What Can Facial Movements Reveal? Depression Recognition and Analysis Based on Optical Flow Using Bayesian Networks

Yu Ma, Jian Shen, Zeguang Zhao, Huajian Liang, Yang Tan, Zhenyu Liu, *Member, IEEE*, Kun Qian, *Senior Member, IEEE*, Minqiang Yang, *Member, IEEE*, and Bin Hu, *Fellow, IEEE*

Abstract— **Recent evidence have demonstrated that facial expressions could be a valid and important aspect for depression recognition. Although various works have been achieved in automatic depression recognition, it is a challenge to explore the inherent nuances of facial expressions that might reveal the underlying differences between depressed patients and healthy subjects under different stimuli. There is a lack of an undisturbed system that monitors depressive patients' mental states in various free-living scenarios, so this paper steps towards building a classification model where data collection, feature extraction, depression recognition and facial actions analysis are conducted to infer the differences of facial movements between depressive patients and healthy subjects. In this study, we firstly present a plan of dividing facial regions of interest to extract optical flow features of facial expressions for depression recognition. We then propose facial movements coefficients utilising discrete wavelet transformation. Specifically, Bayesian Networks equipped with construction of Pearson Correlation Coefficients based on discrete wavelet transformation is learnt, which allows for analysing movements of different facial regions. We evaluate our method on a clinically validated dataset of 30 depressed patients and 30 healthy control subjects, and experiments results obtained the accuracy and recall of 81.7%, 96.7%, respectively, outperforming other features for comparison. Most importantly, the Bayesian Networks we built on the coefficients under different stimuli may reveal some facial action patterns of depressed subjects, which have a potential to assist the automatic diagnosis of depression.**

Index Terms— **Depression recognition, facial expressions, action units, optical flow, Bayesian networks.**

I. Introduction

MENTAL health holds the great importance in daily life and activities. An estimated 280 million people worldwide suffer from depression, which may even lead to suicide [1]. As a kind of complicated mood disorder, Major Depressive Disorder (MDD) may vary in severity and last for even years, resulting in severe social and psychological distress which may impair patient's ability to function in daily and social life [2], [3], [4].

Early detection of MDD is of great importance to patients, which has a potential to provide them with proper antidepressant medication and psychological counseling in case the symptoms of depression get worse. There are some common diagnostic styles including self-reporting questionnaire [5], [6] and interview assessment methods [7]. Depression diagnosis based on self-assessment scale is widely used in various scenarios, such as fundamental health care, but its disadvantages can not be ignored [8], [9]. The results of scales can't reflect the meaning of the feedback in perspective of clinical concept. Since the content of the scale is fixed, it is impossible to properly consider some individual characteristics and distinguish other psychiatric and medical comorbidities from depression. What's worse, clinical diagnosis conducted by experts such as psychologist and psychiatrist usually consist of comprehensive assessment about patient's life circumstances, symptoms, and etc, which highly relies on expert's subjective experiences, risking a range of subjective biases [10]. As the amount of patients grows, this experience-dependent diagnosis procedure needs intensive

labor of experts and the follow-up treatment can be harder by a lack of resources and trained health care providers. To this end, it is of great importance to develop models diagnosing depression combing machine learning and the traditional evaluation methods to improve the diagnosis effectiveness of depression [11], [12], [13]. The idea inspires us to explore strategies of the depression recognition based on objective indicators, which are highly needed to assist clinical doctors in objective and efficient depression diagnosis [14], [15].

DSM-V [16] depicts that the diagnosis of mental disorders should help clinical doctors to make the patient's treatment plan and potential development of mental state. In order to make up for the inadequacy of traditional depression diagnosis methods such as the high reliance on doctors' experience and risk of subjective bias, an objective and efficient method is urgently needed to assist doctors in clinical diagnosis of depression. As the application in artificial intelligence promotes [17], [18], [19], depression assessment based on machine system is of help to ensure that depressed patients can get timely and effective clinical treatment, improving the well-being of patients' lives [20].

Within the past several years, depression recognition based on facial expressions has been paid increasing attention [21], which has a potential to enable an effective and objective diagnosis approach for people suffering depression, so that the valid treatment can be applied to patients in time. DSM-V [16] demonstrates that patients' facial expressions and actions could be an evidence to infer the existence of the depression. As a significant method for expressing emotion, facial expression has been explored by psychologists with various modeling strategies, such as the "Facial Action Coding System" (FACS) [22] with the famous idea that each facial expression can be visually represented as a combination of different facial muscles action. The process of activating each Action Units (AUs) largely take place in a visible way due to changes in facial geometry and appearance. The experiment results of research [23] exploring whether facial movements can identify depression concluded that the model can achieve similar depression detection accuracy utilising FACS parameter or active appearance model, which suggests that observable differences are existed in facial expressions in depressed patients and normal person.

Despite that significant researches have been conducted in depression detection based on facial expressions, due to the complexity of depression symptoms, this paper makes some contributions according to the specific challenges as follows:

- A facial expression dataset was collected, comprising MDD patients and healthy subjects, to capture facial differences and optimize depression recognition.
- Optical Flow (OP) and wavelet transformation coefficients were extracted to reveal the subtle nuances of facial movements.
- Bayesian Networks (BN) were utilized to analyze the inherent correlation between specific facial regions, which had potentials to show underlying facial movements patterns in depressed patients.

The remainder of the paper is organised as follows. Section II briefly introduces the related work, and Section III depicts the data set and our proposed models for depression recognition. Sections IV presents our experimental settings, results and analysis. Finally, we conclude the paper in Section V.

II. BACKGROUND AND RELATED WORK

A. Depression Recognition Based on Facial Expressions

For the past few years, depression recognition [24], [25], [26] has been paid much attention. One of them is automated depression recognition based on facial expressions. Tremendous algorithms have been proposed to conduct novel and efficient face detector [27], [28], [29]. For instance, early face detection with novel hand-craft features such as the Viola and Jones [30] and Dlib library [31], could locate significant facial regions or landmarks. Moreover, some neural networks have been incorporated with algorithms of face detection, generating a series of promising face detectors [32], [33], [34].

Feature extraction demonstrates the process of information representation, aiming to obtain a robust descriptor and reduce the data redundancy [35]. Sariyanidi et al. [36] summarized some facial features such as the Local Binary Pattern (LBP) [37], which excelled in the image texture and highlighted certain facial areas including the mouth, eyebrows and eyes [38]. OP [39], [40] extracts and evaluates targeted object's magnitude and direction on the pixel level, which could effectively measure spatio-temporal intensity changes of facial landmarks or extracted ROIs. Moreover, statistical analyses such as Discrete Wavelet Transformation (DWT) could achieve the location of key frames without time consumption and inevitable redundancy comparing with continuous wavelet transformation [41]. DWT has been extensively applied to feature extraction from one-dimensional signals. Previous studies have explored its classical applications in depression recognition [42], [43] and facial analysis [44], [45]. The approximation and detail coefficients derived from the discrete wavelet transform can serve as effective features for capturing the information of signals at different scales and frequencies. These feature vectors can be utilized as inputs to machine learning algorithms for classification and recognition tasks. Specifically, in the domain of facial image analysis, the approximation and detail coefficients obtained from the discrete wavelet transform have been widely employed for tasks such as facial expression and action recognition. These coefficients capture both the global trends and local details of the images, which are crucial for extracting specific features from facial images.

B. Bayesian Networks Construction

BNs are probabilistic graphical models, which are directed acyclic graphs and could represent the joint probability distribution of variables in an effective form. The network consists of nodes, which represent the variables of interest, and edges, which represent the conditional dependencies between the nodes. The advantage of BNs is that they provide a compact and intuitive representation of complex probabilistic relationships between variables. They can also be leveraged to perform probabilistic inference, which allows

for the calculation of the probability distribution of a subset of variables given evidence about other variables. BNs have been widely applied in scenarios of uncertainty reasoning [46], [47], with advantages of graphical representation and probability [48]. But the structure learning of BNs is NP-hard with its challenging learning process from the data set [49]. The effectiveness of model depends on the BNs structure to a large extent [50]. On the premise of utilising the whole data, in order to find the global score under the optimal structure, a lot of researches leverage the dynamic planning method to find the global optimal structure avoiding the search space [51]. The disadvantages of above methods are high time complexity and big memory consumption, which makes them unsuitable for variables in large amount. Perrier et al. [52] proposed an optimal undirected graph with structure constraint to reduce the search space. Kojima et al. [53] promote the idea of the method utilising variables clustering and ancestors constraints. Structure learning based on constraints is an effective strategy to reduce the search space [54] of BNs. Certain strategies should be performed to measure every two variables and determine whether to add an edge to them by certain measuring methods. Li et al. [55] used constraint-based BNs learning strategy in the process of AUs correlation exploration. Moreover, researchers have explored compound emotion categories by measuring AUs involved in facial expressions [56].

C. Action Units and Facial Expressions

Facial expressions recognition has attracted increasing attention recently [57]. The FACS could utilise the combination of facial action units to describe a host of facial expressions. In previous [58] studies, the spatio-temporal correlation and interaction between facial movements are rarely explored, and those correlations could help us better understand and analyze facial expressions [55]. The FACS divides facial area into 46 AUs, each of which is anatomically related to the contractile movements of a specific set of facial muscles. In general, various AUs representing different facial muscle movements are usually combined to form natural facial expressions to reveal emotions [59]. As depicted in the FACS [60], the intrinsic relationship between AUs can provide the information to better analyze facial expressions.

III. Materials and Methods

A. Participants and Data Acquisition

1) Experimental Paradigm: To explore the facial movements of depressed patients and obtain nuances of expressions, three kinds of emotional stimuli are required. We recruit participants and collect data utilising the video stimuli from the "Chinese Affective Video System" including positive, neutral, and negative movie clips, and each video clip is about 2 minutes. The whole process can be seen as Fig. 1. At the beginning of the experiment, the subject is instructed to prepare to watch the videos. At the end of each video, subject takes a break of about 2 minutes to adjust his/her emotional state, which could eliminate the emotional experience produced by the previous video. Each participant was required to watch all three video clips for once, and their facial expressions
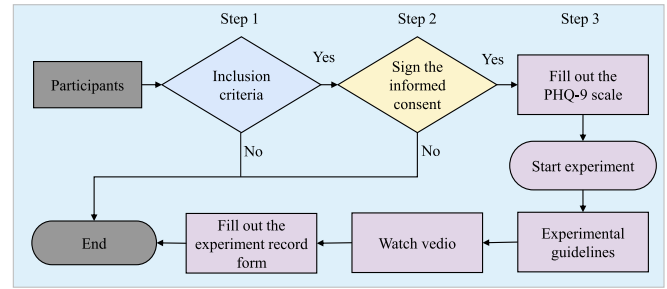

Fig. 1. The flow chart of recruiting participants and collecting data.

were recorded. The experiment is conducted in a quiet and clean room where the subjects' emotions are not affected by the external environment. Experiments are completed in a comfortable chair during the process of watching movie clips and a *Logitech CC2000e* camera is utilised to record videos of the subjects' facial expressions.

Algorithm 1 Depression Recognition Based on Facial Movement Coefficients

---

1: **Input**: Given a video dataset $D = \{w_1, w_2, \ldots, w_n\}$ of depression, where w_i is i th video sequence in dataset
2: **Process**: Feature extraction and depression recognition
3: **I. Data preprocessing**
4: **for** each $w_i, i \in \mathbf{n}$ **do**
5:     Crop a video into a sequence of images $\lambda_i, i \in \mathbf{n}$
6:     Extract facial ROIs
7: **end for**
8: **II. Feature extraction**
9: **for** images in sequence $\lambda_i, i \in \mathbf{n}$ **do**
10:     Compute OP feature for each ROIs ψ_i
11:     Analyse image frames statistically with sliding windows
12:     Compute the DWT approximate and detail coefficients, C_a and C_d based on the data extracted from sliding windows
13:     Set C_a and C_d as facial movements coefficients for the extracted OP
14: **end for**
15: **III. Depression recognition model**
16: Parameters Optimisation
17: Recognise depression utilising the OP feature based on the facial movements coefficients
18: **IV. Facial movements analysis based on the BN**
19: **for** each ROIs $R_i, i \in \mathbf{14}$ **do**
20:     Analysis Pearson correlation on wavelet coefficients
21: **end for**
22: Build BN based on wavelet coefficients
23: **Output**: Recognition results and BN based on ROIs under positive, neutral and negative stimuli

---

2) Inclusion and Exclusion Criteria: Written informed consent was obtained from all subjects prior to the experiment. The local ethics committee approved the consent form and the study for biomedical research at the Second People's Hospital of Gansu Province in accordance with the World Medical Association (WMA) International Code of Medical Ethics. All patients with MDD underwent a structured neuropsychiatric
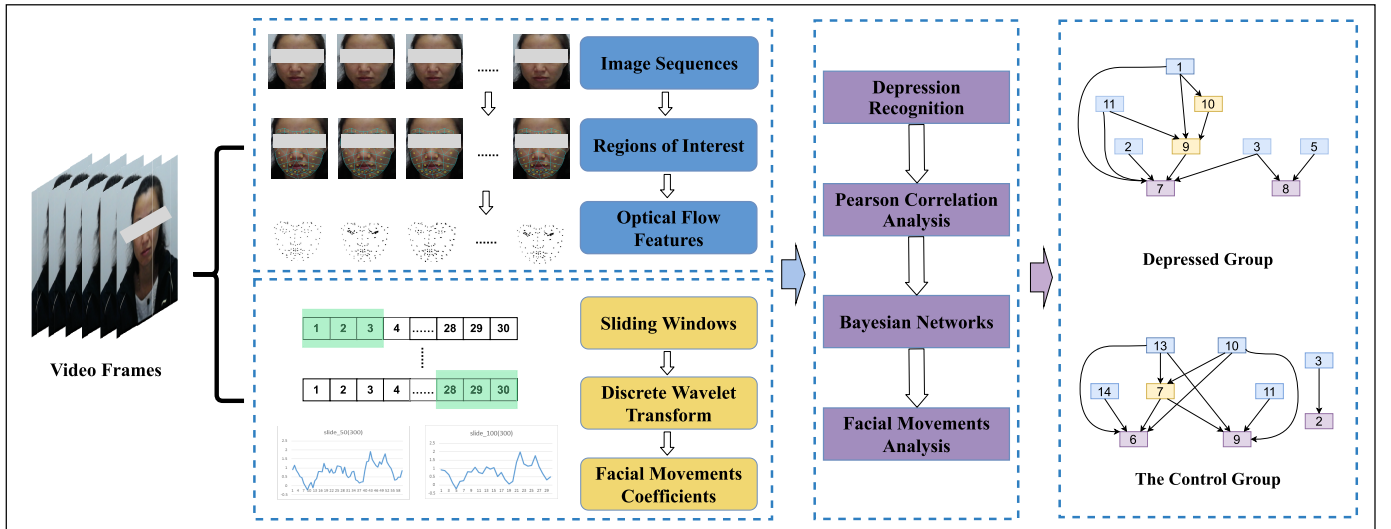
Fig. 2. Depression recognition based on facial expressions.

TABLE I
FACIAL REGIONS OF INTEREST EXTRACTION BASED ON ACTION UNITS

ROIs No	Related AU	Facial Movements	Facial Muscles
1	AU1, AU4	Inner Brow Raiser, Brow Lower	Frontalis, Pars Medialis, Corrugator and Depressor Supercilii
2	AU2	Outer Brow Raiser	Frontalis, Pars Lateralis
3	AU5	Upper Lid Raiser	Levator Palpebrae Superioris
4	AU6	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
5	AU7	Lid Tightener	Orbicularis Oculi, Pars Palpebralis
6	AU9	Nose Wrinkler	Levator Labii, Superioris Alaquae Nasi
7	AU10	Upper Lip Raiser	Levator Labii Superioris
8	AU11	Nasolabial Deepener	Zygomaticus Minor
9	AU12	Lip Corner Puller	Zygomaticus Major
10	AU13, AU14	Cheek Puffer, Dimler	Levator Anguli Oris, Buccinator
11	AU15	Lip Corner Depressor	Depressor Anguli Oris
12	AU16, AU20	Lower Lip Depressor, Lip Stretcher	Depressor Labii Inferioris, Risorius with Platysma
13	AU17	Chin Raiser	Mentalis
14	AU18, AU23, AU24	Lip Puckerer, Lip Tightener, Lip Pressor	Incisivii Labii Superioris and Inferioris, Orbicularis Oris

interview that met the diagnostic criteria for MDD as described in DSM-IV. For all subjects, inclusion criteria are as follows: a) no psychotropic drug treatment could have been taken within the two weeks prior to diagnosis and b) no drug treatment could have been started after diagnosis. The exclusion criteria are: a) chronic somatic-like illness, b) long-term use of cortisol drugs, sedatives, analgesics, drugs for hypertension and heart disease, sleep medications, antiepileptic drugs, etc., c) presence of patients with mental disorders within two generations and three lines, d) neurological problems, alcohol or psychotropic substance abuse or dependence in the past year, and e) pregnant women and those who were breastfeeding or taking birth control pills.

In accordance with the above inclusion and exclusion criteria, this paper collected data based on 34 depression outpatients diagnosed by clinical psychiatrists at the Second People's Hospital of Gansu Province and 32 healthy controls, totaling 66. Excluding some invalid data, such as a part of the face disappearing for a long time during the acquisition process and frequent body movements leading to the face incompleteness, we had 30 outpatients (10 males, 20 females) and 30 healthy controls (16 males, 14 females).

3) Proprecessing: In order to extract more effective OP features, each video is cropped into a sequence of images, and we also delete the video sequences with incomplete faces. For each frame, we utilise the face detector of the Dlib to detect the face in the image, and then leverage 68 face landmarks to calibrate each part of the face to achieve the face crop. The whole process of the method can be seen in Fig. 2.

B. ROIs Extraction

We found that an AU-based approach had a potential for the extraction of facial ROIs. Combining the association between AU and facial expressions, we propose a specific scheme for extracting facial ROIs according to the facial regions involved in AU [61], as is shown in Table I. We also visualised the magnitude of OP in 14 ROIs about the proposed partition scheme. The preliminary facial area extraction scheme is shown in Fig. 3, which represent the box plots of normalized OP magnitude with three different stimuli as the facial regions. It can be seen that the activated areas of different facial regions showed a significant difference which indicated certain movements of different regions.
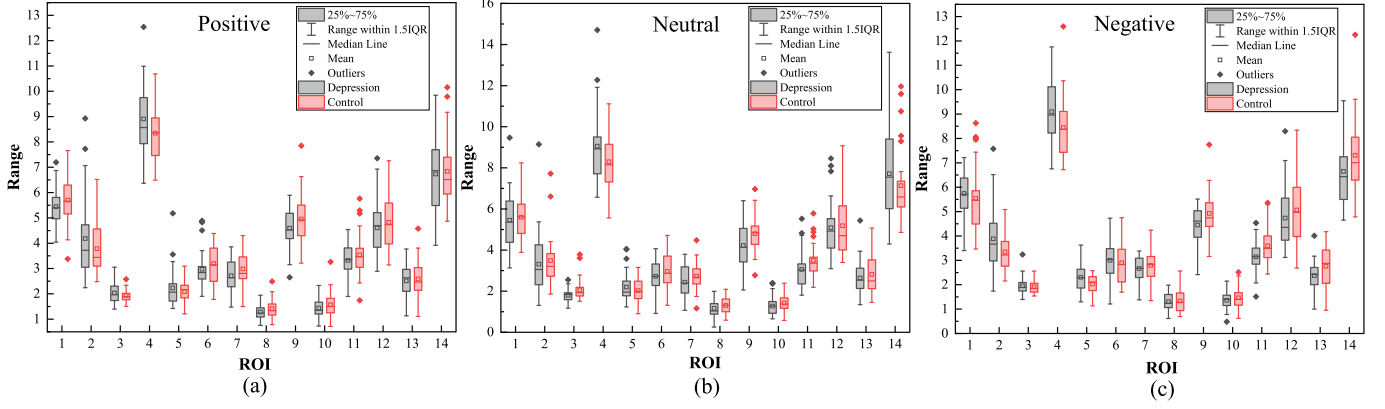
Fig. 3. Magnitude of OP under (a) positive, (b) neutral and (c) negative stimuli in 14 ROIs.

C. Feature Extraction

1) ROIs-Based Optical Flow Feature: OP [39], [62] aims to evaluate the magnitude and direction of motion based on a video frame or image sequences leveraging a series of facial landmarks. Generally, for an object, OP based algorithms compute the pixel-level movements occurring between two frames. The method of OP has a dependence on the selection of the initial tracking points and their position changes, which is highly sensitive to some factors such as noise and occlusion [63].

For an image sequence, the OP infers the motion of a target object by detecting the changes of pixels between two frames over time. For each image, we used these 68 facial landmarks to compute our 14 ROIs referring to action units. Given a facial landmark $L_i, i = 1, 2, \ldots, 68$ derived from the detected face in the image, it has coordinates $L_i(x, y)$ with a pixel intensity as $I_{L_i}(x, y, t)$ at time t, and it moves a certain distance of $\Delta x, \Delta y$ between two frames in the time of Δt. According to the brightness constancy constraint of the OP, the OP of this landmarks can be described as the follows:

$$I_{L_i}(x, y, t) = I_{L_i}(x + \Delta x, y + \Delta y, t + \Delta t), \quad (1)$$

Assuming that the distance of the pixel is small, the OP constraint equation for the pixel intensity $I_{L_i}(x, y, t)$ at i_th point $L_i(x, y)$ can be developed according to the Taylor series as:

$$I_{L_i}(x + \Delta x, y + \Delta y, t + \Delta t)$$
$$= I_{L_i}(x, y, t) + \frac{\partial I_{L_i}}{\partial x}\Delta x + \frac{\partial I_{L_i}}{\partial y}\Delta y + \frac{\partial I_{L_i}}{\partial t}\Delta t + \tau, \quad (2)$$

where τ is a higher order infinitesimal, and from these equations, we can obtain:

$$\frac{\partial I_{L_i}}{\partial x}\Delta x + \frac{\partial I_{L_i}}{\partial y}\Delta y + \frac{\partial I_{L_i}}{\partial t}\Delta t = 0, \quad (3)$$

and:

$$\frac{\partial I_{L_i}}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I_{L_i}}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I_{L_i}}{\partial t}\frac{\Delta t}{\Delta t} = 0, \quad (4)$$

We can concluded that:

$$\frac{\partial I_{L_i}}{\partial x}V_{x_{L_i}} + \frac{\partial I_{L_i}}{\partial y}V_{y_{L_i}} + \frac{\partial I_{L_i}}{\partial t} = 0, \quad (5)$$

where $V_{x_{L_i}}, V_{y_{L_i}}$ are the velocities corresponding to the pixel intensities $I_{L_i}(x, y, t)$ and the x and y components of the OP, respectively. Thus, between two frames at a distance Δt, the OP of a facial landmark $L_i(x, y)$ at t is represented as a two-dimensional vector as follows:

$$\left[V_{x_{L_i}}^t, V_{y_{L_i}}^t\right]^T, \quad i = 1, 2, \ldots, 68, \quad (6)$$

Then, the Euclidean coordinates can be converted into polar coordinates represented by

$$\left[\rho_{L_i}^t, \theta_{L_i}^t\right]^T, \quad i = 1, 2, \ldots, 68, \quad (7)$$

where $\rho_{L_i}^t$ and $\theta_{L_i}^t$ are the magnitude and direction of OP in i_{th} landmarks at time t, respectively. Given an image sequence $f_1, f_2, \ldots, f_m$, for each frame $f_j, j = 1, 2, \ldots, m$, compute the OP of each ROI $R_k^j, k = 1, 2, \ldots, 14$, where m is the total number of frame and k is the number of ROIs. The normalized OP feature based on ROIs for a sequence is represented by:

$$\Psi = \frac{1}{m}\sum_{j=1}^{m}\left[\left(\rho_{R_1^j}, \theta_{R_1^j}\right)^T, \left(\rho_{R_2^j}, \theta_{R_2^j}\right)^T, \ldots, \left(\rho_{R_{14}^j}, \theta_{R_{14}^j}\right)^T\right],$$
$$(8)$$

2) Facial Movements Coefficients Based on Discrete Wavelet Transformation: Wavelet analysis provides a highly flexible strategy to extract features based on the specified time or frequency [64]. We use the approximation coefficients and detail coefficients in DWT as those coefficients, respectively. We first extract optical flow from sliding windows as input signals and leverage DWT to obtain their time-frequency representations. This approach aims to capture the temporal dynamics and variations of optical flow features at different time scales. We then extract the approximation and detail coefficients as facial movements coefficients. The approximation coefficients capture the global trends of the optical flow features, while the detail coefficients capture their local variations. The DWT-based facial movements coefficient offers several advantages. Firstly, it provides a multi-scale representation, enabling us to analyze the variations of optical flow features across different frequency ranges. This helps us capture subtle dynamic patterns and extract features relevant to individual subjects or emotional states. Secondly, the approximation and

detail coefficients provide descriptions of both the overall trends and local details of the optical flow features. This is crucial for understanding the global dynamics and local changes within the optical flow features. The research findings on the application of discrete wavelet transform in one-dimensional signal feature extraction and image analysis provide important insights for our study. By applying DWT to analyze OP, we could help us to better construct the nuances in facial movements exhibited between depressed patients and controls under the experimental stimuli across multiple time scales and frequencies, moreover, to uncover their potential implications for individual-specific and invariant feature representation.

For each image sequence $f_1, f_2, \ldots, f_m$, we utilise an sliding window with size of window as n, and step as $step$, the sub-sequence $\gamma_{[a,b]}$ obtained by sliding window is Γ, and sliding window can be represented as:

$$\Gamma = \gamma_{[1+step \times \eta, n+b \times \eta]}, \tag{9}$$

where a, b is the start and end frame of a sub-sequence, and η can be computed as:

$$n + step \times \eta = m, \tag{10}$$

For a sequence Γ from sliding windows, we compute the average optical flow feature, and then we conduct discrete wavelet transformation based on the feature and obtain the approximation coefficients and detail coefficients, which are denoted as C_a, C_d, respectively. The $pywt$ package from the Python library was employed for DWT calculations using $db3$, and The mean values of the approximation and detail coefficients were separately calculated, resulting in the final facial movement coefficients. We combine these coefficients C with OP as our final feature:

$$\zeta = \left[C_{R_\xi} \times \left(\rho_{R_\xi}, \theta_{R_\xi} \right)^T \right], \quad \xi = 1, 2, \ldots, 14. \tag{11}$$

The two facial movements coefficients extracted in this paper demonstrate the motion characteristics of ROIs. Specifically, the approximate coefficients derived from the sliding window data reflect the process of changes in the subjects' facial OP throughout the experiment, which have significant ability to represent the overall changes. Moreover, the detailed coefficients could reveal the more acute parts of the changes depicting different responses of facial expressions generated by these two groups under the experimental stimuli. The whole algorithm could be seen in Algorithm 1.

3) Pearson Correlation Coefficients: The Pearson Correlation Coefficient measures the correlations of two or more variables. Studies exploring the correlation between different expressions and action units have received much attention in recent years. For example, Wang et al. [65] investigated the analysis of the correlation between six basic expressions and AU based on the Extended Cohn-Kanade (CK+) dataset [66]. Referring to those researches, in our paper, we build and analyse the correlation between our ROIs and various AUs, which could help us understand the nuances generated by different emotions from the perspective of facial muscles by utilising Pearson Correlation Analysis.

TABLE II
PERFORMANCE RESULTS UNDER POSITIVE, NEUTRAL AND NEGATIVE STIMULI. COEF.: FACIAL MOVEMENTS COEFFICIENTS. C_a: APPROXIMATE COEFFICIENT. C_d: DETAIL COEFFICIENT. DT: DECISION TREE. kNN: k NEAREST NEIGHBOURS. SVM: SUPPORT VECTOR MACHINE. NAIVEB: NAIVE BAYES

Coef.	Stimuli	Evaluation	Classifier			
			DT	kNN	SVM	NaiveB
C_a	Positive	Accuracy	**0.82(0.39)**	0.65(0.48)	0.72(0.45)	0.50(0.50)
		Precision	**0.74(0.50)**	0.66(0.47)	0.67(0.50)	0.50(0.49)
		Recall	**0.97(0.50)**	0.63(0.47)	0.87(0.50)	0.80(0.49)
		F1 Score	**0.84(0.50)**	0.64(0.47)	0.75(0.50)	0.62(0.49)
	Neutral	Accuracy	**0.80(0.40)**	0.62(0.49)	0.65(0.48)	0.63(0.48)
		Precision	**0.72(0.50)**	0.60(0.48)	0.68(0.45)	0.59(0.50)
		Recall	**0.97(0.50)**	0.70(0.48)	0.57(0.45)	0.87(0.50)
		F1 Score	**0.83(0.50)**	0.65(0.48)	0.62(0.45)	0.70(0.50)
	Negative	Accuracy	**0.75(0.43)**	0.55(0.50)	0.68(0.47)	0.55(0.50)
		Precision	**0.76(0.48)**	0.53(0.49)	0.69(0.47)	0.53(0.50)
		Recall	**0.73(0.48)**	0.77(0.49)	0.67(0.47)	0.87(0.50)
		F1 Score	**0.75(0.48)**	0.63(0.49)	0.68(0.47)	0.66(0.50)
C_d	Positive	Accuracy	**0.73(0.44)**	0.65(0.48)	0.67(0.47)	0.60(0.49)
		Precision	**0.69(0.49)**	0.62(0.49)	0.68(0.47)	0.64(0.42)
		Recall	**0.83(0.49)**	0.80(0.49)	0.63(0.47)	0.47(0.42)
		F1 Score	**0.76(0.49)**	0.70(0.49)	0.66(0.47)	0.54(0.42)
	Neutral	Accuracy	**0.68(0.47)**	0.52(0.50)	0.57(0.50)	0.57(0.50)
		Precision	**0.66(0.49)**	0.52(0.44)	0.56(0.46)	0.54(0.49)
		Recall	**0.77(0.49)**	0.53(0.44)	0.60(0.46)	0.83(0.49)
		F1 Score	**0.71(0.49)**	0.52(0.44)	0.58(0.46)	0.66(0.49)
	Negative	Accuracy	**0.60(0.49)**	0.60(0.49)	0.63(0.48)	0.57(0.50)
		Precision	**0.58(0.48)**	0.62(0.44)	0.64(0.46)	0.54(0.50)
		Recall	**0.70(0.48)**	0.53(0.44)	0.60(0.46)	0.90(0.50)
		F1 Score	**0.64(0.48)**	0.57(0.44)	0.62(0.46)	0.68(0.50)

Firstly, we used the Wilcoxon signed ranks test to calculate z = -2.59 (the critical value (CV) of z is -1.96). Therefore, the C_a was significantly better than the C_d. Then, we used the Friedman test to calculate the $F_{3,15} = 14.35$ (the CV is 3.29), which satisfied $p<0.05$ and CD=1.91. Thus, according to the post-hoc two-tailed Nemenyi test, the DT performed significantly better than the kNN and NaiveB.

IV. RESULT

A. Classification

Several classifiers including k Nearest Neighbors (kNN), Support Vector Machines (SVM), Naive Bayes (NaiveB), and Decision Trees (DT). Moreover, depression recognition was performed under different kinds of the stimuli for comparison, and all these classifiers underwent a grid parameter tuning step. For instance, in the decision tree, we leveraged the grid search to optimise the parameter selection including max_depth, $min\_samples\_split$, and $min\_samples\_leaf$. In SVM, we used the $sigmoid$ kernel and performed grid tuning for the hyper-parameters. We conducted leave-one-subject-out cross validation, in each fold, we leveraged 59 subjects to train model and the left one to test. After 60 folds, we obtained confusion matrix. The standard deviation was calculated by results of each fold. The results with standard deviation are shown in Table II and the receiver operating characteristic (ROC) curves of optimized models are shown in Fig. 6, where C_a and C_d depict facial movements coefficients.

Additionally, we selected some facial features for comparison. For LBP [67], we extracted the LBP features of each frame image and then used the features as our classification features. For Histogram of Oriented Gradient (HOG) [68], we divided the images into blocks with cells. We computed the HOG features for each frame image and took the average values across all cells within the block. The Main Directional Mean Optical Flow (MDMO) [39] and the Transitional
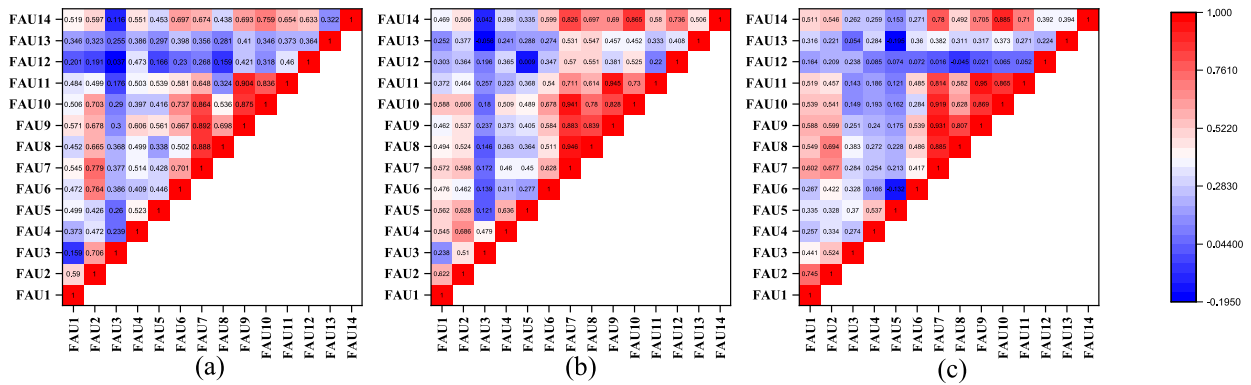
Fig. 4. Pearson Correlation Coefficients of DWT Coefficients in depressed patients under three kinds of stimuli. (a): positive stimuli, (b) neutral stimuli, (c) negative stimuli.
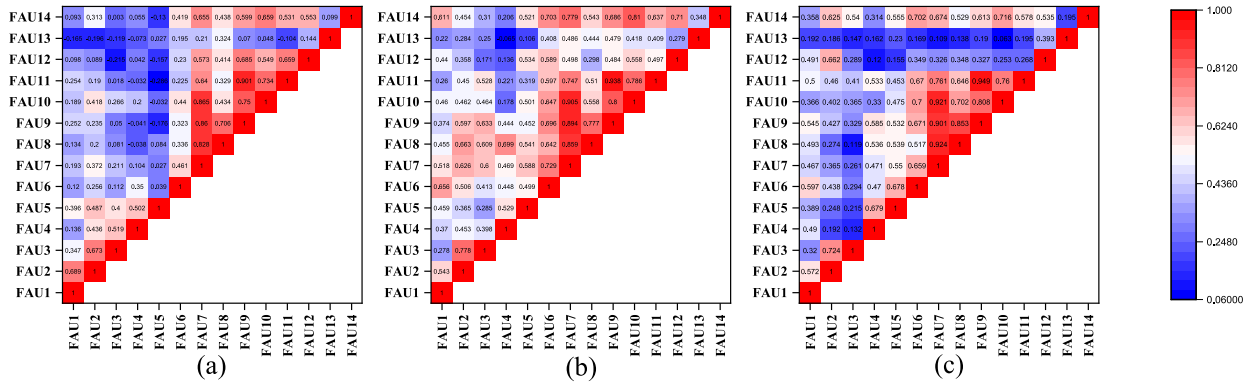

Fig. 5. Pearson Correlation Coefficients of DWT Coefficients in healthy control group under three kinds of stimuli. (a): positive stimuli, (b) neutral stimuli, (c) negative stimuli.

TABLE III
PERFORMANCE RESULTS OF OUR PROPOSED
FEATURE VS OTHER FEATURES

Feature	Accuracy	Precision	Recall	F1 Score
MDMO [39]	0.65(0.48)	0.66(0.47)	0.63(0.47)	0.64(0.47)
LBP [67]	0.62(0.48)	0.65(0.43)	0.50(0.43)	0.57(0.43)
HOG [68]	0.65(0.48)	0.65(0.47)	0.67(0.47)	0.66(0.47)
TOFS [58]	0.78(0.41)	0.72(0.50)	0.93(0.50)	0.81(0.50)
Ours	**0.82(0.39)**	**0.74(0.50)**	**0.97(0.50)**	**0.84(0.50)**

According to the Friedman test, the $F_{4,20} = 17.09$ (the CV is 2.87), which satisfied $p < 0.05$ and CD=2.49. Then, the Nemenyi test revealed our features performed significantly better than MDMO, LBP and HOG.

Fig. 6. ROC of optimized models based on facial movements coefficients under positive, neutral and negative stimuli.

Optical Flow under stimuli (TOFS) [58] are two optical flow features, which are also compared, and results with standard deviation are shown in Table III. From above results, we can see the depression recognition based on multiple classifiers achieved a high recall, which is consistent with our previous experiment [58]. The high recall of some classifiers in our experiment indicated that our proposed feature performed better for the recognition of depressed patients, while the diversity of facial activities of healthy subjects might lead classifiers misjudged. Moreover, about classifiers in Table II, we found that the DT performed significantly ($p < 0.05$) better than the kNN and NaiveB. About features in Table III, our feature performed significantly ($p < 0.05$) better than MDMO, LBP and HOG.

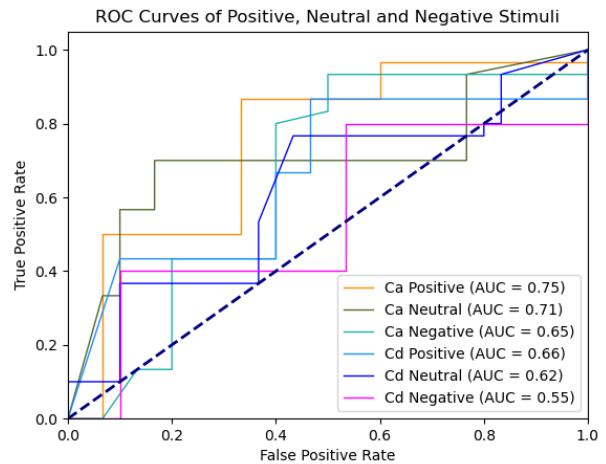We selected the detail coefficients to perform the Pearson correlation analysis of 14 ROIs (as shown in Fig. I) in the depressed and control group under the different stimuli, and the results are shown in Figs. 4 and 5. As can be seen from those figures, under positive stimuli, some ROIs have high correlations in the depressed group such as ROI2, ROI3, ROI7 and ROI9. Among those ROIs, ROI2 involves facial movements of external brow elevation and ROI3 reveals upper
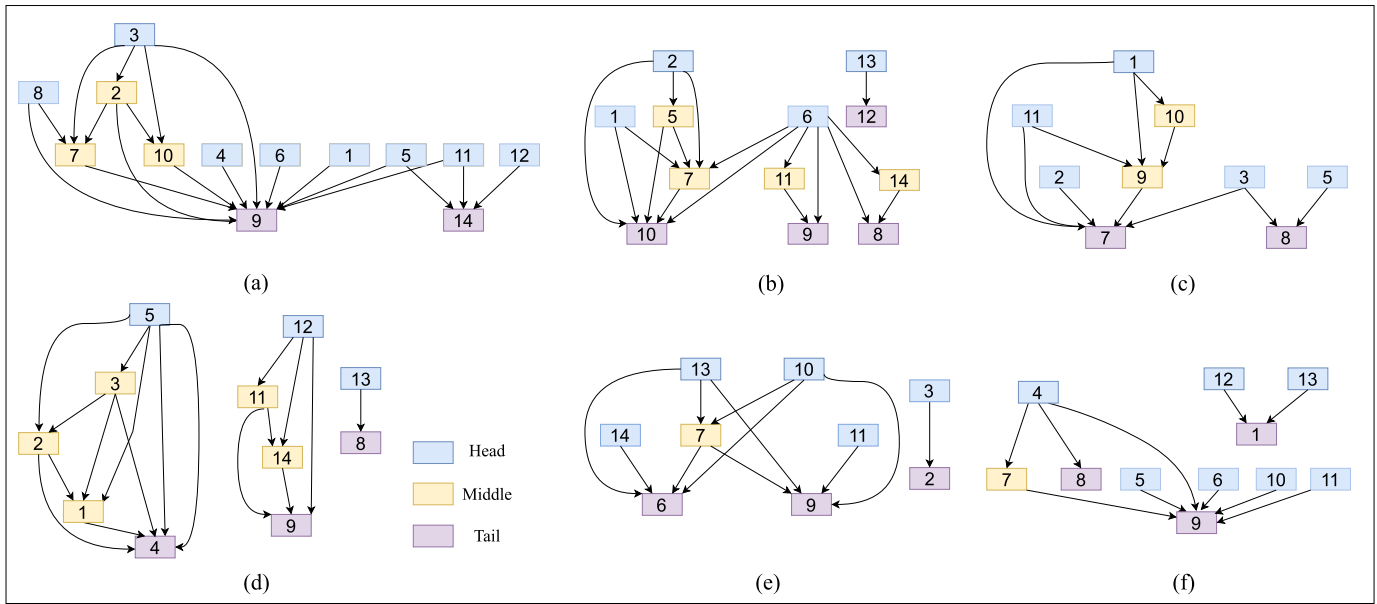
Fig. 7. Bayesian Networks of depressed group and the control group under different stimuli. (a), (b) and (c) are BN of depressed group under positive, neutral and negative stimuli, respectively. (d), (e) and (f) are BN of the control group under positive, neutral and negative stimuli, respectively. As can be seen from BNs in two groups, BNs in depressed group show a strong correlation of facial movements under negative stimulus with ROI7 (lip muscle movement) and ROI9. Moreover, the lip movements of depressed subjects under all three stimuli reveal higher correlations with other facial movements, while the ROIs in the control group show a multiple connection.

eyelid lift. We speculate that the higher correlations of these ROIs reveal the external brow elevation under positive stimuli. Under neutral stimuli of depressed group, ROI6 shows a high correlation with other ROIs, which reveals a facial movement of nasal wrinkling in the depressed subjects. Moreover, it can be seen that ROI7 correlates with most of the ROIs higher in the depressed and control groups than in the depressed group. Under negative stimuli, depressed subjects generally had higher correlation coefficients for ROI7 and other ROIs, but controls did not have this relationship. In order to analyse those patterns in depth, we built BN and showed the correlations of those various ROIs under different stimuli.

B. Bayesian Networks

In addition to the Pearson correlation coefficients for analyzing the differences in facial activity characteristics between the two groups of subjects, in order to further analyze the correlations between these ROIs and thus explore the facial movements characteristics of the depressed and control groups under three different paradigm stimuli, we constructed BN based on the facial movements coefficients. To learn the structure of BN, we leveraged the Pearson correlation coefficients as a network structural constraint. For each group of data, we take its second quartile as a threshold, and if the value of Pearson correlation coefficient of a two ROIs is smaller than this threshold, we set them as independent variables. In order to learn the correlation between ROIs, we do not add edges manually because determining the direction of edges requires significant domain knowledge about facial expressions, which is not necessarily applicable in depressed patients, so we put some constraints on the independence between ROIs based on Pearson correlation coefficients. The final constructed BN

under different stimuli are shown in Fig. 7. Furthermore, it should be noticed that there are various learning algorithms for structures of BN, and the optimal network structure may deviate based on the different structure learning algorithms. In this paper, we construct the network from the detail coefficients under different paradigm stimuli based on the independence constraint and use those networks to analyse the characteristics of facial movements between those two groups.

C. Analysis and Evaluation

Based on the BN, we explored the facial movements of depressed and control subjects under different stimuli. Fig. 7 shows that the depressed group exhibited a series of movements associated with ROI9, corresponding to the facial action of lip pulling tight (AU12). Additionally, ROI7 and ROI14 showed high correlation with other ROIs, both corresponding to lip movements. Under neutral stimuli, the network structure of depressed subjects showed more multiple connections compared to positive stimuli, indicating richer interconnections among ROIs. Depressed subjects also showed a close relationship between ROI10 and ROI7, which correspond to muscle movements in the labial horn muscle decompression and upper lip elevation, respectively. The BN of the control group under negative stimuli showed a multiplicity of connectivity similar to that of ROI9 exhibited by depressed subjects under positive stimuli, and related studies have shown that the lip horn pulling and tightening corresponding to ROI9 occurs in compound classes of emotions, such as the expression [56] of happy and disgust. Furthermore, the depressed group showed a strong correlation of facial movements under negative stimulus with ROI7 (lip muscle movement) and ROI9. To conclude, lip movements of depressed subjects under all three stimuli

revealed higher correlations with other facial movements, while the ROIs in the control group showed a multiple connection.

In some of our cases, Bayesian Networks exhibit independent subnetworks, which might be attributed to several reasons such as the lack of correlation among variables, subdomain specificity, and limitations of the learning algorithm. Furthermore, factors such as dataset characteristics, sample size, and algorithm constraints could contribute to the formation of separate subnetworks. To address these issues, future experiments will involve increasing participant numbers and optimizing the learning algorithm for a more comprehensive BN structure and lead us to an improved understanding of the dependencies among variables.

V. CONCLUSION

Facial expression data contain rich emotional information, and increasing studies have been conducted in recent years to combine facial expressions and depression recognition to utilise some information revealed by patients' faces to help psychiatrists diagnose depression. In this study, we extracted sparse OP features and conducted a series of statistical analysis to explore the time-frequency characteristics and recognise depression. The experimental results showed the validity of our model. Furthermore, we built BN based on wavelet transformation coefficients to analyse facial movements. In order to extract more specific and unique movements patterns of depressed group, ROIs should be explored more in future by exploiting other partition strategies, which could promote the performance of depression recognition models.

VI. LIMITATION AND FUTURE WORK

In this work, we proposed a sparse OP feature based on the coefficients of facial movements, and the classifier constructed based on it achieved effective depression recognition. Reviewing the whole process of depression, there are still certain aspects that deserve our further improvement and exploration. For example, our model achieved a high recall rate, which might be caused by the similar expression feedback of some subjects in the control group as depressed individuals. Moreover, besides the action units, various partition strategies are need to be explored to extract more effective face regions. Furthermore, it is worth exploring how to derive these independence constraints and threshold values to learn the BN structure more effectively.

REFERENCES

[1] World Health Organization. (2021). *Depression*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[2] A. Jan, H. Meng, Y. F. B. A. Gaus, and F. Zhang, "Artificial intelligent system for automatic depression level analysis through visual and vocal expressions," *IEEE Trans. Cognit. Develop. Syst.*, vol. 10, no. 3, pp. 668–680, Sep. 2018.

[3] J. Shen et al., "An optimal channel selection for EEG-based depression detection via kernel-target alignment," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2545–2556, Jul. 2021.

[4] X. Zhang et al., "EEG-based depression detection using convolutional neural network with demographic attention mechanism," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 128–133.

[5] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The PHQ-9: Validity of a brief depression severity measure," *J. Gen. Internal Med.*, vol. 16, no. 9, pp. 606–613, Sep. 2001.

[6] A. T. Beck, R. A. Steer, and G. Brown, "Beck depression inventory—II," *Psychol. Assessment*, vol. 78, no. 2, pp. 490–498, Jan. 1996.

[7] M. W. O'Hara and L. P. Rehm, "Hamilton rating scale for depression: Reliability and validity of judgments of novice raters," *J. Consulting Clin. Psychol.*, vol. 51, no. 2, pp. 318–319, 1983.

[8] S. Gilbody, T. Sheldon, and A. House, "Screening and case-finding instruments for depression: A meta-analysis," *Can. Med. Assoc. J.*, vol. 178, no. 8, pp. 997–1003, Apr. 2008.

[9] E. Stockings et al., "Symptom screening scales for detecting major depressive disorder in children and adolescents: A systematic review and meta-analysis of reliability, validity and diagnostic utility," *J. Affect. Disorders*, vol. 174, pp. 447–463, Mar. 2015.

[10] Y. S. Ben-Porath, "Assessing personality and psychopathology with self-report inventories," in *Handbook of Psychology: Assessment Psychology*, vol. 10. Hoboken, NJ, USA: Wiley, Apr. 2003, pp. 553–577.

[11] X. Zhang, J. Shen, Z. U. Din, J. Liu, G. Wang, and B. Hu, "Multimodal depression detection: Fusion of electroencephalography and paralinguistic behaviors using a novel strategy for classifier ensemble," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 6, pp. 2265–2275, Nov. 2019.

[12] J. Shen et al., "Exploring the intrinsic features of EEG signals via empirical mode decomposition for depression recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 356–365, 2023.

[13] J. Shen, X. Zhang, G. Wang, Z. Ding, and B. Hu, "An improved empirical mode decomposition of electroencephalogram signals for depression detection," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 262–271, Jan. 2022.

[14] J. Shen, S. Zhao, Y. Yao, Y. Wang, and L. Feng, "A novel depression detection method based on pervasive EEG and EEG splitting criterion," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 1879–1886.

[15] B. Hu, J. Shen, L. Zhu, Q. Dong, H. Cai, and K. Qian, "Fundamentals of computational psychophysiology: Theory and methodology," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 2, pp. 349–355, Apr. 2022.

[16] *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, American Psychiatric Association, Washington, DC, USA, 2013.

[17] K. Qian, Z. Zhang, Y. Yamamoto, and B. W. Schuller, "Artificial intelligence Internet of Things for the elderly: From assisted living to health-care monitoring," *IEEE Signal Process. Mag.*, vol. 38, no. 4, pp. 78–88, Jul. 2021.

[18] B. W. Schuller, J. Löchner, K. Qian, and B. Hu, "Digital mental health—Breaking a lance for prevention," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 6, pp. 1584–1588, Dec. 2022.

[19] K. Qian, T. Koike, K. Yoshiuchi, B. W. Schuller, and Y. Yamamoto, "Can appliances understand the behavior of elderly via machine learning? A feasibility study," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 8343–8355, May 2021.

[20] J. Shen et al., "Depression recognition from EEG signals using an adaptive channel fusion method via improved focal loss," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 7, pp. 3234–3245, Jul. 2023.

[21] A. Pampouchidou et al., "Automatic assessment of depression based on visual cues: A systematic review," *IEEE Trans. Affect. Comput.*, vol. 10, no. 4, pp. 445–470, Oct. 2019.

[22] P. Ekman and W. V. Friesen, *Facial Action Coding System (FACS): A Technique for the Measurement of Facial Movement*, vol. 5, no. 3. Palo Alto, CA, USA: Consulting Psychologists Press, 1978, pp. 56–57.

[23] J. F. Cohn et al., "Detecting depression from facial actions and vocal prosody," in *Proc. 3rd Int. Conf. Affect. Comput. Intell. Interact. Workshops*, Sep. 2009, pp. 1–7.

[24] S. Ghosh, A. Ekbal, and P. Bhattacharyya, "What does your bio say? Inferring Twitter users' depression status from multimodal profile information using deep learning," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 5, pp. 1484–1494, Oct. 2022.

[25] F. Hao, G. Pang, Y. Wu, Z. Pi, L. Xia, and G. Min, "Providing appropriate social support to prevention of depression for highly anxious sufferers," *IEEE Trans. Computat. Social Syst.*, vol. 6, no. 5, pp. 879–887, Oct. 2019.

[26] S. Ghosh and T. Anwar, "Depression intensity estimation via social media: A deep learning approach," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 6, pp. 1465–1474, Dec. 2021.

[27] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[28] S. Wan, Z. Chen, T. Zhang, B. Zhang, and K.-K. Wong, "Bootstrapping face detection with hard negative examples," 2016, *arXiv:1608.02236*.

[29] Y. Li, B. Sun, T. Wu, and Y. Wang, "Face detection with end-to-end integration of a ConvNet and a 3D model," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 420–436.

[30] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.

[31] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1867–1874.

[32] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5325–5334.

[33] H. Qin, J. Yan, X. Li, and X. Hu, "Joint training of cascaded CNN for face detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3456–3465.

[34] D. Chen, G. Hua, F. Wen, and J. Sun, "Supervised transformer network for efficient face detection," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 122–138.

[35] L. Zhang and X. Lu, "Feature extraction based on support vector data description," *Neural Process. Lett.*, vol. 49, no. 2, pp. 643–659, Apr. 2019.

[36] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1113–1133, Jun. 2015.

[37] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[38] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.

[39] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, and X. Fu, "A main directional mean optical flow feature for spontaneous micro-expression recognition," *IEEE Trans. Affect. Comput.*, vol. 7, no. 4, pp. 299–310, Oct. 2016.

[40] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, Mar. 2011.

[41] V. J. Samar, A. Bopardikar, R. Rao, and K. Swartz, "Wavelet analysis of neuroelectric waveforms: A conceptual tutorial," *Brain Lang.*, vol. 66, no. 1, pp. 7–60, Jan. 1999.

[42] X. Song, D. Yan, L. Zhao, and L. Yang, "LSDD-EEGNet: An efficient end-to-end framework for EEG-based depression detection," *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103612.

[43] W. Mumtaz, L. Xia, M. A. M. Yasin, S. S. A. Ali, and A. S. Malik, "A wavelet-based technique to predict treatment outcome for major depressive disorder," *PLoS ONE*, vol. 12, no. 2, Feb. 2017, Art. no. e0171409.

[44] A. A. Abdulrahman and F. S. Tahir, "Face recognition using enhancement discrete wavelet transform based on MATLAB," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 23, no. 2, pp. 1128–1136, 2021.

[45] A. Divya, K. B. Raja, and K. R. Venugopal, "Sorting pixels based face recognition using discrete wavelet transform and statistical features," in *Proc. 3rd Int. Conf. Imag., Signal Process. Commun. (ICISPC)*, Jul. 2019, pp. 150–154.

[46] N. Friedman, "Inferring cellular networks using probabilistic graphical models," *Science*, vol. 303, no. 5659, pp. 799–805, Feb. 2004.

[47] W. Chen, B. Zhu, and H. Zhang, "BN-mapping: Visual analysis of geospatial data with Bayesian network," *Chin. J. Comput.*, vol. 39, no. 7, pp. 1281–1293, 2016.

[48] Y. Lv, J. Miao, J. Liang, L. Chen, and Y. Qian, "BIC-based node order learning for improving Bayesian network structure learning," *Frontiers Comput. Sci.*, vol. 15, no. 6, pp. 1–14, Dec. 2021.

[49] D. M. Chickering, D. Heckerman, and C. Meek, "Large-sample learning of Bayesian networks is NP-hard," *J. Mach. Learn. Res.*, vol. 5, pp. 1287–1330, Oct. 2004.

[50] C. P. D. Campos and Q. Ji, "Efficient structure learning of Bayesian networks using constraints," *J. Mach. Learn. Res.*, vol. 12, no. 3, pp. 663–689, 2011.

[51] M. Koivisto and K. Sood, "Exact Bayesian structure discovery in Bayesian networks," *J. Mach. Learn. Res.*, vol. 5, pp. 549–573, May 2004.

[52] E. Perrier, S. Imoto, and S. Miyano, "Finding optimal Bayesian network given a superstructure," *J. Mach. Learn. Res.*, vol. 9, no. 74, pp. 2251–2286, 2008.

[53] K. Kojima, E. Perrier, S. Imoto, and S. Miyano, "Optimal search on clustered structural constraint for learning Bayesian network structure," *J. Mach. Learn. Res.*, vol. 11, no. 9, pp. 285–310, 2010.

[54] X.-W. Chen, G. Anantha, and X. Lin, "Improving Bayesian network structure learning with mutual information-based node ordering in the k2 algorithm," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 5, pp. 628–640, May 2008.

[55] Y. Li, S. M. Mavadati, M. H. Mahoor, Y. Zhao, and Q. Ji, "Measuring the intensity of spontaneous facial action units with dynamic Bayesian network," *Pattern Recognit.*, vol. 48, no. 11, pp. 3417–3427, Nov. 2015.

[56] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 15, pp. 1454–1462, Apr. 2014.

[57] T. Ni, C. Zhang, and X. Gu, "Transfer model collaborating metric learning and dictionary learning for cross-domain facial expression recognition," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 5, pp. 1213–1222, Oct. 2021.

[58] M. Yang, Y. Ma, Z. Liu, H. Cai, X. Hu, and B. Hu, "Undisturbed mental state assessment in the 5G era: A case study of depression detection based on facial expressions," *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 46–53, Jun. 2021.

[59] Y. Tong, W. Liao, and Q. Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1683–1699, Oct. 2007.

[60] E. L. Rosenberg and P. Ekman, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford, U.K.: Oxford Univ. Press, 2020.

[61] J. F. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the facial action coding system," in *The Handbook of Emotion Elicitation and Assessment* (Series in Affective Science), vol. 1, no. 3, J. A. Coan and J. B. Allen, Eds. Oxford, U.K.: Oxford Univ. Press, 2007, pp. 203–221.

[62] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[63] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 43–77, Feb. 1994.

[64] Y. Chen et al., "A method of removing ocular artifacts from EEG using discrete wavelet transform and Kalman filtering," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 1485–1492.

[65] S. Wang, G. Peng, S. Chen, and Q. Ji, "Weakly supervised facial action unit recognition with domain knowledge," *IEEE Trans. Cybern.*, vol. 48, no. 11, pp. 3265–3276, Nov. 2018.

[66] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn–Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 94–101.

[67] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[68] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," *Pattern Recognit. Lett.*, vol. 32, no. 12, pp. 1598–1603, Sep. 2011.