Journal of Network Intelligence ©2023 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **8**, Number **4**, November **2023**

Academic Emotions Cognition Model on Multi-Kernel SVM in Human-Computer Interaction

Chao Yang*

Youth League committee
Zhejiang College of Security Technology
Wenzhou 325000, China
yangchaosundra@126.com

Kypraios Wassim

School of Education
Grenfell Campus, Memorial University of Newfoundland
Corner Brook A2H 5G4, Canada
kysdzmpjnh33@gmail.com

*Corresponding author: Chao Yang

Received May 12, 2023, revised July 28, 2023, accepted September 26, 2023.

Abstract. *The online intelligent learning auxiliary system at the present stage lacks effective emotion cognition function in the human-computer interaction process. To solve this problem, a cognitive model of academic emotions based on an optimised support vector machine is proposed. Firstly, based on theories of emotional psychology and pedagogical psychology, the learning characteristics generated by learners in normal learning states are analysed, and a novel cognitive model of academic emotion with reverse emotion is established from three dimensions. Then, a region chunking strategy is used to initially process the web video and extract the main features of face expressions based on the attribute factors in each region. Then, the support vector machine is optimised by local linear embedding (LLE) data dimensionality reduction and selection of hybrid kernel functions. Finally, the optimised support vector machine was used to build a cognitive model of learners' emotions in an online teaching context. The experimental results show that by setting the number of LLE neighbours and kernel functions appropriately, the optimised support vector machine can achieve high detection performance. Compared with other support vector machines, the optimised support vector machine based emotion cognition model obtained a more accurate academic emotion recognition rate.*
Keywords: Artificial intelligence; Academic emotions; Cognitive models; Support vector machines; Locally linear embedding

---

1. **Introduction.** In 2019, New Crown Pneumonia began to see widespread spread around the world, causing a huge impact on people's lives and work. New Crown pneumonia has also led to a huge crisis in the global economy. The global spread of New Crown Pneumonia has had a huge impact on the education sector [1, 2]. This has had a negative impact on the school curriculum schedule. In order not to delay the progress of the curriculum, most schools have launched online web-based teaching. Online teaching has changed the relationship between teacher and student in traditional teaching [3,4]. Educational psychology suggests that a personalised artificial intelligence system should not only be intelligent, but also emotional.

Due to their inability to sense students' learning emotions, existing online learning systems are unable to process the emotional information generated by students in a timely manner during the teaching process [5,6], thus resulting in teachers being unable to communicate with students emotionally in a timely manner. The problem of emotion deficit in online learning systems leaves a shortfall in emotionally personalised education [7, 8]. If the emotional deficit cannot be addressed, online learning will have a large gap with the actual teaching context. The establishment of this personalized intelligent network teaching system based on cognitive and emotional interaction mode involves research fields such as face detection and expression recognition, artificial emotion, psychology and emotional calculation [9, 10]. How to establish an intelligent learning assistant system with emotional cognitive function by using the technical means in the above fields under the network environment is a hot issue in the field of artificial intelligence at present [11,12].

Expressions are one of the most important ways to understand other people's emotions and emotional states [13,14], and can enable the transmission of emotional information between people. Expressions can be divided into three categories: facial expressions, verbal expressions and physical expressions, among which facial expressions can directly reflect a person's emotional characteristics and emotional state [15,16]. The above phenomenon suggests that analysing facial expressions during learning can provide a basis for making decisions about academic emotion perceptions. The use of learners' expressions to build emotional cognitive models can, to a certain extent, solve the problem of emotional deficits in online teaching [17]. Through personalised emotion teaching, a positive transfer of learning emotions can be achieved, which is of great practical significance in realising the development of personalised digital education.

1.1. **Related Work.** An online learning system with emotional interaction contains many aspects, such as emotion recognition, emotion modelling, emotion and cognition, and pedagogical reasoning [18], among which the most crucial parts are emotion recognition and emotion modelling.

For emotion recognition, Hong [19] proposed to combine virtual reality technology with online teaching and learning. When learners are learning in a virtual learning environment, an intelligent learning partner recognises the learner's emotions and thus behaves accordingly. Jonauskaite et al. [20] studied the theory of artificial emotion and the "emotional software human" model and proposed an architecture for applying anthropomorphic robots. Song et al. [21] studied the precise location technology of learners' faces, and proposed a facial expression feature extraction method, thus realizing effective academic emotion recognition. Zhu et al. [22] proposed a support vector machine (SVM)-based emotion recognition model that can accurately recognise facial expressions and thus obtain the detection of many different emotions.

The above emotion recognition methods all use a single image feature extraction, e.g., luminance features, color features, etc. However, single feature extraction can no longer meet the demand of high accuracy facial expression recognition, so this paper uses multi-feature fusion technique to solve this problem. In addition, as an important method of multidimensional feature correlation analysis, the local linear embedding (LLE) technique [23,24] can correlate variable features of different dimensions in order to remove redundant features.LLE can reduce the noise interference while reducing the dimensionality of variables, which helps to reduce the computational complexity of the model [25]. The LLE can reduce the dimensionality of the variables while reducing the interference of noise, which helps to reduce the complexity of the model and improve the final recognition accuracy.

Emotion modelling is the representation of human emotions using appropriate mathematical models, with the aim of implementing machine systems with artificial emotions.

Katerndahl et al. [26] proposed a simple and effective method for modelling emotions - the state-space approach. The state-space approach considers three simple emotions, namely anger, fear and happiness, and uses these three emotions to construct an emotion space. Shafti et al. [27] proposed a model of emotion that includes 22 basic emotions. This model assumes that emotions are dispositional responses to events, agents and states, thus greatly enhancing emotion generation. In addition to these aforementioned emotion modelling, there are also emotion models based on grey system theory. Most of the existing emotional models are aimed at common emotions, but not at learners' academic emotions.

1.2. **Motivation and contribution.** In order to realise the emotional cognitive interaction function of the online learning system, this paper proposes an academic emotional cognitive model based on optimised SVM. The learner's facial expression features will be used as a basis for determining academic emotions, and the construction of an emotion cognitive model will be implemented in a three-dimensional emotion space in combination with a personalised model. The main objective of this study is to determine the learners' emotional responses to the current learning content by analysing the learner's facial expression features collected in the e-learning context, so as to provide emotional functional support for personalised teaching in online courses.

The main innovations and contributions of this paper include.

(1) Based on theories of emotional psychology and pedagogical psychology, we analyzed the learning characteristics generated by learners in normal learning states, and developed a new cognitive model of academic emotion with reverse emotion in three dimensions

(2) Feature extraction using multi-feature fusion of regional attributes in order to obtain more information about the learner's facial expression features through video.

(3) A multi-core SVM was used instead of the traditional SVM for facial expression recognition, and LLE data was used for dimensionality reduction in order to improve the accuracy of emotion recognition while increasing the generalization ability of the samples.

2. **A cognitive model of academic emotions with reverse emotions.**

2.1. **Definition of emotions and relationship to cognition.** Emotion is the mental state of a series of subjective perceptions that reflect the complex relationship between a person's subjective needs and objective things. Psychologists believe that emotions and cognition are inextricably linked. When we are in a good emotion to work and study, our minds become more alert and open than usual. However, when we are depressed, our thinking becomes much slower than usual. Research on emotions and cognition has led to two conclusions [28].

(1) Emotions arise from the perception of the stimulus situation

When an individual is associated with the external environment, the individual will experience some emotions as a result of past cognitive experiences as a result of stimuli from the external environment. Emotional development and cognition are also closely linked.

(2) Emotions affect cognition

Emotional psychology research shows that positive emotions have a positive effect on cognitive activity. When an individual has a positive emotion, this positive emotion accelerates the individual's thinking and thus drives cognitive activity. Conversely, when an individual is in a negative emotional state, this unhealthy emotion can inhibit or hinder cognitive activity.

2.2. **Academic emotions and perceptions.** There are currently two main types of classification of academic emotions: some researchers classify academic emotions as being related to pleasurable and unpleasant, while other researchers classify academic emotions as positive and negative. According to research, academic emotions are closely related to students' cognitive and physical and mental health. Positive and negative academic emotions have a facilitating and hindering effect on academic performance respectively. Positive academic emotions broaden cognitive scope and increase cognitive flexibility. Academic emotions include all kinds of academic-related emotional experiences, such as happiness, boredom, confusion, etc.

Taken together, academic emotions and cognition go hand in hand in the learning process. If students lack cognition in the learning process, then it will make it difficult to progress emotionally. If students lack more positive emotions in the learning process, then their cognitive abilities will also be greatly diminished.

2.3. **Three-dimensional academic emotion space with reverse emotion.** The emotional state space model considers three simple emotions, namely angry, scared and happy, and uses these three emotions to construct an emotional space, as shown in Figure 1. For affective teaching, it is not at all necessary to examine all emotions of the learner,

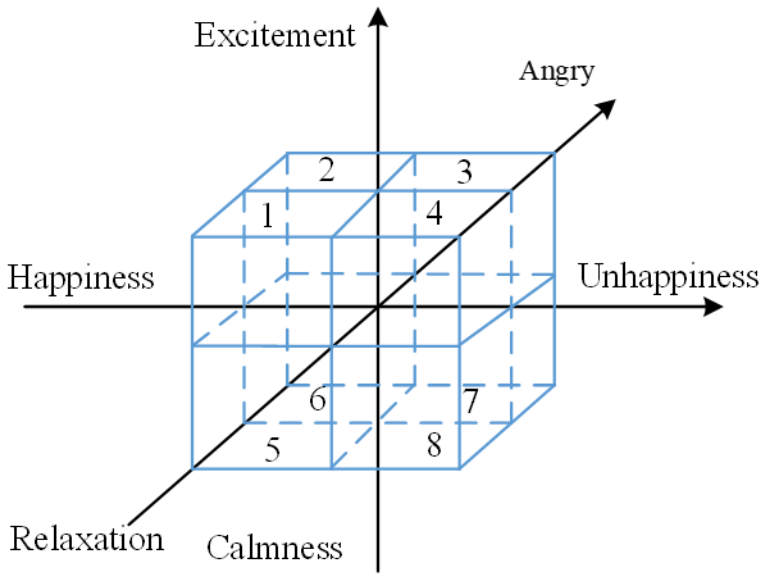


Figure 1. Emotional state space

but only to capture those main emotions that are closely related to the learning process. In line with this view, this paper will develop models of learner interestingness, concentration and happiness and map them onto a three-dimensional academic emotion space with reversed emotions. Academic emotions are defined as three dimensions, grouped into the six most common reciprocal academic emotions, including interest, boredom, exhilaration, fatigue, pleasure and distress [29]. Of these, interest, exhilaration and pleasure are positive emotions. Boredom, fatigue and distress are negative emotions.

Expressions are the external manifestation of emotions and can be used to understand a person's emotional state at the time. Expressions can be divided into facial expressions, verbal expressions and physical expressions. Facial expressions are more important than the other two types of expressions. Therefore, the academic cognitive emotion model established in this paper uses the change of facial expression as an input parameter.

The interestingness model determines how interested the learner is in the current learning content during the learning process by calculating the size of the face contour and the size of the pupil spacing. The interestingness model includes both the reverse academic emotions of interest and boredom. The concentration model is used to determine how focused the learner is on the content at hand by detecting changes in the distance of eyelids. The concentration model includes two inverse academic emotions: exhilaration and fatigue. The happiness model determines how pleasurable the learner is during the learning process by detecting changes in the curvature (i.e. angle) of the corners of the mouth. The happiness model consists of two inverse academic emotions: pleasure and distress.

Defining the range of emotion values for each of the six common academic emotions as [-1,1], it is then possible to map these values into a sphere. The sphere is centred at the origin and has a radius of 1. Because human beings can't have no emotions, no emotions don't conform to normal human emotions. Therefore, the origin of coordinates in the sphere is removed in the emotional space. The three-dimensional academic emotion space is shown in Figure 2.

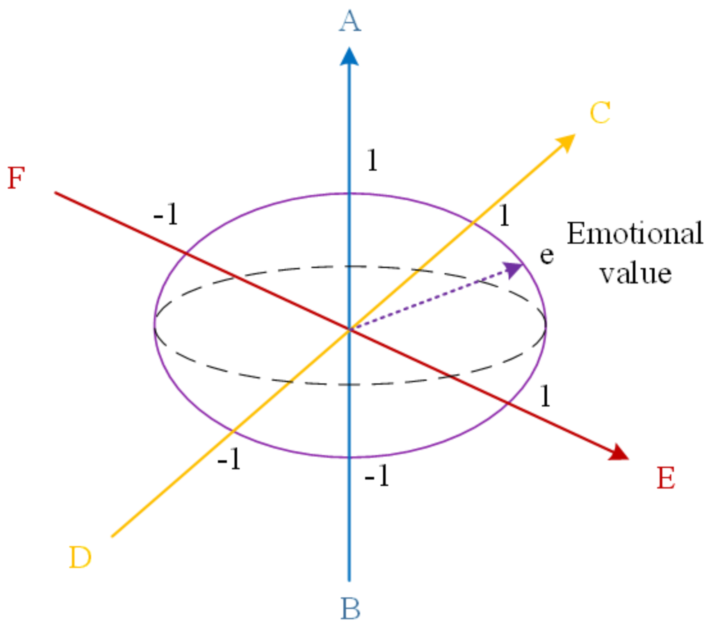


Figure 2. Three-dimensional academic emotion space with reversed emotions

Interest and boredom in the interestingness model are represented by A and B respectively. Exhilaration and fatigue in the concentration model is represented by C and D respectively. Pleasure and distress in the happiness model are represented by E and F respectively. In the emotion space, the origin of the coordinates represents an emotionless state. Since the two emotions in each dimension are each other's inverse emotions, the right-angle coordinate relationship leads to the representation of interest emotion A and boredom emotion B as: $x_A = -x_B$; exhilaration emotion C and fatigue emotion D as: $y_C = -y_D$; pleasure emotion E and distress emotion F as: $z_E = -z_F$. Finally, we can get the calculation method of emotional state in the three-dimensional academic emotional space, $e_s = [x_s, y_s, z_s] = [x_A, y_c, z_E] = [-x_B, -y_D, -z_F]$.

2.4. **Emotional cognitive model construction.** The focus of this paper is on the identification of academic emotions and the construction of an emotion cognitive model.

Based on the aforementioned psychological concepts and affective computational models, this paper combines the special characteristics of e-learning and builds an emotion cognitive model of academic emotions based on facial expression information. The process of expression feature extraction for learners is shown in Figure 3. A video surveillance of

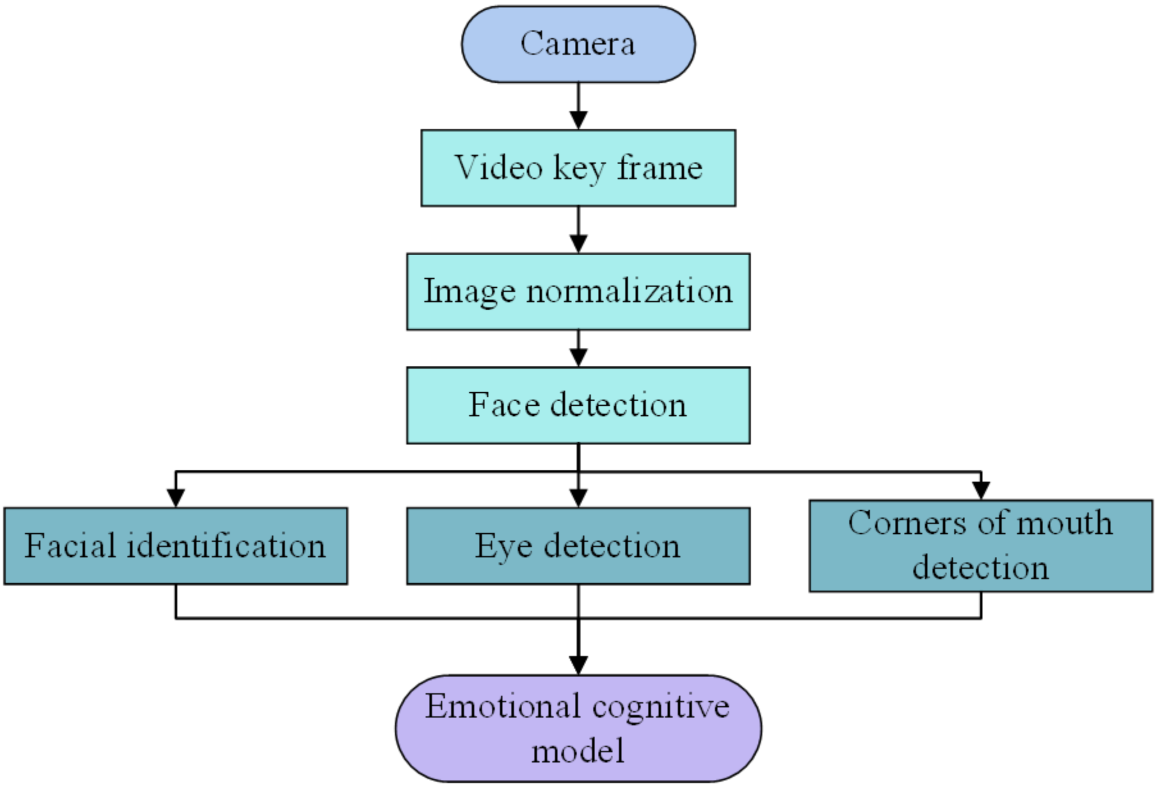


Figure 3. Flow of expression feature extraction for learners

the learner is first performed by means of a camera, thus acquiring a piece of input video, which is then subjected to some processing. Once the face is detected, the facial organs are separated from the skin colour. The features of the eyes and mouth are extracted, which in turn gives the four feature values of face area, eye spacing, pupil spacing and mouth curvature. Finally, these four expression feature values are passed into the emotion perception model.

In the cognitive model of academic emotion, the recognition of emotion is the key, so it is necessary to establish the mapping between the parameters of academic expression features to academic emotion. Firstly, the interestingness, concentration and happiness models were developed to transform from facial expressions to these three dimensions. The values of interestingness, concentration and happiness are then mapped to the three-dimensional academic emotion space. The cognitive model of academic emotion is shown in Figure 4.

3. Cognitive model of academic emotions based on multi-kernel SVM.

3.1. Regional chunking. The key to the cognitive model of academic emotions is to identify the six different academic emotions mentioned above. The focus and difficulty in identifying academic emotions is the need to establish a mapping between the parameters of the learner's facial expression features and the corresponding academic emotions. At

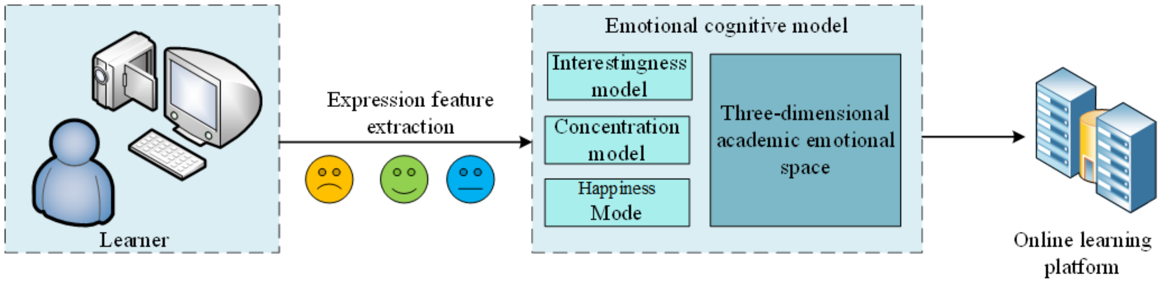


Figure 4. Academic emotional cognitive model

present, there is not much research on the computation of the conversion between expressions to emotions. Some of the more typical methods are curve fitting and neural network methods. The key to the research lies in the design of a good classifier. The recognition of expressions is a non-linear small sample problem. The faces in online web learning videos have features such as high dimensionality and non-linearity. So in order to solve the above problems, this paper uses an improved SVM instead of the two methods mentioned above in order to reduce the difficulty of implementation. Before using the optimised SVM for recognition, the main features of the video need to be extracted using a region chunking strategy for each region of the network video with factors such as motion attributes and texture attributes.

For the input video, the attributes in the different regions are somewhat different. These differences can be better represented through region segmentation. Therefore, the input video of size $M \times N \times T$ is first divided into region chunks of the same size as $K \times L$, noting that the size of all the chunks is fixed. The size of each region chunk is $h \times v$, where $h = M/K$, $v = N/L$, $M \times N$ is the resolution and T is the length of the video sequence.

3.2. **Colour feature extraction for regional chunking.** The colour distribution in the regional chunks of the input video varies significantly due to lighting factors [30], so it is more reliable to use colour features as a basis for keyframe extraction. Let m be the colour component belonging to $\{\mathrm{H}, S, V\}$, then the mean value $\mu_{m,n}$ of the components within the n block is calculated as follow:

$$\mu_{m,n} = \frac{1}{h \times v} \sum_{i=1}^{h \times v} x_{i,m,n} \tag{1}$$

where $x_{i,m,n}$ is the pixel value of the i-th pixel component m in the area block n.

In order to obtain information about the facial expressions in the input video, it is necessary to perform keyframe extraction. Therefore, this process is done based on the colour features of the regional chunks and the main steps are as follows.

Step 1: Select a base frame in the input video i and calculate the distance D_{ij} between it and the subsequent frame j.

$$D_{ij} = \sum_{i=0}^{L} |f_i(l) - f_j(l)| \tag{2}$$

where i and j are the serial numbers of two different frames, L is the dimension of the feature vector, and $f_i(l)$ is the l-th component of the colour feature vector of the i-th frame.

Step 2: When the distance between frames D_{ij} is greater than the set threshold $Thre$, the frame j is replaced with a new base frame.

$$Thre = \frac{1}{N-j-1}\sum_{j=i+1}^{N} D_{ij} \tag{3}$$

Step 3: Determine if the frame j is the last frame, if so end the extraction, otherwise skip to step 1.

3.3. **Facial expression feature extraction.** For facial expression feature parameter extraction, most studies have used the overall image of the key frame as a framework.

However, it is important to note that the way faces move on different region chunks also differs somewhat and facilitates more accurate feature extraction. Therefore, the motion vector fields within the region chunks are calculated based on the above region chunking [31]. After establishing the Cartesian coordinates, the motion vectors of the regional sub-blocks are calculated as follows:

$$\mu_x = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{C} c_{x,i}^t, \quad \sigma_x^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{C}\left(c_{x,i}^t - \mu_x\right)^2 \tag{4}$$

$$\mu_y = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{C} c_{y,i}^t, \quad \sigma_y^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{C}\left(c_{y,i}^t - \mu_y\right) \tag{5}$$

where μ_x and μ_y are the expectation of the motion vector in the regional chunk in the direction of x and y respectively, σ_x and σ_y are the variance of the motion vector in the regional chunk respectively, C indicates the number of macroblocks in each chunk, $c_{x,i}^t$ and $c_{y,i}^t$ are the motion vectors of the i-th macroblock in the regional chunk in the t-th frame.

In order to be more conducive to the task of eye and mouth feature extraction, based on the principle of conditional probability density functions of statistical methods, the five most commonly used texture features are selected to extract the texture attributes of the eyes and mouth.

$$f_1 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \tag{6}$$

$$f_2 = \sum_{i=1}^{N_g}\sum_{j}^{N_g} (i-j)^2 p(i,j) \tag{7}$$

$$f_3 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \times \log(p(i,j)) \tag{8}$$

$$f_4 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p^2(i,j) \tag{9}$$

$$f_5 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{p(i,j)}{1+|i-j|} \tag{10}$$

where $p(i,j)$ is the frequency of the pixel grey level that satisfies the condition (i,j) when constructing the co-occurrence matrix and N_g is the maximum pixel grey level. The expression features extracted in the face recognition process are shown in Figure 5.

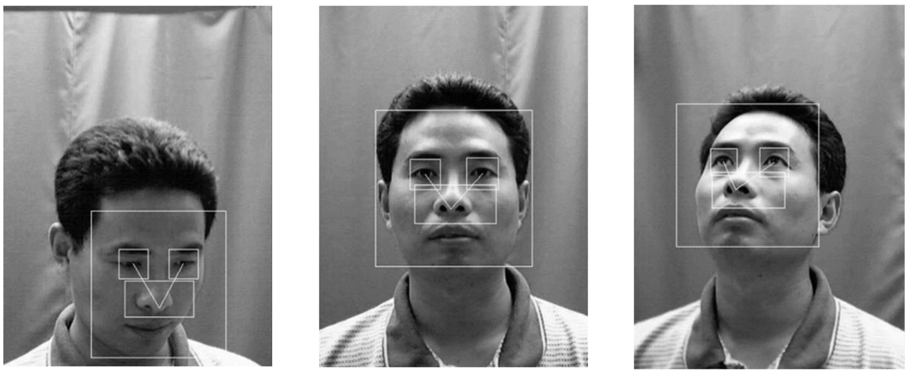


Figure 5. Emotional features

3.4. **SVM optimization and spatial dimensionality reduction.** For the calculation of the transformation between expressions to emotions, this paper uses an SVM with a hybrid kernel function. a typical SVM indicates the existence of a line that separates data mixed in the same space and maximally maintains a large classification interval [32].

The definition of SVM can be expressed as follow:

$$w \cdot x + b = 0 \tag{11}$$

where w is the slope and b is the intercept. For binary classification, the sample set satisfies the conditions shown as follow:

$$y_i(w \cdot x_i + b) - 1 \geq 0, i = 1, 2, \ldots, m \tag{12}$$

In multidimensional spaces, the classification of straight lines becomes the classification of planes. By solving the hyperplane, the binary classification problem can be transformed into a multidimensional spatial classification problem. With the help of Lagrange multipliers, SVM solves for the minimum.

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{13}$$

where α_i is the Lagrangian multiplier corresponding to each sample and $K(x_i, x_j)$ is the kernel function [33].

$$\sum_{i=1}^{n} y_i \alpha_i = 0, \alpha_i \geq 0, i = 1, 2, \ldots, n \tag{14}$$

The classification function of SVM is $f(x)$.

$$f(x) = \mathrm{sgn}\{\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\} \tag{15}$$

A common kernel function in current research on multicore functions is $K(x, y)$.

$$K(x, y) = \exp\left[-\|x - y\|^2 (2\sigma^2)\right] \tag{16}$$

where σ is a constant.

$$K(x, y) = (< x \cdot y > + R)^d, R \geq 0, d \geq 0 \tag{17}$$

where R is a constant, d is a power, and $< x \cdot y >$ represents a polynomial consisting of x and y.

$$K(x, y) = \tanh(g < x \cdot y > + h), g > 0, h < 0 \tag{18}$$

where $\tanh(\cdot)$ is the sigmoid function,g is the coefficient constant and h is the constant.

The mixed kernel function is mainly used because the classification effect of single kernel function is not ideal. Due to the influence of environmental factors such as illumination and occlusion, human faces have high-dimensional and nonlinear characteristics. If a single kernel function is selected in the traditional way, the SVM will not be able to distinguish between complex non-linear cases. In addition, multi-core SVMs require a small number of samples and have better generalisation capabilities.

Then, to further improve the accuracy of expression recognition, data dimensionality reduction is performed in SVM classifiers using LLE. Let the sample x_i can be obtained from its neighbouring samples x_j, x_k and x_l after a linear operation.

$$x_i = \omega_{ij} x_j + \omega_{ik} x_k + \omega_{il} x_l \tag{19}$$

where ω_{ij}, ω_{ik} and ω_{il} are linear coefficients.

In practice, the selection of neighbouring samples for x_i can be multiple. Let the set of k neighbouring samples be Q_i. In order to keep the sample points invariably linear after dimensionality reduction, the objective function needs to be set.

$$\min \sum_{i=1}^{m} \left\| x_i - \sum_{j \in Q_i} \omega_{ij} x_j \right\|^2 \tag{20}$$

LLE is able to keep ω_{ij} constant during the dimensionality reduction. So according to ω_{ij}, the set of samples after dimensionality reduction can be solved.

$$\min \sum_{i=1}^{m} \left\| z_i - \sum_{j \in Q_i} \omega_{ij} z_j \right\|^2 \tag{21}$$

where $\mathbf{Z} = [z_1, z_2, \ldots, z_m]$ and z_i are the reduced dimensional values of x_i. The eigenvectors corresponding to the eigenvalues of ω_{ij} can be solved for by solving for the eigenvectors of $\mathbf{Z}$.

The proposed process of sentiment classification based on regional attribute features and multi-core SVM is shown in Figure 6.

4. **Experimental results and analysis.**

4.1. **Experimental setup.** In order to analyze and verify the video classification method proposed in this paper, specific experiments are conducted. The experimental hardware environment is: Intel Core i7 2.2GHz processor, GTX970@2G video memory and 8G of memory.

Video data from four courses from the MOOC e-learning website (www.cmooc.com) were selected for the experiments, and the video format was adjusted to MPEG-4 and the resolution was adjusted to 352×288. The four courses consist of four datasets, each containing 55, 67, 45, and 72 college students. The parameters of the experimental dataset are shown in Table 1. Each course contains 100 video data, of which 70% are used as training samples and 30% as testing samples.

Table 1. Parameters of the experimental data set

Data sets	Course Type	Number of videos	Video duration
1	University Physics	100	30 minutes
2	Introduction to Physiology	100	30 minutes
3	Computer Programming (C++)	100	45 minutes
4	Fundamentals of Photography	100	45 minutes

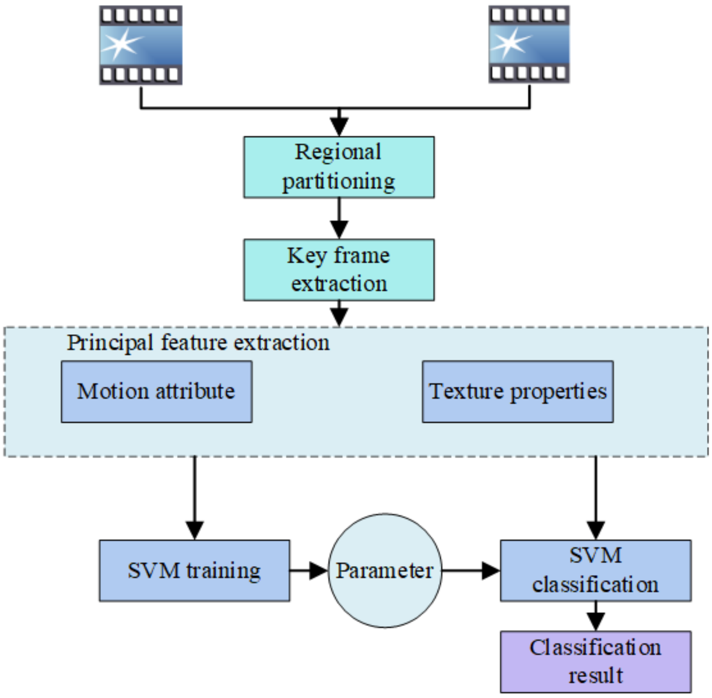


Figure 6. Flow of the proposed emotion classification process

4.2. **Performance verification of the LLE.** First, set different number of neighboring nodes to verify the influence of LLE spatial data dimensionality reduction on detection performance. The results are shown in Table 2. It can be seen that the optimal number

Table 2. Performance of detection under LLE with different number of neighbors

Number of neighbours	Data sets	Accuracy	False alarm rate	Detection time/s
10	1	0.8839	0.0241	47.232
	2	0.8542	0.0449	32.742
	3	0.8722	0.0327	25.358
	4	0.8369	0.0316	21.213
20	1	0.9841	0.0236	59.882
	2	0.9543	0.0438	43.863
	3	0.9725	0.0314	35.668
	4	0.9370	0.0309	29.612
30	1	0.9842	0.0224	77.147
	2	0.9543	0.0431	53.189
	3	0.9727	0.0305	48.227
	4	0.9372	0.0303	37.663
40	1	0.9844	0.0215	93.931
	2	0.9549	0.0424	67.178
	3	0.9733	0.0301	57.556
	4	0.9377	0.0299	44.389

of adjacent nodes in LLE is 20. It can be seen that the detection rate of all four samples is below 90% when the number of neighbours is 10, while the detection rate of the test increases slowly when the number of neighbours reaches 20. This indicates that when the number of neighbouring nodes is small, the dimensionality reduction of the LLE is also smaller. And when the number of nodes reaches 20, the change in the number of neighbouring nodes in the LLE gradually becomes less influential on detection. However, as the number of nodes increases, the detection time increases rapidly because the complexity of the operation of spatial dimensionality reduction increases, and more nodes involved in the linear operation inevitably leads to an increase in the time used.

4.3. **Effect of different kernel functions on SVM performance.** The number of neighbours selected for LLE participation in the calculation was 20, and the test results were compared between the single kernel function and the hybrid kernel function. It can

Table 3. Detection performance of different SVMs

Data sets	Kernel functions	Accuracy	Detection time/s
1	RBF kernel	0.9717	58.991
	Sigmoid nucleus	0.9843	58.893
	RBF kernel + Sigmoid kernel	0.9972	58.601
2	RBF kernel	0.9541	43.863
	Sigmoid nucleus	0.9328	43.842
	RBF kernel + Sigmoid kernel	0.9610	43.625
3	RBF kernel	0.9235	35.572
	Sigmoid nucleus	0.9481	35.634
	RBF kernel + Sigmoid kernel	0.9725	35.102
4	RBF kernel	0.9375	29.594
	Sigmoid nucleus	0.9372	29.602
	RBF kernel + Sigmoid kernel	0.9470	29.318

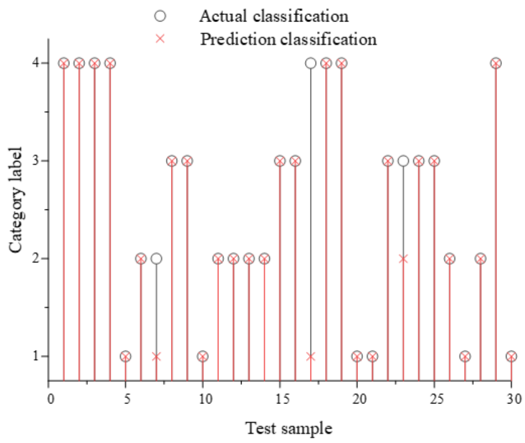
be seen that compared to SVMs based on a single kernel function, SVMs based on hybrid kernel functions show better advantages in terms of both accuracy and detection time.

4.4. **Integrated experiments with the Emotional Cognition Model.** In this paper, the learner's interestingness, concentration and happiness models will be developed and mapped to a three-dimensional academic emotion space with reversed emotions. The features of academic affective state identification are shown in Table 4. The training and

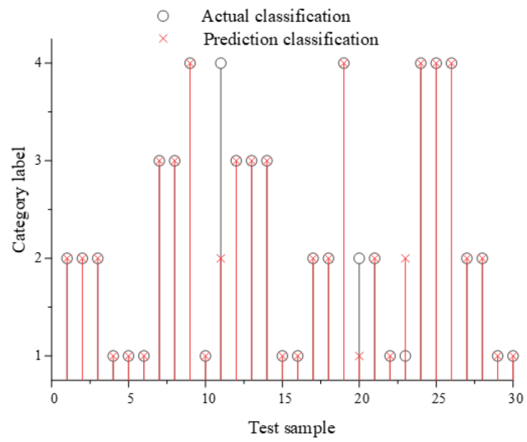
Table 4. Identifying characteristics of academic affective state

Status	Category	Expression recognition feature description
Interestingness	Interested	Body leaning forward (increased face area and pupil spacing)
	Not interested	Body tilted back (face area and pupil spacing reduced)
Concentration	Interested	Eyes open
	Not interested	Eyes closed
Happiness	Interested	Curvature of the mouth becomes larger
	Not interested	Curvature of the mouth becomes smaller

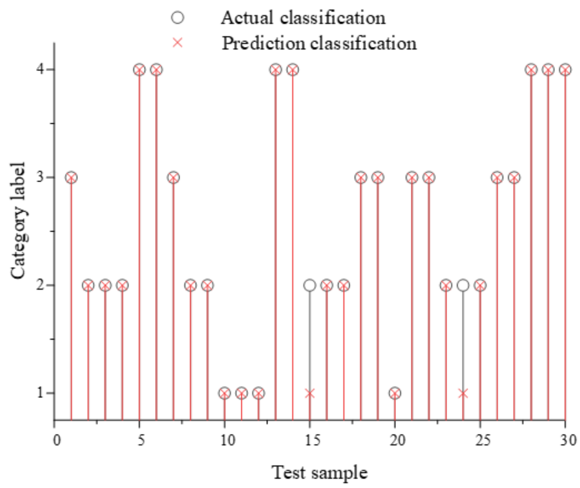
test set data were pre-processed using a normalisation approach. To simplify matters, the emotions in the three states were divided into four categories: very bored, more bored, more interested and very interested. The test results for the three states are shown in Figure 7. From the test results, we can see that the classification accuracy of all three states reached over 94.2%. Basically, the construction of the interestingness model, concentration degree model and happiness model was successfully completed, and the test results were more satisfactory. For comparative analysis, the PSO-SVM [34], GA-SVM



(a) Interestingness model



(b) Concentration model



(c) Happiness model

Figure 7. Test results for the three states

[35] and the proposed method were compared in the same experimental environment. A comparison of the detection performance of the three methods on the video data of the four courses is shown in Table 5. Recall, Precision and F1 were used as evaluation metrics. It is more intuitive to see that the Optimised SVM-based Academic Emotional Cognition

Table 5. Video classification performance comparison (%)

Data	PSO-SVM			GA-SVM			Ours		
sets	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
1	0.68	0.84	0.75	0.88	0.90	0.89	0.92	0.94	0.93
2	0.88	0.80	0.84	0.91	0.89	0.90	0.93	0.92	0.93
3	0.92	0.88	0.90	0.90	0.93	0.91	0.95	0.97	0.96
4	0.90	0.77	0.83	0.90	0.82	0.86	0.93	0.84	0.89
Average	0.84	0.82	0.83	0.89	0.88	0.89	0.93	0.91	0.92

Model performs best on all three assessment metrics. Finally, this paper conducted several comprehensive experiments on the emotional cognition model with 50 different learners. The results of the detected experiments were compared with the current academic emotion states provided by the learners to obtain a detection accuracy rate. The results of the experiments for the detection of abnormal academic emotions are shown in Table 6. As

Table 6. Experimental results of abnormal academic emotion

Unusual academic emotion	**Experimental results**	**Detection accuracy**
Close eyes	45	90%
Not in front of a computer	50	100%
Does not conform to normal study habits	46	92%

can be seen from the above table, the accuracy rate for detecting abnormal academic emotions is high, and the combined accuracy rate can reach over 94

5. **Conclusion.** In order to realise the emotional cognitive interaction function of the online learning system, this paper proposes an academic emotional cognitive model based on optimised SVM. The features of learners' facial expressions will be used as the basis for judging academic emotions, combined with a personalised model to achieve the construction of an emotion cognitive model in a three-dimensional emotion space. To address the high-dimensional, non-linear problem of faces in online e-learning videos, this paper uses an improved SVM to implement the computation of the conversion between expressions to emotions. Before using the optimised SVM for recognition, the main features of the video need to be extracted using a region chunking strategy, with motion attributes, texture attributes and other factors for each region in the web video. Experimental results show that the optimised SVM can achieve high detection performance by reasonably setting the number of LLE neighbours and the kernel function. Compared with other SVMs, the accuracy of the optimised SVM-based emotion perception model for detecting abnormal academic emotions is higher, and the combined accuracy can reach over 94%. In the follow-up studies, SVM and convolutional neural network will be combined to achieve academic emotion recognition.

Acknowledgement. This work is one of the research results of the project of Zhejiang College of Security Technology's 2021 Special Topics of Ideological and Political Education, "Research on the Effectiveness of the 'Post-00' Higher Vocational Students' Discourse of Ideological and Political Education ——Taking Zhejiang College of Security Technology as an Example" (Project No. SZJY2021Z01).

REFERENCES

[1] F. Zhang, T.-Y. Wu, J.-S. Pan, G. Ding, and Z. Li, "Human motion recognition based on SVM in VR art media interaction environment," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, 40, 2019.

[2] F. Zhang, T.-Y. Wu, and G. Zheng, "Video salient region detection model based on wavelet transform and feature comparison," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, 58, 2019.

[3] J. Gao, H. Zou, F. Zhang, and T. Y. Wu, "An intelligent stage light-based actor identification and positioning system," *International Journal of Information and Computer Security*, vol. 18, no. 1/2, 204, 2022.

[4] J. H. Thrall, X. Li, Q. Li, C. Cruz, S. Do, K. Dreyer, and J. Brink, "Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success," *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 504-508, 2018.

[5] E. K. Wang, F. Wang, S. Kumari, J.-H. Yeh, and C.-M. Chen, "Intelligent monitor for typhoon in IoT system of smart city," *The Journal of Supercomputing*, vol. 77, no. 3, pp. 3024-3043, 2020.

[6] J. M.-T. Wu, Q. Teng, S. Huda, Y.-C. Chen, and C.-M. Chen, "A Privacy Frequent Itemsets Mining Framework for Collaboration in IoT Using Federated Learning," *ACM Transactions on Sensor Networks*, vol. 19, no. 2, pp. 1-15, 2023.

[7] C.-M. Chen, S. Lv, J. Ning, and J. M.-T. Wu, "A Genetic Algorithm for the Waitable Time-Varying Multi-Depot Green Vehicle Routing Problem," *Symmetry*, vol. 15, no. 1, 124, 2023.

[8] M. Di Napoli, D. Di Martire, G. Bausilio, D. Calcaterra, P. Confuorto, M. Firpo, G. Pepe, and A. Cevasco, "Rainfall-Induced Shallow Landslide Detachment, Transit and Runout Susceptibility Mapping by Integrating Machine Learning Techniques and GIS-Based Approaches," *Water*, vol. 13, no. 4, 488, 2021.

[9] A. Winckelmann, S. Nowak, S. Richter, S. Recknagel, J. Riedel, J. Vogl, U. Panne, and C. Abad, "High-Resolution Atomic Absorption Spectrometry Combined with Machine Learning Data Processing for Isotope Amount Ratio Analysis of Lithium," *Analytical Chemistry*, vol. 93, no. 29, pp. 10022-10030, 2021.

[10] L. Wang, G. Hu, and T. Zhou, "Semantic Analysis of Learners' Emotional Tendencies on Online MOOC Education," *Sustainability*, vol. 10, no. 6, 1921, 2018.

[11] K. S. Zaikov, A. A. Saburov, A. M. Tamitskiy, and A. S. Nikiforov, "Online Education in the Russian Arctic: Employers' Confidence and Educational Institutions' Readiness," *Sustainability*, vol. 13, no. 12, 6798, 2021.

[12] C. Damşa, M. Langford, D. Uehara, and R. Scherer, "Teachers' agency and online education in times of crisis," *Computers in Human Behavior*, vol. 121, 106793, 2021.

[13] M. W. Bentley, R. Kerr, M. Ginger, and J. Karagoz, "Behavioural change in primary care professionals undertaking online education in dementia care in general practice," *Australian Journal of Primary Health*, vol. 25, no. 3, 244, 2019.

[14] N. Döring, "Sexualbezogene Online-Fortbildung für Fachkräfte: Eine Einführung," *Zeitschrift für Sexualforschung*, vol. 31, no. 01, pp. 97-100, 2018.

[15] J. Chen, Y. Zhang, and Z. Yin, "Education Premium in The Online Peer-To-Peer Lending Marketplace: Evidence from The Big Data in China," *The Singapore Economic Review*, vol. 63, no. 01, pp. 45-64, 2018.

[16] S. McGrew, "Learning to evaluate: An intervention in civic online reasoning," *Computers & Education*, vol. 145, 103711, 2020.

[17] G. Hofree, P. Ruvolo, A. Reinert, M. S. Bartlett, and P. Winkielman, "Behind the Robot's Smiles and Frowns: In Social Context, People Do Not Mirror Android's Expressions but React to Their Informational Value," *Frontiers in Neurorobotics*, vol. 12, 2018.

[18] L. He, D. Jiang, and H. Sahli, "Automatic Depression Analysis Using Dynamic Facial Appearance Descriptor and Dirichlet Process Fisher Encoding," *IEEE Transactions on Multimedia*, vol. 21, no. 6, pp. 1476-1486, 2019.

[19] H. Li, "Design of Multimedia Teaching Platform for Chinese Folk Art Performance Based on Virtual Reality Technology," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 12, no. 09, pp. 28, 2017.

[20] D. Jonauskaite, A. Abu-Akel, N. Dael, D. Oberfeld, and A. M. Abdel-Khalek, "Universal Patterns in Color-Emotion Associations Are Further Shaped by Linguistic and Geographic Proximity," *Psychological Science*, vol. 31, no. 10, pp. 1245-1260, 2020.

[21] X. Song, X. Zhao, L. Fang, and T. Lin, "Discriminative representation combinations for accurate face spoofing detection," *Pattern Recognition*, vol. 85, pp. 220-231, 2019.

[22] L. Zhu, L. Chen, D. Zhao, J. Zhou, and W. Zhang, "Emotion Recognition from Chinese Speech for Smart Affective Services Using a Combination of SVM and DBN," *Sensors*, vol. 17, no. 7, 1694, 2017.

[23] P. Wu, S. Lou, X. Zhang, J. He, and J. Gao, "Novel Quality-Relevant Process Monitoring based on Dynamic Locally Linear Embedding Concurrent Canonical Correlation Analysis," *Industrial & Engineering Chemistry Research*, vol. 59, no. 49, pp. 21439-21457, 2020.

[24] X. Zhang, X. Yu, Y. Liu, Z. Yang, B. Gong, and X. Chen, "Adaptive neighborhood selection based on locally linear embedding for the degradation index construction of traction motor bearing," *Measurement Science and Technology*, vol. 32, no. 11, 115123, 2021.

[25] Z. Cheng, and B. Cai, "Predicting the remaining useful life of rolling element bearings using locally linear fusion regression," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 6, pp. 3735-3746, 2018.

[26] D. Katerndahl, S. K. Burge, R. L. Ferrer, J. Becho, and R. Wood, "Complex Relationship Between Daily Partner Violence and Alcohol Use Among Violent Heterosexual Men," *Journal of Interpersonal Violence*, vol. 36, no. 23-24, pp. 10912-10937, 2020.

[27] M. Shafti, P. J. Taylor, A. Forrester, and D. Pratt, "The Co-occurrence of Self-Harm and Aggression: A Cognitive-Emotional Model of Dual-Harm," *Frontiers in Psychology*, vol. 12, pp. 432-454, 2021.

[28] D. Wang, S. Li, M. Hu, D. Dong, and S. Tao, "Negative Academic Emotion and Psychological Well-being in Chinese Rural-to-Urban Migrant Adolescents: Examining the Moderating Role of Cognitive Reappraisal," *Frontiers in Psychology*, vol. 8, pp. 37-46, 2017.

[29] F. A. Ganotice, J. A. D. Datu, and R. B. King, "Which emotional profiles exhibit the best learning outcomes? A person-centered analysis of students' academic emotions," *School Psychology International*, vol. 37, no. 5, pp. 498-518, 2016.

[30] A. Latif, A. Rasheed, U. Sajid, J. Ahmed, N. Ali, M. Sajid, and T. Khalil, "Content-based image retrieval and feature extraction: a comprehensive review," *Mathematical problems in Engineering*, vol. 2019, 9658350, 2019.

[31] T. Luo, G. Jiang, M. Yu, H. Xu, and W. Gao, "Robust high dynamic range color image watermarking method based on feature map extraction," *Signal Processing*, vol. 155, pp. 83-95, 2019.

[32] M. Khan, M. Q. Reza, A. k. Salhan, and S. P. S. M. A. Sirdeshmukh, "Classification of oils by ECOC based multi-class SVM using spectral analysis of acoustic signals," *Applied Acoustics*, vol. 183, 108273, 2021.

[33] B. T. Atmaja, and M. Akagi, "Two-stage dimensional emotion recognition by fusing predictions of acoustic and text networks using SVM," *Speech Communication*, vol. 126, pp. 9-21, 2021.

[34] R. Li, W. Li, H. Zhang, Y. Zhou, and W. Tian, "On-Line Estimation Method of Lithium-Ion Battery Health Status Based on PSO-SVM," *Frontiers in Energy Research*, vol. 9,693249, 2021.

[35] G. Wang, T. He, Y. Kuang, and Z. Lin, "Optimization of soft-sensing model for ash content prediction of flotation tailings by image features tailings based on GA-SVMR," *Physicochemical Problems of Mineral Processing*, vol. 56, no. 4, pp. 590-598, 2020.