection matrix for channel selection, *NMI* is computed for each pair of channels across all samples. The connection matrix for the ith sample, denoted as *NMIi*, is summed across all samples to obtain the total connection matrix *Gn* (3). In this study, *Gn* is utilised for channel selection based on a predefined threshold. Optimal EEG channels are then selected based on the performance of each channel, aiming to enhance the accuracy of emotion recognition.

3.2 Feature extraction

In the past few years, deep learning has gained significant traction, with many approaches incorporating deep convolutional features to enhance their classification performance [15]. This study employed two distinct neural network architectures, specifically EEG-based Convolutional Neural Network (E-CNN) and Peripheral signals-based Convolutional Neural Network (P-CNN), to extract emotion-related features. These CNNs were trained separately on the DEAP and MAHNOB-HCI datasets, respectively. To automatically extract crucial features, the E-CNN and P-CNN models underwent ten-fold cross-validation on the training sets of their respective datasets. Subsequently, the HC-MFB model was tested using the corresponding testing sets. Through numerous experiments, the internal structure parameters of the HCNNs were determined, albeit with variations in the learning rate and training epoch between the two datasets.

The training of HCNNs on the DEAP dataset utilised a learning rate of 10^{-3}, a batch size of 18 and 30 epochs. Conversely, for the MAHNOB-HCI dataset, the HCNNs were trained with a learning rate of 10^{-4}, a batch size of 15 and 20 epochs. Detailed parameter descriptions for the two CNNs can be found in Table 1 For instance, in the first convolutional layer (Conv 1), there were 16 kernel mappings with a kernel size of 7×7 and a stride length of 1. This was followed by a max-pooling operation with a kernel size of 2×2 and a stride of 2. These specific parameter configurations were selected to achieve optimal training outcomes for each dataset, considering their distinct characteristics and requirements.

Table 1. The structure of CNN

Layer	E-CNN	P-CNN
Conv 4	*	1,3,64
Pool 3	*	2,2
Conv 3	1,2,64	1,5,32
Pool 2	2,2	2,2
Conv 2	1,7,32	1,5,32
Pool 1	2,2	2,2
Conv 1	1,7,16	1,7,16

To enhance the model's performance, a dropout layer was introduced following the last convolutional layer. In a study referenced as [16], it was demonstrated that CNNs can serve as effective emotion classifiers in EEG-based emotion recognition. The researchers highlighted the importance of incorporating batch normalization (BN) layers within the CNN architecture. Consequently, BN layers were added to each CNN utilised in this approach. Stochastic gradient descent (SGD) was employed as the optimisation method, and the loss function employed was cross-entropy. Post-training, the classification layer and softmax activation function were discarded to generate feature vectors. These feature vectors comprise the heterogeneous convolutional features extracted by HCNNs. Subsequently, a process called deep fusion was conducted using the multi-modal fusion block (MFB) to integrate the convolution features effectively.

3.3 Multimodal fusion

In our experiment, we employed a novel fusion technique called MFB. The detailed process is illustrated in Figure 1 After passing through the fully connected layer, the heterogeneous convolution feature vectors were utilised as input for the MFB. This pooling operation involved two feature vectors of different forms: $x \in Rm$ and $y \in Rn$. In this approach, a multimodal bilinear model was employed, which can be mathematically described as follows:

$$Zi = xTWiy \tag{4}$$

Here, $Zi \in R$ represents the output of the bilinear model, while $Wi \in Rm \times n$ represents the projection matrix used in the process. This formulation allows for the effective fusion of the two feature vectors using the MFB technique.

3.4 Classification

This paper utilises ensemble learning for the classification of multimodal signals. Ensemble learning involves training and learning multiple base learners independently and then combining a subset of them based on their individual learning performance. This approach effectively mitigates the issue of overfitting that can arise when using a single base classifier. Various ensemble learning methods, such

as boosting, bagging and stacking, have been developed and employed [17]. In this study, the fusion of the four bands of EEG signals with peripheral signals, or eye movement signals, is conducted. The weak supervised models from all three bands are combined, and a strong supervised model is obtained through majority voting to enhance the effectiveness of the recognition model.

3.5 Mobile application recommendation

The mobile app recommendation module plays a crucial role in the EEG-based emotion recognition recommendation system implemented on a mobile application. This module is responsible for leveraging recognised user emotions based on EEG data to provide personalised app recommendations. The following sections describe the key aspects of the mobile app recommendation module:

Recommendation engine. The recommendation engine is the core component of the module. It utilises the user's emotional state, determined through EEG-based emotion recognition, to generate app recommendations tailored to the user's preferences. The engine employs algorithms such as collaborative filtering, content-based filtering, or hybrid approaches to analyse user data and app characteristics for generating personalised recommendations.

User profile. To provide accurate and relevant app recommendations, the module maintains a user profile that includes information about the user's preferences, past app interactions and emotional states derived from the EEG data. The user profile is continuously updated as the user interacts with the recommended apps and provides feedback, enabling the recommendation engine to refine and adapt its recommendations over time.

App database. The module relies on a comprehensive app database that stores information about various mobile applications, including their features, categories, ratings, reviews and user feedback. This database serves as a knowledge base for the recommendation engine to match user preferences and emotional states with relevant apps.

Real-time processing. The mobile app recommendation module is designed to operate in real-time, providing instant recommendations based on the user's current emotional state. It utilizes efficient algorithms and data structures to handle the computational demands within the mobile application, ensuring a seamless and responsive user experience.

User interface. The recommendation module interfaces with the mobile application's user interface to present the app recommendations to the user. It may display recommended apps in a visually appealing manner, showcasing relevant information such as app names, icons, descriptions, ratings and user reviews. The user interface also allows users to provide feedback on the recommended apps, contributing to the continuous improvement of the recommendation system.

Privacy and data security. The module incorporates measures to ensure user privacy and data security. It adheres to best practices for handling sensitive EEG data, encrypting user information and complying with relevant data protection regulations. User consent and transparency in data usage are emphasised to maintain user trust.

Performance monitoring and evaluation. To evaluate the effectiveness of the mobile app recommendation module, performance metrics such as recommendation accuracy, user engagement and user satisfaction can be monitored and

analysed. These metrics provide insights into the performance of the module and guide further enhancements.

The mobile app recommendation module in the EEG-based emotion recognition recommendation system is responsible for leveraging user emotions derived from EEG data to generate personalised app recommendations. By considering the user's emotional state, preferences and the emotional context of apps, the module enhances the user experience by providing relevant and engaging app recommendations.

4 EXPERIMENT AND RESULTS

4.1 Dataset preparation

In this study, a comprehensive database was created consisting of 32 subjects who underwent examination. To serve as trigger stimuli, 40 videos were carefully selected, each with a duration of 63 seconds. The database also included recordings of the participants' central nervous system activity, peripheral physiological systems and facial expressions. Following the viewing of each video, participants were requested to conduct self-assessments related to valence, arousal, dominance and liking.

The objective of the MAHNOB-HCI database is to capture multimodal emotions by recording signals from 30 individuals as they watch a series of 20 videos. These signals encompass various aspects, including central nervous system activity, peripheral physiological signals and eye movement signals. After each video, participants provide ratings for arousal, valence, control and predictability dimensions to describe their emotional experiences. To ensure data consistency, the middle 30 seconds of each video were chosen for analysis, accounting for variations in video duration. Unfortunately, three individuals encountered issues with equipment and recordings, resulting in corrupted data, while data from two individuals were incomplete. Consequently, the analysis focused on the remaining 25 participants. The signals were down sampled to 256 Hz, and the EEG channels AF3, FC1, F4, CP1, CP2 and PZ were utilised. The processing methods employed mirrored those of the DEAP dataset.

4.2 Results and discussions

In our research, we implemented the fusion technique discussed in Section III to combine the convolutional features of our models. Specifically, we utilized the HCNNs model to extract features from both the EEG signals and PPS signals. To examine the interactions between the bands of EEG signals for each emotion, we employed an ensemble classifier to perform emotion recognition on different combinations of these bands after fusing the multimodal signals. The classification accuracies achieved on the DEAP dataset and MAHNOB-HCI dataset are presented in Figures 2 and 3, respectively. Each point on the graphs represents the average accuracy obtained through ten-fold cross-validation on the respective datasets. The solid lines, displayed in different colours, illustrate the combinations of various wavebands, while the dotted lines depict the final average accuracy obtained from five 10-fold cross-validations.

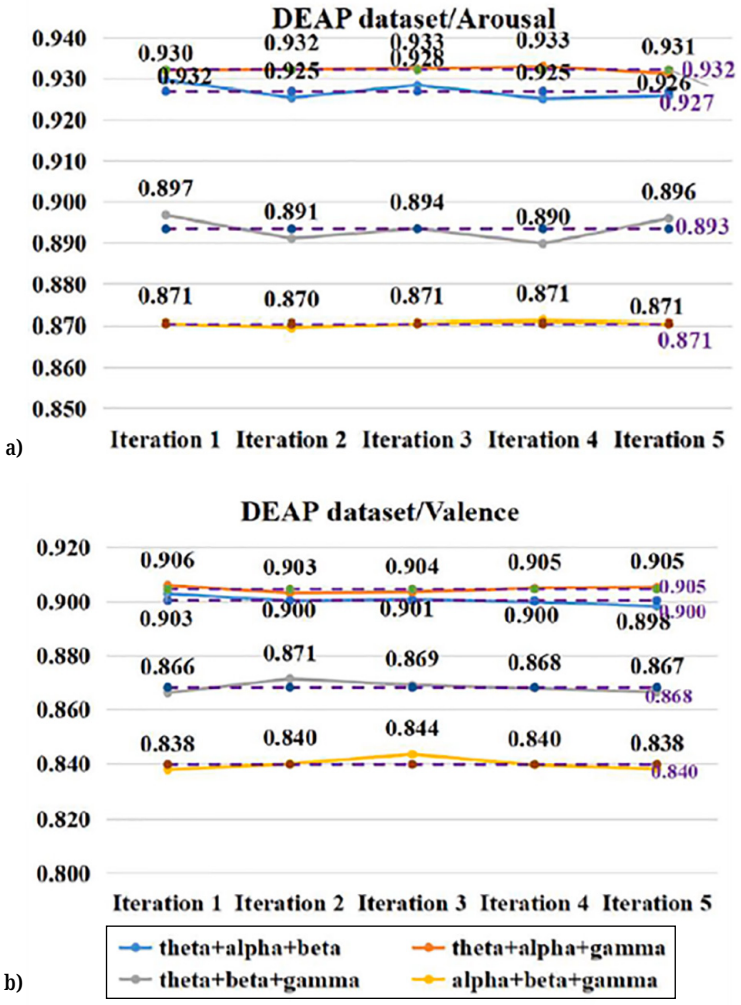


Fig. 2. The results of tenfold cross validation DEAP dataset for EEG a d PPS for a) arousal b) valence

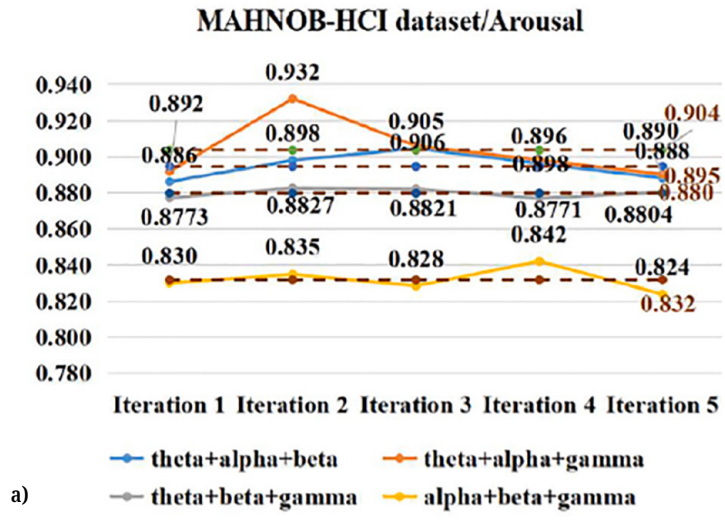
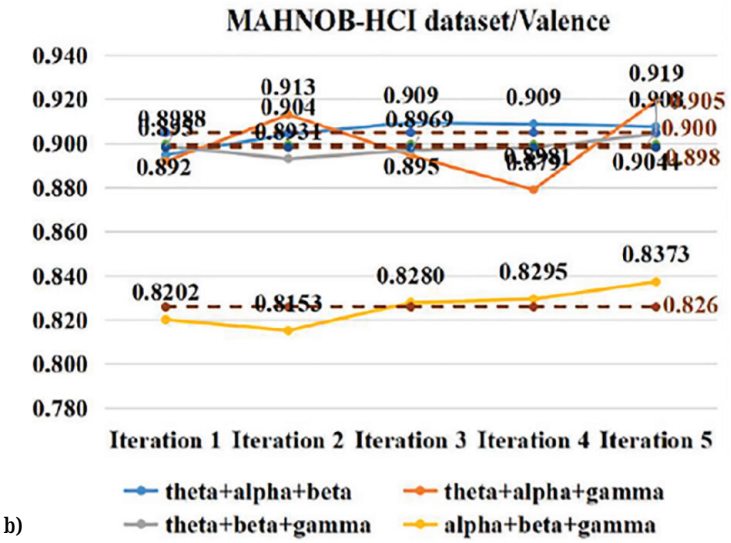


Fig. 3. *(Continued)*



b)

Fig. 3. The results of tenfold cross validation MAHNOB-HCI dataset for EEG a d PPS for a) arousal b) valence

Based on the classification outcomes illustrated in Figure 3 for the MAHNOB-HCI dataset, the combinations of theta, alpha and gamma bands resulted in the highest average accuracies for the arousal dimension and valence dimension, achieving 90.37% and 90.50%, respectively. Specifically, concerning the arousal dimension, the fusion of theta, alpha and gamma bands led to the highest accuracy, while for the valence dimension, the fusion of theta, alpha and beta bands yielded the highest accuracy.

Table 2 presents the classification results of various methods that fuse multimodal DEAP datasets for emotion classification. In our study, we employ multimodal bilinear pooling neural networks to conduct ensemble classification of emotional states. Our proposed method achieves the highest accuracy of 93.22% in the arousal dimension and an accuracy of 90.46% in the valence dimension. When comparing our method to the DCNN approach, we observe that the HCNNs and MFB models outperform the DCNN method specifically in the arousal dimension. This improvement can be attributed to the HCNNs and MFB models' capability to automatically extract and fuse deep features. Interestingly, our method achieves superior classification performance by exclusively utilizing the combination of three bands.

Table 2. The comparison results on DEAP dataset

Fusion Method	Arousal	Valence
MDBN	87.32%	83.69%
DCNN	92.92%	92.24%
HC_MFB	93.21%	90.46%

Table 3 presents the classification results of various methods that fuse multimodal MAHNOB_HCI datasets for emotion classification. The best results are shown in the Table 3, using our proposed method HC_MFB for the arousal and valence dimensions.

Table 3. The comparison results on MAHNOB-HCI dataset

Fusion Method	Arousal	Valence
Maltitask CNN	74.17%	75.21%
Deep Learning	80.41%	80.76%
HC_MFB	90.36%	90.49%

Now we can integrate our more accurate results with the following mobile application designed for emotion recommendation.

Emotion tracker. This mobile application utilises advanced machine learning algorithms to precisely recognise and track emotions based on user input, such as facial expressions, voice recordings, or text analysis. By analysing these inputs, the application provides detailed insights into emotional patterns and trends over time, helping users gain a better understanding of their emotions.

MoodMeter. This mobile app employs a combination of self-assessment and machine learning techniques to classify and track users' emotions. It allows users to input their emotional state through a user-friendly interface and provides real-time feedback and suggestions for managing emotions effectively.

EmoSense. This mobile application incorporates multimodal emotion recognition, including facial expressions, speech analysis and physiological signals, to provide a comprehensive understanding of users' emotional states. It offers personalised recommendations and techniques for emotional well-being based on the classification results.

Feelings diary. This app allows users to keep a digital diary of their emotions throughout the day. By capturing text entries, images and audio recordings, the app employs sentiment analysis and machine learning algorithms to classify and analyze emotional patterns, helping users identify triggers and manage their emotions effectively.

5 CONCLUSION

In conclusion, the combination of heterogeneous CNNs and multimodal factorised bilinear pooling has demonstrated promising results in emotion recognition, particularly when applied to EEG and eye movement data. By leveraging these techniques, mobile applications can harness the power of EEG signals and eye movement data to provide users with accurate and real-time emotion recognition capabilities. These applications have the potential to offer personalised insights into users' emotional states, enhance self-awareness and promote emotional well-being. The integration of EEG and eye movement data in emotion recognition allows for a more comprehensive understanding of users' emotional experiences. Mobile applications utilizing this approach can provide valuable feedback and recommendations tailored to individual users based on their unique patterns of brain activity and eye movements.

The recommendation of mobile applications that incorporate heterogeneous CNNs and multimodal factorised bilinear pooling for EEG and eye movement data empowers users to actively monitor and manage their emotions. These applications can provide valuable tools for self-reflection, emotional tracking and fostering emotional resilience. As research in this field continues to advance, it is expected that

mobile applications for emotion recognition using EEG and eye movement data will become increasingly accurate, user-friendly and accessible. Such advancements have the potential to revolutionise the way individuals understand, monitor and regulate their emotions, ultimately contributing to improved mental well-being and emotional health.

6 REFERENCES

[1] A.T. Lopes and D.R. de Almeida, "Emotion recognition from facial expressions: A survey. Computer Vision and Image Understanding," vol. 189, p. 102824, 2020.

[2] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the Munich open-source multimedia feature extractor," In *Proceedings of the 21st ACM international conference on Multimedia*, pp. 835–838, 2013. https://doi.org/10.1145/2502081.2502224

[3] D. Zhang, L. Yao, and S.M. Zhou, "Emotion recognition from EEG signals using multi-dimensional information in EMD domain," *IEEE Transactions on Affective Computing*, vol. 11, no. 1, pp. 140–152, 2019.

[4] T.Y. Lin, A. Roy Chowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1449–1457, 2018.

[5] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3128–3137, 2015. https://doi.org/10.1109/CVPR.2015.7298932

[6] J.Y. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4694–4702, 2015.

[7] G. Litjens, T. Kooi, B.E. Bejnordi, A.A. Setio, F. Ciompi, M. Ghafoorian, and C.I. Sanchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017. https://doi.org/10.1016/j.media.2017.07.005

[8] S. Nemati, R. Rohani, M.E. Basiri, M. Abdar, N.Y. Yen, and V. Makarenkov, "A hybrid latent space data fusion method for multimodal emotion recognition," In *IEEE Access*, vol. 7, pp. 172948–172964, 2019. https://doi.org/10.1109/ACCESS.2019.2955637

[9] H. Huang, Z. Hu, W. Wang, and M. Wu, "Multimodal emotion recognition based on ensemble convolutional neural network," *IEEE Access*, vol. 8, pp. 3265–3271, 2020. https://doi.org/10.1109/ACCESS.2019.2962085

[10] Y. Zhang, C. Cheng, and Y. Zhang, "Multimodal emotion recognition using a hierarchical fusion convolutional neural network," *IEEE Access*, vol. 9, pp. 7943–7951, 2021. https://doi.org/10.1109/ACCESS.2021.3049516

[11] M. Wu, W. Su, L. Chen, W. Pedrycz, and K. Hirota, "Two-stage fuzzy fusion based-convolution neural network for dynamic emotion recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 805–817, 2022. https://doi.org/10.1109/TAFFC.2020.2966440

[12] J. Shukla, M. Barreda-Ángeles, J. Oliver, G.C. Nandi, and D. Puig, "Feature extraction and selection for emotion recognition from electrodermal activity," *IEEE Transactions on Affective Computing*, vol. 12, no. 4, pp. 857–869, 2021. https://doi.org/10.1109/TAFFC.2019.2901673

[13] L. Chen and P. Pu, "A survey of personalization approaches for mobile recommendation," *ACM Computing Surveys (CSUR)*, vol. 51, no. 3, pp. 1–34, 2018. https://doi.org/10.1145/3190507

[14] Z.-M. Wang, S.-Y. Hu, and H. Song, "Channel selection method for EEG emotion recognition using normalized mutual information," *IEEE Access*, vol. 7, pp. 143303–143311, 2019. https://doi.org/10.1109/ACCESS.2019.2944273

[15] L. Wu, Y. Wang, X. Li, and J. Gao, "Deep attention-based spatially recursive networks for fine-grained visual recognition," *IEEE Transactions on Cybernetics*, vol. 49, no. 5, pp. 1791–1802, 2019. https://doi.org/10.1109/TCYB.2018.2813971

[16] W.-C. Fang, K.-Y. Wang, N. Fahier, Y.-L. Ho, and Y.-D. Huang, "Development and validation of an EEG-based real-time emotion recognition system using edge AI computing platform with convolutional neural network system-on-chip design," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 4, pp. 645–657, 2019. https://doi.org/10.1109/JETCAS.2019.2951232

[17] Chen Wei, Lan-lan Chen, Zhen-zhen Song, Xiao-guang Lou, and Dong-dong Li, "EEG-based emotion recognition using simple recurrent units network and ensemble learning," *Biomedical Signal Processing and Control*, vol. 58, p. 101756, 2020. https://doi.org/10.1016/j.bspc.2019.101756

7 AUTHORS

D. Saisanthiya received a B.Tech. Degree in CSE from Arulmigu Meenakshi Amman College of Engineering, Thiruvannamalai, affiliated to Anna University, Tamil Nadu, in 2009 and M.Tech Degree in CSE from Sastha Institute of Science and Technology, Chembarambakkam, affiliated to Anna University, Tamil Nadu in 2011. She is currently working towards the Ph.D. degree at the Department of Networking and Communications, SRM University, Kattankulathur, Tamil Nadu, India. Her research interests include deep learning and Machine learning algorithms (E-mail: saisantd@srmist.edu.in).

Dr. P. Supraja is currently working as Associate Professor, Department of Networking and Communications in SRM Institute of Science and Technology Kattankulathur, India. She is a recipient of AICTE Visvesvaraya Best Teacher Award 2020. She completed Indo-US WISTEMM Research fellowship at University of Southern California, Los Angeles, USA, funded by IUSSTF and DST Govt., of India. She served as a Post-Doctoral Research Associate at Northumbria University, Newcastle, UK and completed her Ph.D. from Anna University in 2017. She has published more than 50 research papers in reputed national and international level journals/conferences. She received university-level Best Research Paper Award in 2019 and 2022. Also, she has received funding from AICTE for conducting STTP. Her research interests include Cognitive Computing, Optimization algorithms, Machine learning, Deep Learning, Wireless Communication, and IoT. She is a reviewer in IEEE, Interscience, Elsevier and Springer journals. She is also a member of several national and international professional bodies including IEEE, ACM, ISTE, etc. In addition, she has received the young women in Engineering award and Distinguished Young Researcher award from various international organizations (E-mail: suprajap@srmist.edu.in).