

Theme Article: Head

Postgraduate Student Depression Assessment by Multimedia Gait Analysis

Haifeng Lu

Lanzhou University

Edith Ngai

The University of Hong Kong

Bin Hu*

Lanzhou University

Shihao Xu

Lanzhou University

Yi Guo

Shenzhen People's Hospital

Xiping Hu*

Lanzhou University

Wei Wang

Sun Yat-sen University

Abstract—In recent years, mental health (especially depression) of university students has aroused general concern. Fast detection of students at risk of depression using multimedia data is a challenge. However, existing methods require the cooperation of participants such as using their speech or facial expression, which are inconvenient to collect and difficult for large-scale screening. In this paper, we propose an integrated gait assessment framework that contains the collection and analysis of multimedia data to assess risk of depression for postgraduate students. First, the rigid body representation is realized by analyzing kinetic energy (KE) and potential energy (PE) generated during walking. Then, we use the Fast Fourier transform (FFT) to analyze KE and PE in the frequency domain for extracting JE-feature. Compared with the conventional methods, our method has significantly increased the objectivity of depression assessment in both clinical theory and practice.

■ **DEPRESSION** has become a common mental disorder in modern society. It is characterized by persistent sadness and a lack of interest or pleasure in previously rewarding or enjoyable activities. According to the World Health Organization (WHO), there are more than 264 million people worldwide suffering from depression. It seriously endangers patients' physical and mental health, and leads to suicidal behavior in the worst case [1]. Recently, emerging research is finding

that postgraduate students are prone to pressures in life and study. They are easily troubled by depression and may have some extreme behaviors (e.g., suicide) [2]. Therefore, a quick method to detect the risk of depression in postgraduate students is demanded.

Due to the fact that consulting psychiatrists is uncommon, it is still difficult to perform extensive mental health counseling for students. Hence, there is a requirement for low cost, quick, and

Department Head

non-contact method that can perform large-scale depression risk screening. Fortunately, the emergence of multimedia technology provides a new opportunity for automatic depression detection. At present, detection methods based on audio and video have been thoroughly studied [3]. However, when collecting audio and video data, participants need to complete different paradigms in a restricted environment.

Besides the above detection methods, recent research has shown that the gait of patients with depression can be observed as abnormal (*e.g.*, gait velocity, stride length, body sway) [4]. Thus, we focus on gait and propose a new method to detect risk of depression. We use Kinect to collect gait information. Kinect V2 can extract human skeleton data and represent it with 25 points, which plays an important role in action recognition and person re-identification [5], [6]. Most researchers use raw data directly for analyze, but they have not studied the importance of different joints for gait analysis in depth [7]. It limits common analysis methods to statistical characteristics such as step length and step width, and fails to describe the changes in the movement process well [8]. To address the above shortcomings, we use Kinect camera to collect gait data and propose a novel gait data analysis method for assessing risk of depression in postgraduate students.

We collected gait data of 95 participants using Kinect camera, including 43 scored-depressed participants and 52 non-depressed participants. They were freshman of postgraduate students from Lanzhou University. First, we preprocess the skeleton data collected by the Kinect camera. Second, we calculate kinetic energy (KE) and potential energy (PE) generated by the participants during walking. Third, a skeleton representation method called rigid-body representation is proposed based on KE and PE, which reduces redundant data. Then, we use the Fast Fourier transform (FFT) to extract the joint energy feature (JE-feature) from KE and PE. Finally, we take advantage of the postgraduate students depression gait dataset, and leverage Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest (RF) to test our approach [8]. Experimental results show that our proposed method can effectively distinguish high risk students with depression up to 91.21%.

Our main contributions are summarized as follows:

- We provide a new approach for detecting postgraduate students at risk of depression through gait information. Our solution is inexpensive, contactless, and is suitable for a wide range of depression risk detection, which provide a new idea for the application of multimedia data.
- Combining clinical experience and joint energy, we propose a rigid-body representation method, which effectively reduces noise and improves classification accuracy.
- We propose a novel gait feature named JE-feature, which is extracted by using FFT from joint energy matrix. The JE-feature can be taken as a new feature in gait analysis for depression detection.

It should be mentioned that this manuscript is a significant extension of our previous in the IEEE Healthcom 2020 conference [9]. As an extension, we propose a new feature named JE-feature, which is extracted by using FFT. Importantly, the new method is more accurate than the previous work. Then, we conduct detailed analysis of joint energy, and increase the discussion of the difference of joint energy between postgraduate students at risk of depression and postgraduate students.

Related work

Automatic depression detection

Many researchers attempted to detect depression in terms of speech, facial expression, EEG, *etc.* In the famous Audio-Visual Emotional Challenge (AVEC), depression detection is a competition theme that has come up many times [3]. Participants used AI technology to judge whether a person is depressed through audio, video, and text in the AVEC 2019. Mel Frequency Cepstral Coefficients (MFCCs) and the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) are often extracted in audio processing. When processing the video data, these features are extracted such as Facial Action Units (FAUs), Bag-of-VisualWords (BoVW), *etc.* Researchers also used pre-trained Convolutional Neural Networks (CNNs) to extract DEEP Spectrum feature [3]. Many machine learning methods are used in the study of EEG for detecting depression. In the

study of EEG signal, filtering algorithm is used to distinguish different wavebands and remove noise. Then, the linear and nonlinear features are extracted for different electrodes to obtain the features matrix of EEG signals [10].

Although researchers found that depression can cause changes in human cognitive abilities, and lead to differences in movements of depression patients from normal people [4], there are still few studies using AI technology to recognize depression through gait.

Depression and gait

Depression patients are typically in a long-term negative emotional state. Their brains reduce neurotransmitters, such as dopamine, serotonin, and noradrenaline [11]. The negative state is not necessarily related to the environment. It may be caused by the imbalance of internal body regulation [12]. People with depression may have different gait from normal people [4].

Researchers found that depression patients displayed reduced gait velocity, vertical head movement, arm swing, and stride length *etc.* [13]. Michalak *et al.* [14] pointed out that the gait pattern of patients with depression is characterized by larger lateral swinging upper body and more slumped posture. In [8], the authors extracted twelve features, such as walking speed, arm swing, stride length, vertical head movement and so on in gait data, and analyzed the relationship between these features and depression. Finally, machine learning algorithms were used to classify the features extracted from gait information into scored-depressed and non-depressed. The authors of [15] focused on extracting a novel time-domain and frequency-domain feature (TF-feature) and a spatial geometric feature (SG-feature), and proposed a novel gait assessment framework to automatic depression detection using Kinect. In [7], Zhao *et al.* extracted features from the frequency domain, which provides us with new ideas.

At present, the research on gait in depression mainly focuses on statistical characteristics, such as stride length, step width, gait velocity, vertical movements, *etc.* [4]. These findings indicate that patients with depression have lower energy consumption movements, which are reflected in a slumped posture during walking [14]. Hence, we proposed a novel method to distinguish between

Table 1: Dataset information.

	Male	Female	SDS(mean)	PHQ-9(mean)	Fragments
Scored-depressed	23	20	62.73	13.04	138
Non-depressed	28	24	36.85	2.04	136

* Standard of division : PHQ-9 score ≥ 10 and SDS score ≥ 60 .

people at risk of depression and normal people through joint energy.

Method

Dataset

In this study, participants were freshman of postgraduate students from different colleges of Lanzhou University in 2018. Their ages range from 22 to 28 years old. The details of dataset are shown in the Table 1. According to the scores of the Physical Health Questionnaire (PHQ-9) and the Self-rating Depression Scale (SDS) [8], they were divided into a scored-depressed group and a non-depression risk group. Since there is no doctor's diagnosis, we only detect postgraduate students at risk of depression according to the scale score. The protocol had obtained permission from the Institutional Review Board of Lanzhou University.

We arranged a 10-meter-long path in the data collection site. Participants needed to walk back and forth on the path two times. Two Kinect cameras that were four meters apart and facing each other. Kinect Software Development Kit (SDK) generated the skeleton joint coordinate streams with a frame rate of 30Hz.

Data Pre-processing

Angle transformation The camera has an angle to the ground, as shown in Figure 1. One can see that the participants do not walk horizontally with Z-axis in Kinect's coordinate system from Figure 2.a. On the one hand, the angle between each camera and the ground is slightly different. On the other hand, the raw data is not in the same coordinate system, and the system errors are imported during analysis, especially when filtered the data of Z-axis. For the purpose of analyzing the data precisely, the coordinate system of raw data has to be transformed. However, due to installation errors and inconvenience of actual

Department Head

measurement, we need to explore a method to automatically compute the camera tilt angle.

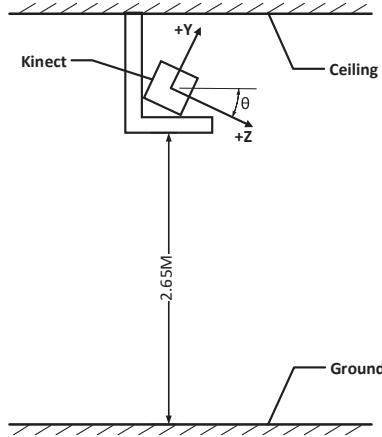


Figure 1: The schematic diagram of Kinect installation.

We use SpineBase joint as a reference to obtain the angle between the Kinect camera and the ground. First, we divide all fragments into two groups based on the two Kinect cameras. Then, we extract the value of Z-axis of SpineBase and Y-axis of SpineBase in each fragment, and use the Ordinary Least Square (OLS) to fit the slope. The formula is shown in Eq 1. Finally, the slopes of all clips under the same camera are added and averaged to get θ .

$$\tan \theta = \frac{n \sum z_i y_i - \sum z_i \sum y_i}{n \sum z_i^2 - (\sum z_i)^2}, \quad (1)$$

where $\tan \theta$ is the slope between the Kinect camera and the ground, n denotes the number of video frames, z_i and y_i represent the values of Z-axis of SpineBase and Y-axis of SpineBase, respectively.

After the angle between the camera and the ground is obtained, we multiply the transformation matrix with each frame. The transformation matrix is shown in Eq 2. The result of raw data rotation is shown in Figure 2.b.

$$\begin{bmatrix} x' & y' & z' \end{bmatrix} = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \quad (2)$$

where $[x' \ y' \ z']$ is defined as the joint position after coordinate transformation, $[x \ y \ z]$ is raw data.

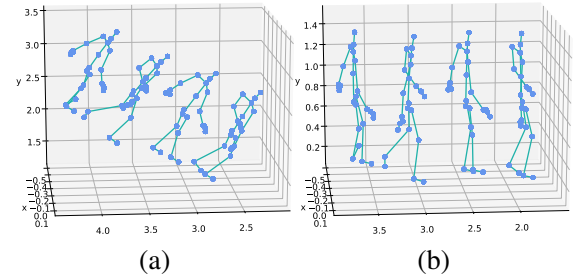


Figure 2: Comparison before and after angle transformation.

Re-sampling and De-noising During the data acquisition, many factors, such as light and color of participants's clothes, influenced the quality of data collected by Kinect. We use Gaussian Filter to reduce noise and smooth the data [7]. We choose a sliding window of length 5 to filter the three axes, respectively.

The participants need to walk back and forth during the data collecting, hence the part of data is that the participants walk towards the cameras, and the others are that the participants walk away from the cameras. Previous research has shown that data toward Kinect is more accurate [8]. We also find that when the participants walk away from the cameras, and skeleton data is prone to deformation. Thus, we only keep the front segments in feature extraction.

Each participants's gait fragment varies in length, and the data quality at the beginning and the end of the fragments are usually bad. While retaining the necessary information, it should crop the fragment. Finally, 64 frames of each fragment are reserved based on the paper of Li *et al.* [16].

KE and PE

In physics, KE is defined as the energy produced by the movement of object, and PE is defined as the energy stored in a system. Gait velocity, step length, step width *etc.* are important indicators in gait research on depression. The above indicators are closely related to KE and PE. For example, increasing gait velocity inevitably produces greater KE, and increasing movement

range produces greater PE. Consequently, we try to calculate KE and PE in each frame, and analyze the difference between Scored-depressed groups and Non-depressed groups from KE and PE.

KE calculation formula is shown in Eq (3). Considering that human body is represented by points in the raw data, we ignore the mass of limbs when calculating KE of joints, and only focus on velocity [9]. The v_i^j is the velocity of the same joint between adjacent frames, which is obtained by dividing d_i^j by the sampling time interval Δf , the calculation formula is shown in Eq (4). d_i^j is the Euclidean distance of same joint in adjacent frames, the calculation formula is shown in Eq (5). Import (4) and Eq (5) into Eq (3) to obtain Eq (6).

$$E_k = \frac{1}{2}mv^2, \quad (3)$$

where m is mass of the body, v represents velocity of the body.

$$v_i^j = \frac{d_i^j}{\Delta f}, \quad (4)$$

where i is index of frame, j means different joints.

$$d_i^j = \sqrt{\sum_{a=1}^3 \left(J_{i,a}^j - J_{i-1,a}^j \right)^2}, \quad (5)$$

where $J_{i,a}^j$ corresponds to the values of three coordinate axes.

$$E_{k,i}^j = \frac{\sum_{a=1}^3 \left(J_{i,a}^j - J_{i-1,a}^j \right)^2}{2\Delta f^2}, \quad (6)$$

We define a gait potential energy by analogy with gravitational potential energy, the calculation formula of gravitational potential energy is shown in Eq (7). We also ignore the mass of limbs when calculating PE of joints. Since the acceleration of gravity is a common variable, we do not take it into account. Therefore, we only focus on h and h_0 .

$$E_p = mg(h - h_0), \quad (7)$$

where g is gravitational acceleration, h is the height of the object, h_0 is defined as the height of the zero of potential energy.

When a person is walking, the joint position

changes relative to the zero of potential energy. We call the relative displacement as gait potential energy. Since the height and walking track of each participant are different, it is not properly to define the ground as the zero of potential energy. We propose to define gait neutral pose (GNP) as the zero of potential energy to calculate PE during walking [9], [17]. We select a special frame in each fragment as GNP. The GNP frame should be consistent with the standing posture as far as possible, keeping both hands on both sides, stretching both legs, and keeping the body upright. Figure 3 shows the comparison between GNP and normal walking process. Finally, PE is calculated by subtracting the corresponding GNP in each fragment, Its formula is shown in Eq (8).

$$E_{p,i}^j = F_i - G, \quad (8)$$

where F_i is different frame, G is gait neutral pose.

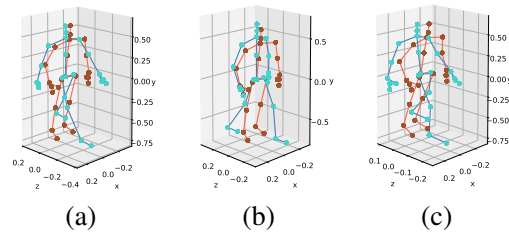


Figure 3: Comparison of instant poses (steelblue lines) and their corresponding pseudo neutral poses (sienna lines) in different individuals.

Rigid-body representation

Some joints have small motion ranges and almost the same motion cycle, such as SpineBase joints and SpineMid joints. By analyzing movement and joint energy, raw data can be rationally simplified to achieve the purpose of removing redundant data and improving accuracy. Eq (9) in [17] is used to calculate total joint energy.

$$E = \sum_{i=1}^n (E_{k,i} + E_{p,i}), \quad (9)$$

According to the calculation results of joint energy in Eq (9), drawing the joint energy diagram, as shown in Figure 4. The darker joint color in the picture, the greater joint energy.

Most people keep their hands in the same posture during walking, but the calculation results

Department Head

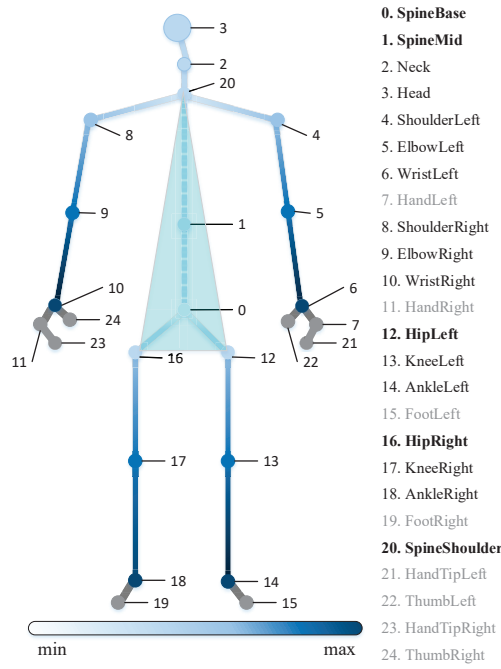


Figure 4: The diagram of joints energy.

show that the energy of four joints of hand is obviously different. It is pointed out in some researches that Kinect's positioning of the hand joints is prone to errors. Since the cameras are placed above the trail, there is also a certain distortion in the positioning of the participants's foot joints. In summary, we take the average of the four joints of hand as the hand joint, and take the average of foot joints and ankle joints as foot joint. After the above processing, 8 joints are removed, as shown in gray in Figure 4.

Trunk joints have less relative movement, and maintain similar movement periods during walking. Therefore, Figure 4 shows that all of the trunk joints have the least energy joint. In [18], the authors proposed that seven joints (0,1,4,8,12,16,20) can be regarded as a rigid body to simplify subsequent calculation. However, in [8], the authors proposed that the range of motion of shoulder is also a very important gait feature. Thus, we propose a rigid-body representation method, which defines five joints of trunk (0,1,12,16,20) as a rigid-body. First, five trunk joints are averaged and named torso joint, as shown by the triangle in Figure 4. Second, all the joints minus torso joint. Finally, there are only 12 joints to be kept.

JE-feature extraction and Classification

Given that the rigid-body representation re-estimates joint positions, KE and PE of different joints in each frame should be recalculated. Two matrices that are KE matrix 12×64 and PE matrix 12×64 are obtained for each fragment after recalculation. Since the value range of the KE matrix and PE matrix are too large, which affects the results of subsequent feature extraction, it is also necessary to min-max normalize the energy matrix.

In [16], the authors used FFT to extract features from skeletal data. This article also uses the same method. We use FFT (defined as Eq (10)) on each column of the PE matrix and KE matrix, and get 64 amplitude coefficients from each column. Then, we select the 20 largest phases and their corresponding frequencies from each column as features (2×20). Next, we use the Pearson correlation to select the 10 features that have the largest absolute value of correlation coefficients in each column. Finally, these features from PE and KE are combined and normalized to $(-1, 1)$ range, and called as the JE-feature.

$$JE = \sum_{m=0}^{n-1} E_m e^{-i2\pi k \frac{m}{n}}, k = 0, \dots, n-1. \quad (10)$$

where JE is defined as extracted features, N is the length of the data fragment, E_m represents the energy of the m th frame.

A 12×20 -dimension JE-feature is obtained for each fragment. We use Principal Components Analysis (PCA) to reduce dimension, and keep the vectors that contribute more than 80%. In order to minimize the contingency and uncertainty of the classification results, leave-one-out cross-validation strategy is adopted. We randomly select a person, and use the all fragments of the person as the test data, and other people's data as the training data. According to the above strategy, each person's fragment is classified, and the final classification result is obtained through voting. The classifiers are SVM ($C = 1.0$, kernel = 'sigmoid'), KNN ($k = 7$), and RF ($nTree = 100$) [9].

Table 2: Performance of each recognition model in various skeletal representation.

	Classifier	ACC	SEN	SPE	F-score
Raw	KNN	0.7564	0.8960	0.5963	0.7972
	SVM	0.9011	0.9762	0.8367	0.9011
	RF	0.8462	0.9762	0.7347	0.8542
Lower body	KNN	0.6709	0.8640	0.4495	0.7372
	SVM	0.8681	0.9762	0.7755	0.8723
	RF	0.7363	0.9286	0.5714	0.7647
Upper body	KNN	0.6624	0.7280	0.5872	0.6973
	SVM	0.7582	0.9524	0.5918	0.7843
	RF	0.7143	0.8810	0.5714	0.7400
Rigid-body	KNN	0.8034	0.9200	0.6697	0.8333
	SVM	0.9121	0.9762	0.8571	0.9111
	RF	0.8571	0.9762	0.7551	0.8632

The bold indicates the optimal result under the same index, the same as follow.

Experiment results and Discussion

The following evaluates our approach in the dataset. These evaluation indicators are used, including accuracy (ACC), Sensitivity (SEN), Specificity (SPE), and F-score [19].

Experimental Results of Rigid-body Representation

In previous studies, there are three main skeletal represent methods. First, the raw Kinect skeleton (19 joints) data is used after averaging hands and feet. Second, the skeleton data of the upper body (10 joints) is used. Third, the skeleton data of the upper body (6 joints) is used [8]. In order to verify the validity of rigid-body representation, we use the above representation methods respectively according to the feature extraction method of this article. Table 2 shows the classification results.

One can see that the classification result of rigid-body representation is the best among four representation methods from the data in Table 2. Comparing with raw data, the accuracy of rigid-body representation is improved by 2.3% on average. Although the accuracy is not significantly improved, the classification time of rigid-representation is shorter than other representation methods. When only upper body data or lower body data are used, it significantly reduces model performance. The above experiments show that rigid-body representation can reduce data size and

Table 3: Classification results of different energy features.

	Classifier	ACC	SEN	SPE	F-score
PE	KNN	0.7094	0.8000	0.6055	0.7463
	SVM	0.7473	0.8571	0.6531	0.7579
	RF	0.7473	0.9286	0.5918	0.7723
KE	KNN	0.7137	0.7760	0.6422	0.7433
	SVM	0.8352	0.9048	0.7755	0.8352
	RF	0.7363	0.9524	0.5510	0.7692
PE+KE	KNN	0.8034	0.9200	0.6697	0.8333
	SVM	0.9121	0.9762	0.8571	0.9111
	RF	0.8571	0.9762	0.7551	0.8632

ensure that effective information is not lost.

Experimental Results of JE-feature

KE and PE can reflect the movement changes of body during walking from different aspects. In order to prove that JE-feature can distinguish high-risk of depression and low-risk of depression, three groups of experiments are set up. The first and second groups perform FFT on KE matrix and PE matrix respectively, and analyze frequency-domain features of KE and PE, which called KE-feature and PE-feature. The third group combines KE matrix and PE matrix, then performs FFT on the combined matrix. The features obtained by FFT are called JE-feature. Table 3 shows the experimental results of the three groups of experiments.

We can observe when PE-features and KE-features are used alone, performing the classifier is not ideal from Table 3. This shows that it is not comprehensive to analyze participants's gait only from KE and PE. Compared with PE-features and KE-features, the average accuracy of classification after using JE-feature is increased by 12.30% and 9.56% respectively. Experimental results show that JE-feature can be used as a new feature to assess postgraduate students at risk of depression.

Comparison With Different Methods

We compare the performance of time domain features, frequency domain features, and JE-feature in order to better evaluate our method using simply SVM as a classifier. Twelve time-domain features mainly include walking speed, stride length, step width, arm swing *etc.* [8]. The

Table 4: Comparison of the results between our method and other methods.

	ACC	SEN	SPE
Time-domain features	0.8526	0.7907	0.9038
Frequency-domain features	0.8571	0.8595	0.8544
Our method	0.9121	0.9762	0.8571

extraction of frequency-domain featurese mainly refer to Li *et al.* [20]. The authors used FFT to analyze the frequency domain feature of gait data. Table 4 reports the accuracy, sensitivity, and specificity of the difference between the proposed method and other methods.

As shown in Table 4, it can be easily observed that our method achieve the state-of-the-art results. Although the accuracy of classification after using time-domain features is very close to our method, the sensitivity is only 79.07%. This may lead to a high rate of missed diagnosis, and is not suitable for the diagnosis of depression. The accuracy of classification after using frequency-domain features has decline, which is only 85.71%. In summary, our method has more balanced indicators in all aspects, and has more application space.

Discussion

The experiment in Table 2 shows that rigid-body representation can reduce redundant data and improve classification accuracy. Combining joint energy and clinical experience, we simplify the joints, and reduce data size by half, and decrease unnecessary computation time. Yet, the reduction of data size does not mean loss of effective data. According to the results in Table 2, we find that after using the rigid-body representation model, accuracy and F-score of two classifiers improve significantly, which indicate that the deleted information may contain a lot of redundant.

The experiment in Section demonstrates that JE-feature are workable as a gait feature for classifying risk of depression. In order to further analyze the impact of energy of each joint on the classification results, the joint energy of a person's same joint is added and compared with all samples to obtaining the mean and standard deviation. Table 5 shows the calculation results.

The p-value is obtained using the independent sample T-test. False discovery rate (FDR) is used for correcting after the p value is obtained.

In previous studies, many researchers focused on movement of the lower limbs. Participants' gait velocity and step length are important indicators. We calculate the total energy of each person's ankle joint that is closely related to the above indicators. Surprisingly, there is not much difference in ankle joint energy between scored-depression students and non-depression students during walking from the data. The p-value of ankle joint energy in Table 5 also shows that there is no significant difference between two groups. Although the PE of KneeRight is significantly different ($p < 0.05$), this difference is not enough to be called an indicator to distinguish postgraduate students at risk of depression. Since different people's leg movements during walking are greatly related to their daily habits, height, gender, *etc.*. Even if some people at high risk of depression, energy changes in the leg joints are still normal. Therefore, it is difficult for us to effectively distinguish the difference between scored-depression students and non-depression students based on the changes in joint energy of leg.

However, we find that after the total joint energy of all joints are added, scored-depression students and non-depression students show obvious differences. We find that the joint energy of wrist and elbow joints is significantly different ($p < 0.001$) between scored-depression students and non-depression students. Figure 6 shows the quartile of wrist joint energy, Figure 5 shows the quartile of elbow joints energy. One can see that scored-depression students and non-depression students have enormous differences ($p < 0.05$) in KE and PE, which can be used as an indicator for distinguishing.

Combining the above analysis, we conclude that joint energy is a very important indicator for gait analysis. Regardless of scored-depression students and non-depression students, there is no significant difference in the energy changes in the legs of the subjects during walking, but we should focus on the swing of the participants' arms. Participants in scored-depression students tend to reduce arm swings and habitually kept their arms at their sides.

Table 5: Energy comparison of different joints.

	PE			KE		
	Non-depression	Scored-depressed	p-value	Non-depression	Scored-depressed	p-value
All of joints	131±13.96	122.92±11.30	p < 0.001	81.98±7.47	75.05±7.97	p < 0.001
Neck	1.67±0.87	1.9±1.05	p = 0.094	0.41±0.34	0.41±0.33	p = 0.725
Head	2.18±1.01	2.53±1.04	p < 0.05*	0.44±0.29	0.48±0.32	p = 0.086
ShouderLeft	1.55±0.9	1.74±1.06	p = 0.13	0.29±0.26	0.33±0.24	p < 0.05*
ElbowLeft	6.71±1.78	6.05±2.47	p < 0.05*	2.18±0.79	1.89±1.06	p < 0.001*
WristLeft	19.2±4.59	16.49±6.3	p < 0.001*	11.67±4.26	9.1±4.84	p < 0.001*
ShoulderRight	1.54±0.75	1.57±0.9	p = 0.59	0.27±0.22	0.3±0.21	p = 0.136
ElbowRight	6.22±2.24	4.63±2.37	p < 0.001*	1.91±0.88	1.33±0.95	p < 0.001*
WristRight	17.16±6.02	12.8±5.9	p < 0.001*	9.81±4.62	6.16±4.45	p < 0.001*
KneeLeft	12.47±2.67	12.14±2.56	p = 0.094	5.8±1.84	5.97±2.0	p = 0.779
AnkleLeft	25.82±2.34	26.23±2.84	p = 0.093	21.74±2.4	21.61±2.5	p = 0.677
KneeRight	12.6±2.32	11.87±2.9	p < 0.05*	5.96±1.72	5.76±1.98	p = 0.252
AnkleRight	24.81±2.62	24.97±2.88	p = 0.088	21.49±2.45	21.72±2.82	p = 0.058

* There is a significant difference after FDR corrected.

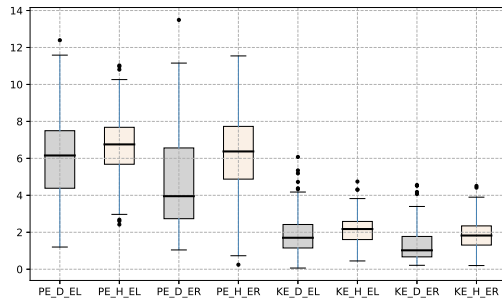


Figure 5: PE and KE in elbow joint. (D - Scored-depressed, H - Non-depression, EL - EblowLeft, ER - EblowRight)

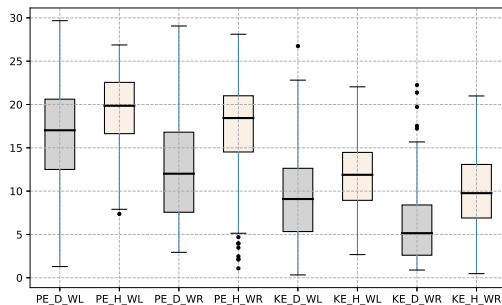


Figure 6: PE and KE in wrist joint. (D - Scored-depressed, H - Non-depression, WL - WristLeft, WR - WristRight)

Conclusions and Future Work

In this paper, we propose a novel approach using multimedia data for classifying postgraduate students at risk of depression. Experimental results show that the accuracy of our method for identifying the students at high-risk of depression can achieve 91.21%. In addition, we analyze the characteristics of joint energy in the students at high-risk of depression, it is found that the joint energy of the elbow and wrist joints of the students at high-risk of depression is significantly less than that of the normal students. In future work, we would like to achieve faster and more accurate recognition. Meanwhile, we will further investigate the relationship between gait and depression, and consider regression of scale data for high-risk groups to make more accurate risk detection. Furthermore, our research is not limited to the student groups. Our next step is to extend our study to other age groups and special populations.

ACKNOWLEDGMENT

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFA0706200), in part by the National Natural Science Foundation of China (Grant No.61632014, No.61627808, No.61210010), in part by the National Basic Research Program of China (973 Program, Grant

Department Head

No.2014CB744600), and in part by the Program of Beijing Municipal Science & Technology Commission (Grant No.Z171100000117005).

REFERENCES

1. "World health organization," <https://www.who.int/news-room/fact-sheets/detail/depression>, accessed on 11 August 2020.
2. F. Tang, M. Byrne, and P. Qin, "Psychological distress and risk for suicidal behavior among university students in contemporary china," *Journal of Affective Disorders*, vol. 228, pp. 101–108, 2018.
3. F. Ringeval, B. Schuller, M. Valstar, N. Cummins, R. Cowie, L. Tavabi, M. Schmitt, S. Alisamir, S. Amiriparian, E.-M. Messner *et al.*, "Avec 2019 workshop and challenge: state-of-mind, detecting depression with ai, and cross-cultural affect recognition," in *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, 2019, pp. 3–12.
4. F. Deligianni, Y. Guo, and G.-Z. Yang, "From emotions to mood disorders: A survey on gait analysis methodology," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, no. 99, pp. 1–1, 2019.
5. "Microsoft kinect," <https://msdn.microsoft.com/>, accessed on 11 January 2021.
6. H. Rao, S. Wang, X. Hu, M. Tan, Y. Guo, J. Cheng, B. Hu, and X. Liu, "A self-supervised gait encoding approach with locality-awareness for 3d skeleton based person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
7. N. Zhao, Z. Zhang, Y. Wang, J. Wang, B. Li, T. Zhu, and Y. Xiang, "See your mental state from your walk: Recognizing anxiety and depression through kinect-recorded gait data," *PloS one*, vol. 14, no. 5, 2019.
8. J. Fang, T. Wang, C. Li, X. Hu, E. Ngai, B.-C. Seet, J. Cheng, Y. Guo, and X. Jiang, "Depression prevalence in postgraduate students and its association with gait abnormality," *IEEE Access*, vol. 7, pp. 174 425–174 437, 2019.
9. H. Lu, W. Shao, E. Ngai, X. Hu, and B. Hu, "A new skeletal representation based on gait for depression detection," in *2020 IEEE International Conference on E-health Networking, Application Services (HEALTHCOM)*, 2021, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/9399002>
10. H. Cai, Z. Qu, Z. Li, Y. Zhang, X. Hu, and B. Hu, "Feature-level fusion approaches based on multimodal eeg data for depression recognition," *Information Fusion*, vol. 59, pp. 127 – 138, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1566253519302143>
11. H. Lvheim, "A new three-dimensional model for emotions and monoamine neurotransmitters," *Medical Hypotheses*, vol. 78, no. 2, pp. 341–348, 2012.
12. M. Kalia, "Neurobiological basis of depression: an update," *Metabolism Clinical & Experimental*, vol. 54, no. 5-suppl-S, pp. 24–27, 2005.
13. M. R. Lemke, T. Wendorff, B. Mieth, K. Buhl, and M. Linnemann, "Spatiotemporal gait patterns during over ground locomotion in major depression compared with healthy controls," *Journal of psychiatric research*, vol. 34, no. 4-5, pp. 277–283, 2000.
14. J. Michalak, N. F. Troje, J. Fischer, P. Vollmar, T. Heidenreich, and D. Schulte, "Embodiment of sadness and depression–gait patterns associated with dysphoric mood," *Psychosomatic Medicine*, vol. 71, no. 5, pp. 580–587, 2009.
15. T. Wang, C. Li, C. Wu, C. Zhao, J. Sun, H. Peng, X. Hu, and B. Hu, "A gait assessment framework for depression detection using kinect sensors," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3260–3270, 2021.
16. B. Li, C. Zhu, S. Li, and T. Zhu, "Identifying emotions from non-contact gaits information based on microsoft kinects," *IEEE Transactions on Affective Computing*, vol. 9, no. 4, pp. 585–591, 2016.
17. C. Tang, W. Li, P. Wang, and L. Wang, "Online human action recognition based on incremental learning of weighted covariance descriptors," *Information Sciences*, vol. 467, pp. 219–237, 2018.
18. M. Raptis, D. Kirovski, and H. Hoppe, "Real-time classification of dance gestures from skeleton animation," in *Proceedings of the 2011 ACM SIGGRAPH/Eurographics symposium on computer animation*, 2011, pp. 147–156.
19. M. Woźniak, M. Wiecek, J. Siłka, and D. Polap, "Body pose prediction based on motion sensor data and recurrent neural network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2101–2111, 2021.
20. Y. Yuan, B. Li, N. Wang, Q. Ye, Y. Liu, and T. Zhu, "Depression identification from gait spectrum features based on hilbert-huang transform," in *International Conference on Human Centered Computing*. Springer, 2018, pp. 503–515.

Haifeng Lu is currently pursuing the Ph.D. degree at the Gansu Provincial Key Laboratory of Wearable Computing, School of information Science and Engineering, Lanzhou University. His research interests

include affective computing, machine learning. Contact him at luhf18@lzu.edu.cn.

Shihao Xu is currently pursuing the Ph.D. degree in computer application technology with the School of Information Science and Engineering in Lanzhou University, Lanzhou, China. His major research interests include skeleton based action recognition and contrastive learning. Contact him at xushh16@lzu.edu.cn.

Xiping Hu received the Ph.D. degree from the University of British Columbia, Vancouver, BC, Canada. He is currently a professor with Lanzhou University, China. He has more than 100 papers published and presented in prestigious conferences and journals. He was the Co-Founder and CTO of Bravolol Ltd., Hong Kong, a leading language learning mobile application company with over 100 million users, and listed as the top 2 language education platform globally. Contact him at huxp@lzu.edu.cn. He is a co-corresponding author of this article.

Edith Ngai is currently an Associate Professor in Department of Electrical and Electronic Engineering, The University of Hong Kong. She received her Ph.D. from The Chinese University of Hong Kong in 2007. She did her post-doc in Imperial College London, United Kingdom. Her research interests include Internet-of-Things, machine learning, smart cities and health informatics. She was a project leader of the national project, GreenIoT, for sustainable development in Sweden. She is a senior member of ACM and IEEE. Contact him at chngai@eee.hku.hk.

Yi Guo received the Ph.D. degree from the University of Greifswald, Greifswald, Germany, in 1997. He is currently the Chief of Neurology with the Second Clinical Medical College, Jinan University, Guangzhou, China, the Chairman of the Shenzhen Medical Association of Neurology and the Shenzhen Medical Association of Psychosomatic Medicine. His major research interests include cerebrovascular diseases, dementia, movement disorder diseases, sleep disorder, depression, and anxiety. Contact him at xuanyi_guo@163.com.

Wei Wang received the B.Sc. degree from Shenyang University, Shenyang, China, in 2012, and the Ph.D. degree in software engineering from the Dalian University of Technology, Dalian, China, in 2018. His research interests include big scholarly data, social network analysis, and computational social science. Contact him at wangw328@mail.sysu.edu.cn.

Bin Hu received Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Science, China in 1998. Since 2008, He has been a professor of School of Information Science and Engineering, Lanzhou University, China. His research interests include Pervasive Computing, Computational Psychophysiology, and Data Modeling. Contact him at bh@lzu.edu.cn. He is a co-corresponding author of this article.