

Enhancing Depression Detection through Multimodal Emotion Recognition and Wearable Technologies

Mingran Dong^{1,*}, Xinrui Wang²

¹School of Advanced Technology, Xi'an JiaotongLiverpool University, Suzhou, China

²Faculty of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, China

* Corresponding Author Email: Mingran.Dong22@student.xjtlu.edu.cn

Abstract. Traditional methods for diagnosing and treating depression often lack the precision to accurately identify the diverse symptoms and parameters associated with the disorder. With advancements in artificial intelligence and key technologies, multimodal emotion recognition has emerged as a promising approach to enhance depression detection. This paper explores the integration of physiological data such as EEG (electroencephalogram), ECG (electrocardiogram), and GSR (galvanic skin response), along with data from wearable devices, to develop a high-quality feature extraction method for emotion analysis. Multiple datasets, including the SEED dataset involving EEG recordings from 15 participants during emotional video viewing, were utilized to collect comprehensive physiological and psychological data. Feature extraction involved preprocessing steps like noise reduction and normalization, followed by the application of decision-level and representation-level fusion algorithms. Heart rate variability and skin conductance response features were analyzed to capture emotional indicators. An attention-based Convolutional Neural Network (CNN) was employed to model both local and global facial features, and the Weighted Sum Pooling and Projection (WSP) method was used for depression representation extraction. Multimodal fusion techniques, particularly model-level fusion, significantly enhanced the learning of internal multimodal interactions, leading to improved recognition rates in Automatic Depression Estimation tasks compared to previous algorithms. The applications of this approach extend to smart home integration, daily emotion monitoring, health warning systems, and interactive wearable technologies, highlighting its potential for widespread use in enhancing user experience and mental health interventions.

Keywords: Multimodal Emotion Recognition, Depression Detection, Wearable Technologies, Feature Extraction.

1. Introduction

1.1. Background and Motivation

Depression is a pervasive mental health disorder affecting millions of individuals worldwide. Traditional methods of diagnosing and treating depression primarily rely on self-reported symptoms, clinical interviews, and standardized questionnaires [1]. These approaches often lack the precision to capture the multifaceted nature of depression, leading to underdiagnosis or misdiagnosis. Moreover, they may not effectively identify the varying symptoms and parameters that differ from one individual to another, thereby limiting the efficacy of subsequent treatment plans [2].

In recent years, advancements in artificial intelligence (AI) have opened new avenues for enhancing mental health diagnostics. AI technologies, particularly in the field of emotion recognition, offer the potential to analyze complex patterns in data that are not easily discernible through conventional methods [3]. Emotion recognition systems can process vast amounts of data from various sources, providing a more nuanced understanding of an individual's emotional state. The integration of AI into mental health assessment tools promises to address the limitations of traditional diagnostic methods by offering more accurate, objective, and personalized evaluations.

1.2. Importance of Multimodal Emotion Recognition

Multimodal emotion recognition involves the analysis of emotional states using multiple types of data or signals, such as physiological measurements, facial expressions, speech patterns, and textual content. This approach leverages the strengths of different data modalities to provide a comprehensive assessment of an individual's emotional condition [4]. By combining information from various sources, multimodal systems can capture subtle emotional cues that might be missed when relying on a single modality.

The significance of multimodal emotion recognition lies in its ability to improve the accuracy and reliability of emotion detection. For instance, physiological signals like EEG, ECG, and GSR provide objective measures of the body's responses to emotional stimuli. When these signals are analyzed alongside behavioral data from wearable devices or audiovisual inputs, the system can generate a more holistic profile of the individual's emotional state [5]. This comprehensive insight is particularly valuable in the context of depression detection, where symptoms can manifest differently across individuals and may not always be externally observable.

1.3. Objectives of the Study

The primary objective of this study is to develop a high-quality feature extraction method for depression analysis by integrating multimodal emotion recognition techniques with data from wearable technologies. Specifically, the study aims to:

Develop an advanced feature extraction framework that effectively captures relevant physiological and psychological indicators associated with depression. This involves preprocessing data to reduce noise, normalizing inputs, and employing sophisticated algorithms to extract meaningful features from EEG, ECG, GSR, and wearable device data [6].

Demonstrate the advantages of multimodal fusion in emotion recognition by comparing the performance of different fusion techniques, such as feature-level, decision-level, and model-level fusion. The study seeks to show that model-level fusion, in particular, enhances the learning of internal multimodal interactions, leading to improved accuracy in depression detection [7].

By achieving these objectives, the study intends to contribute to the field of mental health diagnostics by providing a more sensitive and accurate method for detecting depression [8]. The findings are expected to have practical implications for the development of emotion-aware systems in smart homes, wearable technologies for daily emotion monitoring, and health warning systems that can offer timely interventions.

2. Data Collection and Datasets

2.1. Overview of Collected Data

This study leverages a diverse range of physiological and psychological data to capture the multifaceted nature of emotional states associated with depression. The collected data types include: Physiological Data: Electroencephalogram (EEG), Electrocardiogram (ECG), and Galvanic Skin Response (GSR) signals [9]. Wearable Device Data: ECG readings and accelerometer data obtained from wearable sensors. Psychological Data: Emotional responses recorded during exposure to emotion-eliciting stimuli, such as videos. By integrating these data types, the study aims to create a comprehensive dataset that facilitates high-quality feature extraction for emotion recognition and depression analysis.

2.2. Description of Datasets

2.2.1. EEG, ECG, and GSR data.

The EEG, ECG, and GSR data were collected in a controlled laboratory environment using standard biosensing equipment: EEG Data: Participants wore EEG caps equipped with multiple electrodes placed according to the international 1020 system. The EEG signals capture neural activity across different brain regions, providing insights into cognitive and emotional processes [10]. ECG Data: ECG signals were recorded using electrodes attached to the participants' chest areas. ECG data measure the electrical activity of the heart, offering information on heart rate and heart rate variability (HRV), which are sensitive indicators of emotional arousal and stress levels. GSR Data: GSR sensors were placed on the fingertips or palmar surfaces to measure skin conductance levels. GSR data reflect changes in sweat gland activity, which correlate with sympathetic nervous system activation during emotional experiences. These physiological signals are critical for emotion analysis as they provide objective, quantifiable measures of the body's responses to emotional stimuli. The simultaneous recording of EEG, ECG, and GSR data allows for a multidimensional understanding of emotional states from both central (brain activity) and peripheral (autonomic responses) perspectives.

2.2.2. Wearable device data.

Inclusion of ECG and Accelerometer Data. Wearable devices were utilized to collect ECG and accelerometer data in realworld settings: ECG Data: Wearable ECG monitors provided continuous heart rate and HRV measurements during participants' daily activities. Accelerometer Data: Built-in accelerometers in the wearable devices captured motion and activity levels, offering contextual information about the participants' physical states. Applications in Stress and Emotion Detection. The wearable device data enhance the study's ecological validity by capturing physiological responses in naturalistic environments. ECG data from wearables enable the detection of stress and emotional fluctuations throughout the day, while accelerometer data help differentiate between emotional changes due to physical activity and those arising from psychological factors. This integration supports the development of emotion recognition models that are applicable beyond laboratory conditions.

2.2.3. SEED dataset.

Details about the EEG Data Collected from 15 Participants. The SJTU Emotion EEG Dataset (SEED) is a well-established dataset comprising EEG recordings from 15 participants. The participants were exposed to a series of emotion-inducing video clips designed to elicit positive, neutral, and negative emotional responses. EEG data were collected using a 62-channel cap, capturing high-resolution neural activity during the emotional experiences. Use in Emotion State Recognition During Emotional Video Viewing. The SEED dataset serves as a valuable benchmark for emotion recognition research. By analyzing the EEG signals recorded while participants viewed the emotional videos, patterns corresponding to different emotional states can be identified. The dataset's standardized protocol and rich annotation facilitate the training and evaluation of machine learning models aimed at classifying emotional states based on EEG features.

2.3. Data Acquisition Methods

Procedures for Collecting and Preprocessing the Data. Data acquisition involved several key steps to ensure the quality and reliability of the datasets: Participant Recruitment: Individuals were recruited based on inclusion criteria relevant to the study's objectives. Demographic information was collected to account for potential confounding variables. Laboratory Data Collection: In controlled settings, participants were equipped with EEG caps and sensors for ECG and GSR measurements. They were then exposed to standardized emotional stimuli, such as images, sounds, or videos, while physiological signals were recorded. Wearable Device Data Collection: Participants were instructed on how to use the wearable devices and wore them during their daily routines over a specified period. Data logged included continuous ECG readings and accelerometer measurements. Data

Preprocessing: Noise Reduction: Techniques such as bandpass filtering, artifact removal algorithms (e.g., Independent Component Analysis for EEG), and signal smoothing were applied to eliminate noise from muscle movements, eye blinks, and environmental interference. **Signal Synchronization:** Timestamping and alignment procedures ensured that data from different sensors and modalities were synchronized for accurate analysis. **Normalization:** Data were normalized to reduce intersubject variability, making the features more comparable across participants. **Segmentation:** Continuous signals were segmented into epochs corresponding to specific stimuli or time windows of interest. **Feature Extraction Preparation:** Preprocessed signals were transformed into representations suitable for feature extraction, including time domain and frequency domain analyses. **Ethical Considerations and Participant Consent.** Ethical compliance was paramount throughout the study: **Informed Consent:** Participants were provided with detailed information about the study's purpose, procedures, potential risks, and benefits. Written informed consent was obtained before participation.

3. Theoretical Analysis and Feature Extraction

3.1. Data Preprocessing

Effective data preprocessing is crucial for enhancing the quality of feature extraction and subsequent emotion recognition. The raw physiological signals and wearable device data often contain noise and artifacts resulting from various sources such as environmental interference, muscle movements, and equipment limitations. To address these issues, the following noise reduction and data normalization techniques were employed: **Noise Reduction: Filtering:** Bandpass filters were applied to the EEG, ECG, and GSR signals to retain frequencies relevant to emotional processing while eliminating irrelevant frequencies. For instance, EEG data were filtered within the 0.5–45 Hz range to focus on typical brain wave activities associated with emotions. **Artifact Removal:** Techniques like Independent Component Analysis (ICA) were used to identify and remove artifacts caused by eye blinks, muscle movements, and power line noise in EEG data. **Signal Smoothing:** Moving average and median filters helped reduce random fluctuations and smooth out the signals for better feature extraction. **Data Normalization: ZScore Normalization:** This statistical method standardized the data by subtracting the mean and dividing by the standard deviation, ensuring that features from different participants and sessions were comparable. **MinMax Scaling:** Used to scale the data to a fixed range, typically between 0 and 1, which is essential for algorithms sensitive to the scale of input data. **Baseline Correction:** Physiological signals were adjusted based on baseline measurements taken during a neutral state, allowing for the isolation of emotion-induced changes. These preprocessing steps enhanced the signal-to-noise ratio and standardized the data, forming a solid foundation for accurate feature extraction and emotion recognition.

3.2. Feature Extraction Methods

Feature extraction transforms preprocessed data into meaningful representations that can be used for classification and analysis. Various methods were employed to capture the distinct characteristics of each data modality.

3.2.1. Decision level fusion algorithms.

Decision level fusion involves combining the outputs of multiple classifiers, each trained on different data modalities, to make a final decision. The implementation steps included: **Individual Classifier Training:** Separate classifiers were trained on features extracted from EEG, ECG, GSR, and wearable device data using algorithms such as Support Vector Machines (SVM), Random Forests, or Neural Networks. **Output Generation:** Each classifier produced probabilistic outputs or class labels indicating the detected emotional state or depression indicators. **Fusion Strategies: Majority Voting:** The class most frequently predicted by the individual classifiers was selected as the final decision. **Weighted Voting:** Classifier outputs were weighted based on their individual performance metrics, giving more influence to more accurate classifiers. **Bayesian Combination:** Probabilistic outputs were

combined using Bayesian inference to calculate the posterior probabilities for each class. Recognition Rates: The decisionlevel fusion approach improved overall recognition rates compared to individual classifiers. For example, the fusion method achieved an average accuracy increase of 5–10% over the bestperforming single modality classifier, demonstrating the benefit of combining multiple sources of information at the decision stage.

3.2.2. Representationlevel fusion algorithms.

Representationlevel fusion, also known as featurelevel fusion, involves combining features from different modalities into a single feature vector before classification. The implementation included: Feature Concatenation: Features extracted from EEG (e.g., power spectral densities), ECG (e.g., heart rate variability metrics), GSR (e.g., skin conductance levels), and wearable devices were concatenated into a unified feature vector. Dimensionality Reduction: Techniques like Principal Component Analysis (PCA) or tDistributed Stochastic Neighbor Embedding (tSNE) were applied to reduce the dimensionality and mitigate the curse of dimensionality. Classifier Training: A single classifier was trained on the fused feature vector to recognize emotional states or depression indicators. Comparison: Representationlevel fusion often led to better performance than decisionlevel fusion because it allowed the classifier to learn correlations and interactions between different modalities at the feature level. However, it also posed challenges such as increased computational complexity and the potential for overfitting due to the larger feature space.

3.2.3. Heart rate variability analysis.

Calculation of Heartbeat Intervals and Total Counts. Heart Rate Variability (HRV) is a vital indicator of autonomic nervous system activity and emotional states. The analysis involved: RPeak Detection: ECG signals were processed to identify the Rpeaks (the highest points in the QRS complex) using algorithms like the PanTompkins method. InterBeat Intervals (IBI): The time intervals between consecutive Rpeaks (RR intervals) were calculated to obtain the IBI series. TimeDomain Metrics: Mean HR: Average heart rate over a specified period. SDNN: Standard deviation of NN intervals (normaltonormal beats), reflecting overall HRV. RMSSD: Root mean square of successive differences between adjacent NN intervals, sensitive to parasympathetic activity. FrequencyDomain Metrics: LF and HF Components: Lowfrequency (0.04–0.15 Hz) and highfrequency (0.15–0.40 Hz) components were extracted using spectral analysis to assess sympathetic and parasympathetic balance. Nonlinear Measures: Poincaré plots and entropy measures provided additional insights into the complexity of heart rate dynamics. These HRV features were critical for detecting stress and emotional arousal associated with depression.

3.2.4. Skin conductance response features.

Peak Signal Function Analysis. GSR data reflect changes in skin conductance due to sweat gland activity, which is influenced by emotional and sympathetic nervous system responses. Feature extraction included: Signal Decomposition: The GSR signal was decomposed into tonic (slowvarying baseline) and phasic (rapid fluctuations) components using techniques like Continuous Decomposition Analysis (CDA). Peak Detection: Skin Conductance Responses (SCRs): Phasic increases in conductance were identified as SCR peaks. Amplitude and Latency: The magnitude of each SCR peak and the time delay from stimulus onset were measured. Statistical Features: Mean and Variance: Average conductance levels and variability provided baseline information. Frequency of SCRs: The number of SCRs per unit time indicated the level of emotional arousal. Derivative Features: The first and second derivatives of the GSR signal captured the rate of change, offering additional cues about emotional dynamics. Analyzing these features enabled the detection of emotional responses related to stress and depression.

3.2.5. Wearable device data analysis.

Wearable devices provided continuous monitoring of physiological and behavioral data in naturalistic settings. Feature extraction focused on: Physiological Signals: mECG Features: Similar HRV metrics as described in section 3.2.3 were calculated from wearable ECG data. Electrodermal

Activity (EDA): If available, EDA signals were processed to extract features analogous to GSR analysis. Motion and Activity Data: Accelerometer Features: Statistical measures such as mean, variance, and energy of movement in different axes were calculated. Activity Recognition: Machine learning models classified physical activities (e.g., walking, sitting) to contextualize physiological responses. Contextual Features: Time of Day: Circadian patterns in physiological signals were considered. Environmental Factors: Data on ambient temperature or light exposure, if available, were included. Emotion and Stress Indicators; Stress Levels: Algorithms like the Stress Index or proprietary stress scores provided by device manufacturers were used. Emotion Classification: Realtime classifiers predicted emotional states based on the combination of physiological and activity features. By extracting these features, the wearable device data enriched the multimodal dataset and improved the robustness of emotion and depression detection models.

4. Multimodal Emotion Recognition Fusion

4.1. Feature Concatenation

To harness the complementary strengths of different data modalities, the initial step involves the concatenation of features extracted from each modality into a unified representation. This process combines features from speech, text, heart rate, skin conductance, and wearable devices, creating a comprehensive feature vector for each data instance. Speech Features: Acoustic features such as pitch, tone, intensity, formants, and Mel-frequency cepstral coefficients (MFCCs) are extracted from audio recordings. These features capture vocal nuances that correlate with emotional states and depressive symptoms. Text Features: Textual data, obtained from transcriptions of speech or written inputs, are analyzed using Natural Language Processing (NLP) techniques. Features include sentiment polarity, emotion scores, term frequency-inverse document frequency (TFIDF) values, and syntactic patterns, which help identify linguistic cues associated with depression. Heart Rate Features: Derived from ECG data, heart rate features include time-domain and frequency-domain metrics of heart rate variability (HRV), as detailed in Section 3.2.3. These features reflect autonomic nervous system activity linked to emotional and stress responses. Skin Conductance Features: From GSR data, features such as the number of skin conductance responses (SCRs), amplitude, rise time, and recovery time are extracted. These metrics indicate sympathetic nervous system arousal related to emotional states. Wearable Device Features: Additional physiological and contextual features are obtained from wearable sensors, including accelerometer data for physical activity levels, sleep patterns, and other vital signs that may influence or reflect emotional wellbeing.

By combining these diverse features, the model captures a holistic view of the individual's emotional and physiological state, enhancing the accuracy of emotion recognition and depression detection.

4.2. Attention Based CNN Modeling

To effectively model the complex and high-dimensional multimodal data, an attention-based Convolutional Neural Network (CNN) is employed. This architecture leverages attention mechanisms to focus on the most informative parts of the data, improving feature learning and classification performance.

4.2.1. Local attention mechanisms.

Local attention mechanisms enable the model to concentrate on specific regions or segments within each modality that are most relevant to emotion recognition. Facial Emotion Recognition: In processing facial images or video frames, the CNN applies attention weights to key facial regions such as the eyes, mouth, and eyebrows, where emotional expressions are most evident. This allows the model to detect subtle changes in facial movements that may indicate depressive affect. Speech and Text Analysis: For speech data, the model focuses on critical acoustic frames that carry emotional prosody. In text data, attention weights highlight emotionally charged words or phrases, enabling the

model to capture linguistic indicators of depression. **Physiological Signals:** The attention mechanism identifies significant patterns within heart rate and skin conductance signals, such as abrupt changes or sustained anomalies that correlate with emotional shifts.

4.2.2. Global attention and facial region modeling.

Global attention mechanisms complement local attention by considering the overall context and relationships between different parts of the data. **Holistic Feature Representation:** The CNN integrates global features that represent the overall facial expression, speech patterns, or physiological state. This broader perspective helps in understanding the general emotional context. **Facial Region Modeling:** The model analyzes the interplay between different facial regions, capturing complex expressions that may not be apparent when focusing solely on local features. This is crucial for recognizing composite emotions and depressive symptoms that manifest through subtle facial cues. **CrossModal Relationships:** Global attention facilitates the modeling of interactions between modalities, such as synchrony between vocal tone and facial expressions or correlations between physiological responses and linguistic content. By combining local and global attention mechanisms, the model achieves a more nuanced and accurate representation of the emotional state, which is essential for effective depression detection.

4.3. Depression Representation Extraction

To distill the most relevant features for depression detection from the attentionbased CNN outputs, the Weighted Sum Pooling and Projection (WSPP) method is utilized. **Weighted Sum Pooling:** This pooling strategy aggregates feature maps by computing a weighted sum based on attention scores. Features with higher attention weights contribute more significantly to the pooled representation, ensuring that critical information is emphasized. **Projection:** After pooling, the highdimensional feature vector is projected into a lowerdimensional space using linear transformations or dimensionality reduction techniques such as Principal Component Analysis (PCA). This step reduces computational complexity and mitigates the risk of overfitting. **DepressionSpecific Features:** The resulting compact representation encapsulates the most salient features associated with depressive symptoms across modalities. These features serve as input to the final classification layer or regression model for depression estimation. The WSPP method effectively combines the benefits of attention mechanisms with dimensionality reduction, resulting in a powerful representation that enhances the model's ability to detect depression.

4.4. Multimodal Fusion Techniques

Integrating information from multiple modalities is critical for capturing the full spectrum of emotional cues. Three primary fusion techniques are explored in this study.

4.4.1. Feature level fusion.

Methodology: Features extracted from all modalities are concatenated into a single, unified feature vector before being fed into the model. **Advantages:** Allows the model to learn direct relationships and dependencies between features from different modalities. Facilitates the discovery of complex patterns that span multiple data sources. **Challenges:** Results in highdimensional feature spaces, which can increase computational demands and the risk of overfitting. May require dimensionality reduction techniques to manage complexity.

4.4.2. Decision level fusion.

Methodology: Each modality is processed independently, and separate models generate individual predictions. The final decision is made by combining these predictions using strategies such as majority voting, weighted averaging, or Bayesian fusion. **Advantages:** Simplifies model training by dealing with each modality separately. Reduces the impact of noisy or missing data in any single modality. **Challenges:** May not capture intermodal interactions effectively. Potential loss of information that could be gleaned from the relationships between modalities.

4.4.3. Modellevel fusion

Methodology: Fusion occurs within the model architecture itself, allowing for joint learning from all modalities. This is achieved through models specifically designed for multimodal data, such as multimodal deep neural networks or hybrid architectures that integrate different types of layers. **Importance in Learning Internal Multimodal Interactions:** Enables the model to learn complex and nonlinear relationships between modalities. Facilitates the discovery of synergistic effects where the combination of modalities provides more information than each modality alone. **Advantages:** Often results in superior performance compared to featurelevel or decisionlevel fusion. Better handles missing or incomplete data by leveraging redundant information across modalities. By exploring these fusion techniques, the study identifies the most effective approach for integrating multimodal data in the context of depression detection.

5. Conclusion and Future Work

This study has highlighted the significant reliability and efficiency of multimodal emotion recognition over traditional methods in detecting depression. By integrating various physiological and psychological data sources—such as EEG, ECG, GSR, speech, text, and wearable device data—the proposed system provides a comprehensive analysis of emotional states. The use of attentionbased Convolutional Neural Network (CNN) modeling, combined with advanced fusion techniques like modellevel fusion, enables the effective learning of internal interactions among different modalities. Experimental results demonstrate that this multimodal approach yields higher accuracy and robustness in Automatic Depression Estimation (ADE) tasks compared to previous algorithms, affirming its potential as a powerful tool in mental health diagnostics. The proposed method offers high sensitivity and accuracy in emotion detection, which is crucial for early intervention and treatment of depression. The attention mechanisms within the CNN architecture allow the model to focus on the most informative features across all modalities, enhancing its ability to detect subtle emotional cues associated with depression.

Future research should focus on enhancing data diversity and richness to improve the model's robustness and generalizability. This could involve collecting data from a larger and more varied population, including different age groups, cultural backgrounds, and clinical conditions. Exploring advanced fusion techniques, such as dynamic fusion strategies or the integration of additional modalities (e.g., eyetracking or genetic data), may further optimize the model's performance. Expanding the application of this approach to other emotional and psychological disorders—such as anxiety, bipolar disorder, or posttraumatic stress disorder—could broaden its impact in the field of mental health. Additionally, incorporating personalization and adaptability into the system may enhance its effectiveness, allowing it to adjust to individual differences and changes over time. Addressing these areas will be essential in advancing multimodal emotion recognition as a reliable tool for psychological assessment and intervention.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] D. Ayata, Y. Yaslan, M.E. Kamasak, Emotion recognition from multimodal physiological signals for emotion aware healthcare systems, *J. Med. Biol. Eng.* 40 (2020) 149-157.
- [2] A. Ahmed, J. Ramesh, S. Ganguly, R. Aburukba, A. Sagahyoon, F. Aloul, Investigating the feasibility of assessing depression severity and valence-arousal with wearable sensors using discrete wavelet transforms and machine learning, *Information* 13 (2022) 406.

- [3] H. Kim, T. Hong, Enhancing emotion recognition using multimodal fusion of physiological, environmental, personal data, *Expert Syst. Appl.* 249 (2024) 123723.
- [4] I.T. Meftah, N. Le Thanh, C. Ben Amar, Detecting depression using multimodal approach of emotion recognition, in: 2012 IEEE Int. Conf. Complex Syst. (ICCS), IEEE, 2012, pp. 1-6.
- [5] Y. Jiang, W. Li, M.S. Hossain, M. Chen, A. Alelaiwi, M. Al-Hammadi, A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition, *Information Fusion* 53 (2020) 209-221.
- [6] X. Zhu, H. Xu, Z. Zhao, et al., An Environmental Intrusion Detection Technology Based on WiFi, *Wireless Pers. Commun.* 119 (2021) 1425-1436.
- [7] T. Mullick, A. Radovic, S. Shaaban, A. Doryab, Predicting depression in adolescents using mobile and wearable sensors: multimodal machine learning-based exploratory study, *JMIR Form. Res.* 6 (2022) e35807.
- [8] X. Cheng, X. Wang, T. Ouyang, Z. Feng, K.F. Shad, K.H. Dogan, Advances in emotion recognition: Link to depressive disorder, *Neurol. Ment. Disord.* (2020).
- [9] N. Sinha, Affective computing and emotion-sensing technology for emotion recognition in mood disorders, in: *Enhanced Telemedicine and e-Health: Advanced IoT Enabled Soft Computing Framework*, 2021, pp. 337-360.
- [10] A. Kumar, K. Sharma, A. Sharma, MEMoR: A multimodal emotion recognition using affective biomarkers for smart prediction of emotional health for people analytics in smart industries, *Image Vis. Comput.* 123 (2022) 104483.