Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 7691586, 12 pages
https://doi.org/10.1155/2022/7691586

Research Article

An Artificial Intelligence-Based Approach to Social Data-Aware Optimization for Enterprise Management

Weiwei Zhang

School of Public Economics and Administration, Shanghai University of Finance and Economics, Shanghai 200433, China

Correspondence should be addressed to Weiwei Zhang; 2016310065@163.sufe.edu.cn

Received 29 June 2022; Revised 20 August 2022; Accepted 2 September 2022; Published 23 September 2022

Academic Editor: Kuruva Lakshmanna

Enterprise management has always been a hot issue in society. In today's society, enterprises are no longer individuals cut off from society and no longer have profit as their sole purpose, but need to exist and develop in combination with social information. With the increasing competition among enterprises, the original enterprise management methods can no longer meet the needs of sustainable development of enterprises, nor can they effectively utilize social data. To better understand the daily emotions of corporate workers, we use a multimodal emotion recognition method in this paper. Multimodal emotion recognition refers to the recognition of human emotional states through different modal information such as speech, visual, and text related to human emotional expressions, which has important research significance in the fields of human-computer interaction, artificial intelligence, and emotional computing and has received much attention from researchers. Given the great success of deep learning methods developed in recent years for various tasks, various deep neural networks are now used to learn high-level representations of emotional features for multimodal emotion recognition. The analysis of employee sentiment is complemented by traditional management methods that make full use of social data. In this paper, based on the study of a single enterprise management model, the proposed model contains five substructure modules, starting with feature inputs, extracting features at different levels through three convolutional modules and outputting recognition results through a softmax classifier. The focus is on how to utilize social data, while combining deep learning with traditional enterprise management methods to fill the research gap in this area in academia.

1. Introduction

The relationship between corporate management and social data was first studied by Moskowitz in the 1970s, and scholars have conducted many exploratory studies since then but still have not reached a unified answer. According to the stakeholder theory and the incentive theory, the enterprise is a contract between stakeholders, and the business activities of the enterprise actually depend on the input of the stakeholders, and they need to provide the material base and environmental protection for the enterprise, so the enterprise has an incentive to perform and disclose the social data in time [1]. In order to maintain the existence of the contract, enterprises want to achieve their own economic interest goals, they must meet the needs of others in exchange; on the other hand, based on the information transfer theory, enterprise performance management can deliver positive information to the relevant parties, thus winning the trust and support of shareholders, consumers, and other stakeholders to the enterprise. The social impact hypothesis, based on this, suggests that the good corporate reputation formed by the former will have a positive impact on corporate performance. The motivation of this paper is that business management has always been a topical issue in society. In today's society, enterprises are no longer individuals cut off from society and no longer have profit as their sole purpose, but need to exist and develop in conjunction with social information. At the same time, with the intensification of competition among enterprises, the original enterprise management methods cannot effectively utilize social data and meet the needs of sustainable development of enterprises.

There is a large body of literature from foreign scholars that empirically tests the positive relationship between the

two. The earliest empirical study was conducted by [2], who found a strong correlation between corporate governance and social data through the definition and measurement of the two concepts; [3] argued that corporate governance can improve risk resistance and corporate reputation; [4] studied more than 200 listed companies in Finland, focusing on the impact of corporate governance environment on corporate performance [5]. Heikkurinen's et al. [6] empirical study found that corporate governance can lead to easier access to investment and better strategic planning, which indirectly demonstrates the positive impact of social data on corporate governance. Some scholars gradually began to study the issue by industry: [7] found that the work environment, work climate, and environmental commitment of financial firms have a positive impact on performance through a study of the financial industry; [8] study covered most industries and firms in Korea and reached a consistent conclusion through an empirical study of seven years of panel data that fulfilling the corresponding regulations would lead to improved firm performance; and some other scholars explore the lag of the relationship, [9] finds through her study that there is a certain lag period in managing the impact of corporate undertakings on performance in the context of financial crisis.

Some scholars have also found that social behavior in some cases can have a negative impact on corporate performance. In terms of enterprise microcomposition, when some managers of enterprises pursue short-term interests, they will make the social fulfillment no longer take the maximization of company's interests and social benefits as the standard; since the realization of social data to enterprise performance is a long-term strategy, ordinary employees are forced to perform the enterprise management actively due to short-term performance pressure; in terms of enterprise macroresource allocation, enterprise resources are scarce and limited. The negative impact of social data on business was first identified by [10], who studied the business environment and found that companies that undertake environmental aspects need to invest in environmental protection equipment, which brings higher business. Lioui et al. [11]also conducted a similar study and reached the same conclusion; [12] and other scholars believe that social data and performance form a mutually constraining relationship, and social data add to corporate costs, increase budgets at the beginning of the period, and bring about a performance burden.

In order to make the study more precise, scholars started to classify corporate performance into long-term and short-term and to study the lags in order to do so. The realization of the role of corporate management on performance is not fully realized in the same period and requires a slower process. On the one hand, the market is not yet ideal, and the asymmetry of information resources makes it difficult for all relevant groups to grasp the multilevel and diversified information of enterprises in detail and accurately; on the other hand, although enterprises can release social signals to obtain external support, it also requires a slow transmission and transformation process [13]. Therefore, it is important to identify the emotions of corporate staff and use them as social information to assist in corporate management. There are two ways to recognize emotions, one is to detect physiological signals (e.g., heart rate, EEG, and body temperature) and the other is to detect emotional behaviors (e.g., facial features, verbal features, and posture). In order of accuracy, the main unmoral modalities currently used for emotion detection are physiological parameters (EEG), facial expressions, speech, and body movements, and in order of difficulty and practicality of acquisition, speech, facial expressions, body movements, and physiological parameters (EEG) [14]. Among them, body movements are usually used as an auxiliary recognition method for other modalities because of their low accuracy and general practicality, while the recognition accuracy of physiological parameters is very high, but they are rarely used in practical scenarios because of the high difficulty and general practicality of acquisition due to the need for professional equipment. The recognition of speech and facial expressions is a popular research method because of its moderate difficulty and high recognition accuracy. Li et al. [15] used the LSTM-RNN network model for sample training and the conditional attention fusion strategy for emotion recognition of face expression and speech to improve the real-time performance of the emotion recognition model [16]. The multimodal fusion recognition can be performed at the signal, feature, and decision levels, and different fusion strategies can be adopted for different modal signals to achieve the best recognition results.

Psychologist Mehrabian [17] found that words reflect 7% of emotion, voice and its characteristics (e.g., intonation and speed of speech) reflect 38% of emotion, and facial expressions and body language reflect 55% of emotion in everyday conversations. This indicates that facial expressions and voice convey the main information in the study of emotion recognition. In this paper, we use a modified convolutional neural network to train and complete the model building for the video image channel, a long- and short-term memory artificial neural network modified by back propagation algorithm to train the training set of speech signals from the video emotion database to build the model, and fuse the recognition results at the decision level to output the emotion classification and the possibility on different emotion classifications [18]. In addition to verifying the effectiveness of the proposed method, this paper also implements the real-time analysis of enterprise employees' emotion: by calling the camera and microphone to capture a video and speech, using LBPH algorithm to identify and target the face region, and then analyzing the user's emotional state by SAE+CNN neural network model to complete the recognition of the image channel, using Spleeter and FFmpeg separation tools. After the preprocessing of filtering and windowing of speech signals, the acoustic features are extracted and classified by the openSMILE tool to complete the recognition of speech modalities, and finally the classification results of the two modalities are fused in the decision layer and the final results are output [19]. Through the identification of enterprise employees' emotions, the existence of larger enterprise management and social data are combined to form a study of the relationship between the two, which can combine important factors affecting the development of enterprises and help to establish

an inclusive, innovative, and diversified development of corporate culture and atmosphere, stimulate the R&D efficiency of employees, develop long-term technological innovation strategies, and form a harmonious and sustainable economic development environment and social environment.

The main contributions of this paper are as follows: business management has always been a hot issue in society. In today's society, enterprises are no longer individuals cut off from society and no longer have profit as their sole purpose, but need to exist and develop in conjunction with social information. In order to better understand the daily emotions of enterprise employees, we adopt a multimodal emotion recognition method in this paper. Multimodal emotion recognition refers to the recognition of human emotional states through different modal information such as speech, visual, and text related to human emotional expressions and has received attention from researchers as it has important research significance in the fields of human-computer interaction, artificial intelligence, and emotional computing. This paper focuses on how to utilize social data based on the study of a single enterprise management model, while combining deep learning with traditional enterprise management methods to fill the research gap in this area in academia.

2. Related Works

2.1. Current Status of Social Research on Business Management. From the point of view of the incentive theory, companies also exist in the society and act on the environment as people do and need to respond to the surrounding environment. The quality and efficiency of the contract can be improved to a large extent by the social performance of the company to meet the needs of the stakeholders in order to achieve the equilibrium in the contract and to obtain resources from the stakeholders. From the perspective of the game theory, carrying out social activities is essentially an act of continuously creating long-term value [20]. Fulfilling social responsibility not only satisfies the needs of relevant people in the internal and external environment and creates corresponding social benefits but also satisfies the enterprise's own development needs. By taking responsibility for different internal and external related groups, the company can accordingly improve its solvency, visibility, profitability, productivity, and social reputation, which ultimately contribute to the performance of the company. Yin et al. [21] used a sample of nearly nine hundred A-share listed companies with their two-year panel data and found a positive relationship between the two; [22] reached the same conclusion through a study of food and beverage companies listed in Shanghai and Shenzhen.

From the existing literature, it seems that the theoretical relationship between corporate management and social data is shown in Figure 1. Most scholars, after studying it, believe that the former has a positive effect on the latter, some find that the former has a negative effect on the latter, and the remainder finds that there is no association between the two. According to economist [23], innovation in socioeconomic growth and development can bring about a new allocation of production factors and production conditions, establish a new production function, achieve a change in resource allocation mode, and gradually realize a highly efficient production method [24]. Bar et al. [25] focused on manufacturing industry, exploring the relationship between technological innovation and firm performance by introducing a production function or a case study. The conclusion is that R&D investment in manufacturing industry has a positive effect on performance, and the return on R&D investment in high-tech manufacturing industry is higher than that in low-tech enterprises. In addition, [26] also reached the same conclusion by studying various industries, where [27] found that the effect of R&D investment on market share is more significant for multinational firms than for SMEs. Some scholars have studied the lag period of R&D investment considering that it takes some time to turn into performance [28]. As early as the end of the last century, Chambers chose high-tech enterprises as a sample and after an empirical study; they concluded that technological innovation investment in high-tech industries has a lagged positive effect on firm performance. The results show that firms that engage in technological innovation are positively influenced by their R&D investment over the next five years.

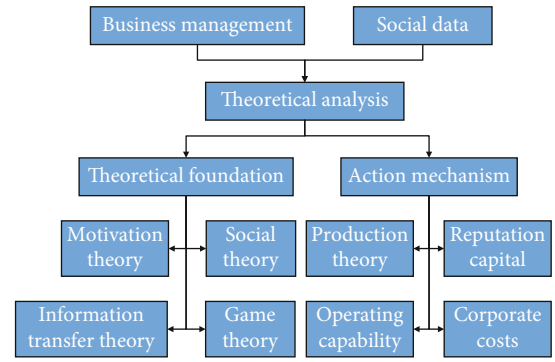

Figure 1: Theoretical analysis chart.

Along with the increasing depth of the research, scholars have started to divide the research by the industry in which the firms are located. Wang et al. [29]argues that high R&D intensity will bring better performance for firms compared to low R&D intensity, where Chun suggests that the relationship between the two is not evident on a sample of large SOEs [31]. Huang et al. [32] conducted a comparative analysis of five industries in China, and the analysis yielded similar conclusions, and the higher the knowledge intensity, the greater the degree of impact; Lu proposed that R&D investment has a positive impact on business performance through a research analysis conducted on a sample of high-tech industries and manufacturing industries. Through R&D, companies are able to introduce new products and technologies to consumers and the market, and these new products and technologies give them an advantage over other competitors in the market, which can greatly increase their sales and revenues and improve their performance. However, research and development is a long-term task, which is constrained by technical theories, costs and the quality of researchers, etc. It takes a long time from the

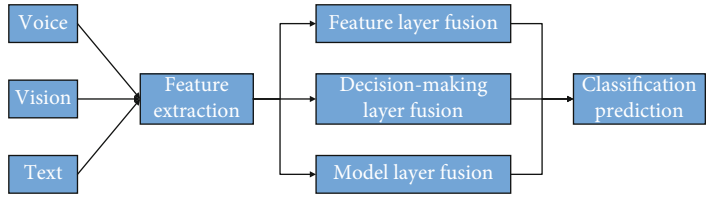

Figure 2: Multimodal emotion recognition framework.

establishment of a project, successful development to commercialization. Moreover, even if R&D is successful, the improvement of enterprise performance is not likely to be immediate. Wang et al. [33] conducted a study on high-tech industries in China, which showed that the current investment in technological innovation has a negative effect on enterprise performance, and it is necessary to control the R&D investment within a certain amount in order to have a positive impact on the performance in the lagged period; there is a lag in the impact of [34] R&D investment on performance, and [35] found that the lag period is three years after the study [36]. Wang used three years of data from manufacturing listed companies as the research object and found that R&D intensity has a positive impact on corporate performance and still has a positive impact on performance in the lag period.

Some scholars found no significant relationship between technology R&D investment and financial performance or an inverted U-shape. Ocean Zhang et al. [37] studied 3 years of data from 34 industrial classifications in China and found that the amount of research expenditures did not affect firm productivity and future performance; [38] found after their study that the number of researchers was not related to firm profitability; similarly, [39] found an insignificant relationship after their study of new energy listed firms [40]. Zhang et al. [41] conducted a study on China's machinery manufacturing industry and private enterprises and concluded that R&D expenditures and corporate financial performance show an inverted U-shape: [42] believes that technological innovation has a positive impact on corporate performance only when the investment in technology is moderate, while [43] believes that the most significant impact of R&D investment on financial performance is in the third quarter after R&D investment, i.e., an inverted U-shape in terms of time period. After combing through the existing literature, it can be seen that scholars, when analyzing the relationship between technological innovation investment and enterprise performance, choose similar indicators for technological innovation expenditure, and the indicators chosen are mainly the amount of R&D expenditure invested and the number of patents obtained, but the selection of indicators for enterprise financial performance is more different, and the different dimensions of specific sample industries, periods, and countries selected by scholars also make scholars reach different conclusions.

2.2. Current Status of Multimodal Emotion Recognition. Multimodal emotion recognition has considerable promise for applications in social robotics, educational quality assessment, security control, human-computer interaction systems, etc. To promote the development of emotion recognition tasks, different multimodal emotion task challenges have emerged in the last decade, including AVEC, EmotiW, MuSE, and MEC. A general multimodal emotion recognition framework is given in Figure 2. As shown in Figure 2, a general multimodal emotion recognition system consists of three steps: feature extraction, multimodal information fusion, and emotion classifier design [44]. Feature extraction is to extract feature parameters related to emotion expression for different modal information such as speech, visual, and text, respectively. Multimodal information fusion refers to the fusion of two or more unmoral information using different fusion strategies.

In recent years, deep learning techniques have been widely used in speech emotion recognition tasks for deep speech emotion feature extraction. Common deep learning methods used for speech emotion recognition are CNN, DBN, RNN, etc. Dutta proposed a speech recognition model based on linear predictive coding and MFCC. LPC and MFCC features are extracted by two different RNN networks for recognizing Assamese. Mao proposed to apply CNN to feature extraction for speech emotion recognition [45]. Chen et al. [45]proposes a new multigranularity feature extraction method [46]. The method is based on different time units, including short-time frame granularity, medium-time granularity, and long-time window granularity. To fuse these multigranularity features, a feed-back neural network based on cognitive mechanisms is proposed. Cirnn combines different temporal-level features to simulate the step-by-step processing of audio signals by humans and achieves multilevel information fusion by highlighting the role of both temporal sequences of emotions and content information. Yu et al. [47] proposed a feature extraction method for the original speech signal, using the SincNet filter to extract some important narrowband emotional features from the original speech waveform, and then using the encoder of the transformer model to extract deep features containing global contextual information. Zhang et al. [48] uses DBN to perform unsupervised feature learning on the extracted low-order acoustic features and initializes a multilayer perception based on the learning results of the DBN implicit layer, which is used for Chinese speech emotion classification. Structures for image recognition pretraining; in addition, 6373-dimensional manual feature representations are extracted using the openSMILE tool, including speech quality features such as jitter and shimmer, as well as spectral, MFCC, and low-level descriptors associated with vocalization. Finally, early and late fusion of deep features and manual features was performed [49]. From the existing literature

on manual and deep speech emotion features mentioned above, (1) the extraction of higher dimensional LLD features using the openSMILE tool has become the mainstream approach for manual speech emotion features. (2) Using CNN to directly extract high level speech emotion features from the original speech signal has become the mainstream method for deep speech emotion features. (3) Manual speech emotion features and deep speech emotion features have their own advantages and disadvantages. The fusion of these two features for speech emotion recognition has been a meaningful research direction in recent years.

Although traditional face recognition methods have achieved remarkable success by extracting manual features, in recent years deep learning methods are gradually applied to emotion recognition for extracting advanced features due to their highly automatic recognition capabilities. (1) Static facial images: for deep feature extraction of static facial images, some model frameworks based on convolutional neural networks are mainly used. Yolcu et al. [50] proposes a method to detect important parts of the face using three structurally identical CNNs, each of which detects a part of the face, such as the eyebrows, eyes, and mouth. Before the images are introduced into the CNNs, cropping and detection of facial key points are performed, and the iconic faces obtained by combining the original images are introduced into a second class of CNNs to detect facial expressions [51]. The researchers show that this method is more accurate than using the original image or imaged face alone. The results show that the method can effectively improve the performance of face expression recognition in static images. Zhang et al. [52] proposes a face expression recognition method based on a multistage feature attention mechanism, which uses two convolutional layers to extract shallow feature information. Secondly, a null convolution is added in parallel to the inception structure for extracting multiscale features, and then a channel attention mechanism is introduced to enhance the model's utilization of useful feature information. (2) Dynamic video expression sequences: for deep feature extraction of dynamic video expression sequences, commonly used methods include CNN, RNN, and LSTM [53]. Kim studied the changes in facial expressions under emotional states and they proposed a framework that combines CNN and LSTM. The features of facial expressions are encoded in two parts. In the first part, CNN learns the spatial features of facial expressions in all frames of emotional states; in the second part, LSTM is used to learn temporal features. Pan et al. [54] proposes a deep spatiotemporal network-based facial expression recognition method for video. Firstly, spatial convolutional neural network and temporal convolutional neural network are used to extract high-level spatiotemporal features in video sequences, respectively. Then the combined extracted spatial and temporal features are input to the fusion network for the video-based facial expression classification task [53]. From the above existing literature on manual visual emotion features and deep visual emotion features: (1) vision-based emotion recognition can be divided into expression recognition based on static facial images and expression recognition based on dynamic video sequences. (2) For the manual feature extraction of static facial images, the facial expression features are mainly obtained by extracting the geometric and appearance features in the image information, and the commonly used methods include LBP, HOG, SIFT, and their improved methods; for the depth features of static facial images, the CNN-based network model is mainly used for the depth feature extraction of facial images; for manual feature extraction of dynamic video expression sequences, capturing the dynamic information of video sequences in order to represent the useful information of facial expressions more effectively, the commonly used methods mainly include the optical flow method and the model method; for depth feature extraction of dynamic video expression sequences, considering the spatiotemporal nature of video sequences, CNN- and RNN-based models are usually used to extract spatial depth features and temporal depth features, respectively.

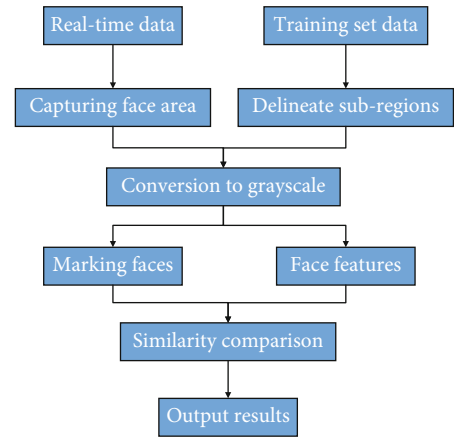

Figure 3: Flow chart of LBPH algorithm.

Driven by the traditional word embedding, OpenAI proposes the transformer-based language model GPT. Unlike ELMo, GPT uses the above to predict the next word. GPT uses a two-stage process, first learning the initial parameters of the neural network model using language modeling goals on unlabeled data and subsequently adapting these parameters to the target task using the corresponding supervised goals. GPT achieved previous state-of-the-art results on many sentence-level tasks of the GLUE benchmark test. Xu et al. [55] proposed Emo2Vec, which encodes sentiment semantics as a word-level representation of fixed-size real-valued vectors, and used a multitask learning approach to train Emo2Vec on six different emotion-related tasks [56]. From the existing literature on manual text sentiment features and deep text sentiment features mentioned above, (1) the commonly used manual text sentiment feature extraction uses the bag-of-words model BoW, but it suffers from high-dimensional sparse and missing interword relationships, which is a low-level text feature representation. In order to improve the BoW model, a series of improved models such as LSA, PLSA, and LDA have emerged successively. (2) Deep text sentiment features are mainly represented in the form of word embeddings, and some pretrained word embedding models for deep learning are widely used in text sentiment extraction tasks, which are mainly divided into typical word embeddings

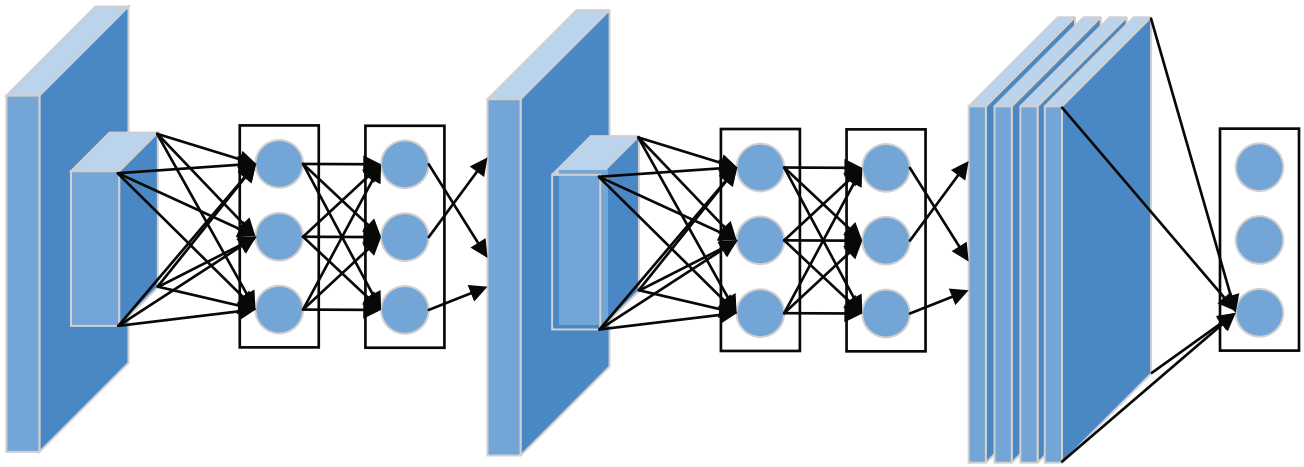

Figure 4: Schematic diagram of GAP working principle.

and sentiment word embeddings. The commonly used word embeddings are word2vec, GloVe, BERT, etc.

3. Algorithm Design

3.1. Video Image Modal Design. The local binary method was proposed in 1996 by Ojala to define the LBP operator in a 3×3 neighborhood of pixels, where the center pixel of the neighborhood is used as the threshold and the grayscale values of the eight adjacent pixels are compared with the center, and the position of the pixel is marked as 1 if it is greater than the center pixel value, and 0 otherwise. When the scale of the image changes, the LBP feature coding will make errors in reflecting the texture information around the pixel. In view of this situation, this paper uses extended LBP features, and the improved method uses circular, expandable neighborhoods [57]. When the scale of the image changes, the LBP feature coding makes an error in reflecting the texture information around the pixel point. Ahonen et al. [58] proposes the LBPH method, which divides the LBP feature image into local blocks and extracts histograms, and then connects these histograms in turn to form a statistical histogram called LBPH. The LBPH algorithm used in this paper adds the function of acquiring face feature data in real time, and its flow is shown in Figure 3.

The edge information of human face expressions has rich emotional features, and this paper incorporates a sparse autoencoder to obtain the emotional details of the images. The main idea of SAE is to impose a sparse constraint on the hidden layer to force the number of hidden nodes to be smaller than the input nodes, so that the network can learn the key features of the image. The SAE network pursues the output data to be approximately equal to the input data, and the network cost function is calculated by back propagation to train the model. The specific implementation process of sparse autoencoder is to first calculate the average activity of hidden neurons as shown

$$\widehat{\rho}_j = \frac{1}{n}\sum_{i=1}^{n} x_i \alpha_j^2. \tag{1}$$

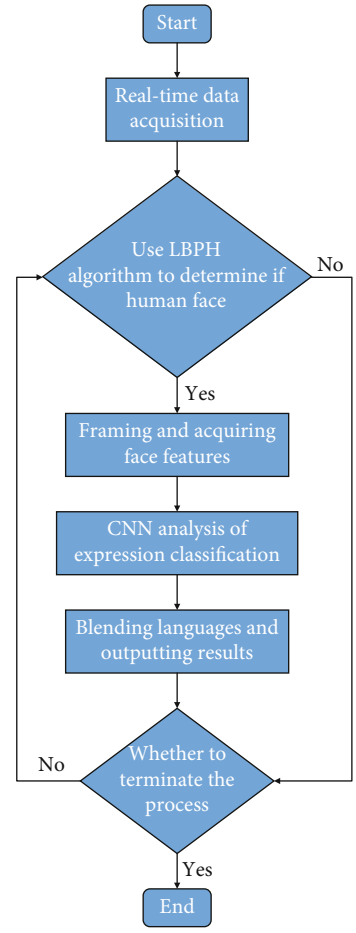

Figure 5: Video image modal workflow diagram.

Then, the parameters of the SAE network are trained to minimize the total cost function so that detailed features of the input image can be captured. A convolutional neural network is a feed-forward network that includes convolutional computation and has a deep structure with local connections and shared weights among neurons. It consists of a convolutional layer, a pooling layer, a fully connected

layer, and an output layer. Increasing the depth of the neural network model results in more features, but when too many features are obtained, it consumes more time and is prone to overfitting because of the connection to each feature in the fully connected layer. In order to overcome this problem, this paper uses the Global Average Pooling (GAP) layer instead of the fully connected layer, which is a summation of spatial information and is more robust to spatial variations. The GAP from Figure 4 includes 3 different subnetwork structures, each consisting of a convolutional layer and multiple fully connected layers.

In addition, in order to reduce the computation of parameters, the convolution operation used in this paper is deep separable convolution. The first five convolutional layers of the CNN are convolved twice and normalized, then pooled and connected to the next layer, with filtering times from 8 to 128. The last convolutional layer performs one convolution and connects to the GAP layer with filter number 1, and then enters the output layer to obtain the classification results. The global convolutional kernel is 3×3 and the ReLU activation function is chosen; the pooling method is maximum pooling, the GAP layer is used instead of the fully connected layer, and the output layer uses softmax to do the classification of expressions. The video image channel workflow is shown in Figure 5.

3.2. Speech Modal Design. In this paper, we choose to use a combination of FFmpeg and Spleeter audio separation tools, where Spleeter can extract the sound signal from the video captured by the camera, and FFmpeg can further process the audio to distinguish the human voice from the background music. Both tools can be called using the python toolkit. The speech signal is a time-varying signal and its feature parameters are constantly changing, but from a microscopic point of view, the features can be kept in a stable state on a short time scale, and these short speech fragments become frames, which are usually 10 ms to 30 ms long. In this paper, we use traditional features (such as rhyme features, sound quality features, spectral features, and Mel frequency campestral coefficients) to achieve good recognition results in the experiments, but the speech signal is unstable, and the recognition effect is limited by using only these traditional features. Therefore, this paper selects rhythmic features, Mel campestral coefficients, and introduces nonlinear attributes and nonlinear geometric features in the feature layer for fusion. The specific is implemented with the depth-constrained Boltzmann machine.

DBM is a type of restricted Boltzmann machine, and RBM consists of a visual layer and a hidden layer. Multiple RBMs are stacked from bottom to top, and the output of the lower layer becomes the input of the upper layer to form the DBM, thus obtaining a deeper representation of the input features. In this paper, a three-layer RBM is used to form a DBM, where the energy function is

$$E\left(v, h^{(1)}, h^{(2)}, h^{(3)}; \theta\right) = -v^T W^{(1)} h^{(1)} - h^{(1)T} W^{(2)} h^{(2)} - h^{(2)T} W^{(3)} h^{(3)}. \tag{2}$$

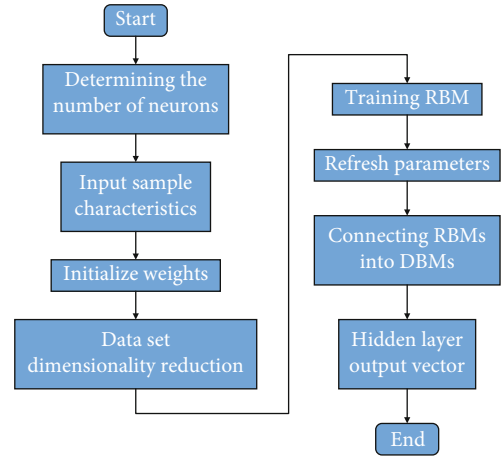

Figure 6: DBM training process diagram.

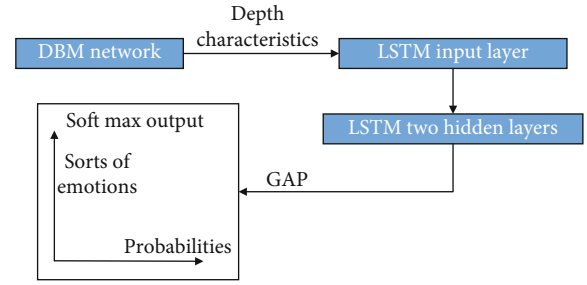

Figure 7: Structure of language channel neural network.

Table 1: Experimental environment.

Name	Versions
Python	3.7
TensorFlow-gpu	2.0.0rc0
CUDA	10.0
cuDNN	v7.5.0
OpenCV-Python	4.4.0.46
Keras	2.3.1

Table 2: CHEAVD2.0 data set.

Classification	Training set	Validation set	Test set
Natural	1400	200	400
Angry	884	128	252
Happy	828	119	236
Sad	462	67	132
Scared	1024	147	293
Surprised	175	25	51
Disgust	144	21	42

Its loss function is

$$L(W, a, b) = -\sum_{i=1}^{m} \ln\left(P\left(v^{(i)}\right)\right). \tag{3}$$

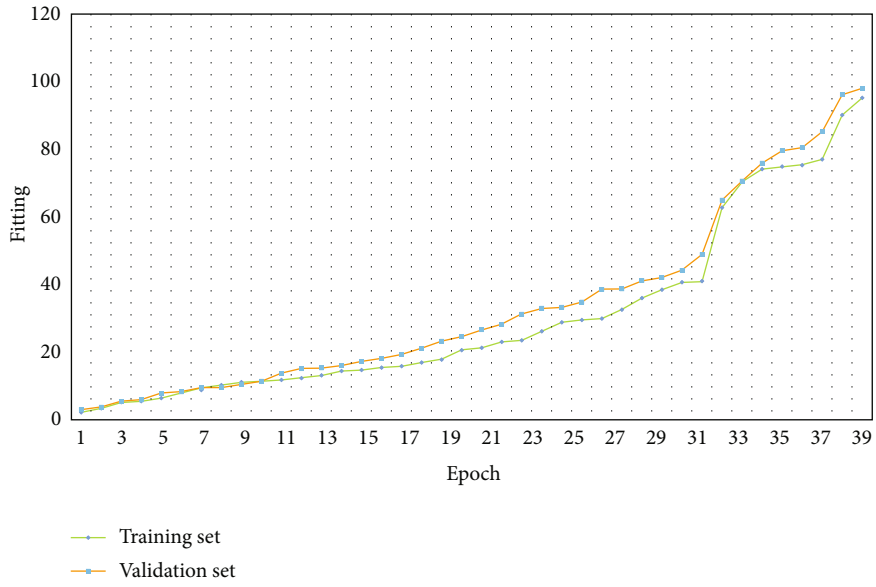

Figure 8: Schematic diagram of training process performance improvement.

After inputting the samples into the RBM, the output feature vector is composed based on the activation probability and expectation of each neuron in the hidden layer. The training process is shown in Figure 6.

A three-layer DBM network is built, and the selected four types of features are fused in the DBM to obtain the deep fused features. Each DBM layer is composed of three RBM layers. First, the features are input to DBM1 layer for deep fusion and dimensionality reduction, and the hidden layer output features 1, 2, 3, and 4; features 1, 2, 3, and 4 are linearly spliced and input to DBM2 layer, and features 5 and 6 are obtained after deep fusion and dimensionality reduction; the process is repeated, and features 5 and 6 become fused features in DBM3 layer, which is the deep representation of the input features.

After the fused features are obtained using the DBM network, the speech emotion needs to be classified [59]. In this paper, we use a modified long short-term memory network, LSTM, which can store useful information over a long period of time and optimize the classification task for time series and has better performance than traditional models (temporal recurrent neural networks, hidden Markov models, etc.) for speech recognition applications. The advantage of LSTM is that the output of the current moment is influenced by the input and the output of the previous moment and can take into account the temporal characteristics of the features. The DBM and LSTM networks are optimized using a backpropagation algorithm with variable weights [45, 60–62]. The addition of BP to the network for the language channel increases the nonlinear mapping capability of the network for processing the acquired nonlinear features. BP uses gradient descent to adjust the internode weights ω_{ij} and node b thresholds, and the functional is

$$\omega_{ij} = \omega_{ij} - \eta_1 \times \frac{\partial E(\omega, b)}{\partial \omega_{ij}}, b_j = b_j - \eta_2 \times \frac{\partial E(\omega, b)}{\partial b_j}. \quad (4)$$

In order to avoid overfitting and improve the processing speed, the fully connected layer is replaced by the GAP layer and finally connected to the softmax layer [63–66]. The input is the fused features processed by the DBM layer, and the output is the classification and probability of emotional affiliation by the softmax layer. The structure of the language channel neural network is shown in Figure 7.

3.3. Experimental Results and Analysis

3.3.1. Experimental Dataset. In this paper, the data sources mainly include internal nonpublic data from a domestic Chinese economic development and business management research institute. fer2013 image dataset and CHEAVD2.0 video dataset are used for the experiments. fer2013 consists of 34658 face expression images, which is the most extensive face expression database covering different countries and ages, with a large number of samples and preprocessed, which is of higher quality compared with the images taken from CHEAVD2.0 video. The CHEAVD2.0 speech dataset consists of 7,156 emotional video clips from movies and variety shows, covering a large amount of data and close to the real environment, with an average length of 3.3 s. The emotion labels are natural, angry, happy, and sad. The average length is 3.3 s. The emotion labels are natural, angry, happy, sad, worried, anxious, surprised, and disgusted. The two databases are very similar in terms of emotion classification, and worry and anxiety are categorized as worry in the preliminary data processing, so that the two databases are consistent in terms of emotion classification for integration at the decision level. The experimental environment is shown in Table 1.

The composition of the processed CHEAVD2.0 data is shown in Table 2. The training process performance enhancement and loss convergence are shown in Figures 8 and 9.

3.3.2. Simulation Experiments. Table 3 shows the comparison of the recognition effects of the unmoral improvement

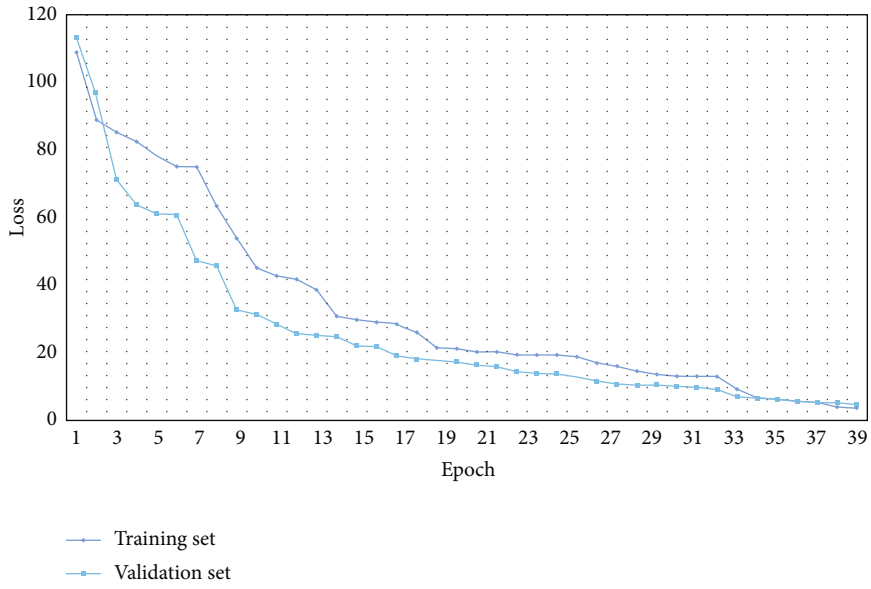

Figure 9: The training process loss convergence schematic.

Table 3: Comparison of identification results on single mode.

Modal	Feature selection	Method selection	Recognition accuracy
Voice	MFCC	SVM	83.2%
	GFCC	LSTM	87.5%
	Depth (ours)	DBM (ours)	91.3%
Image	Convolution	67	73.6%
	VGG	147	72.8%
	Integration (ours)	21	76.4%

Table 4: Comparison of single-mode and multimode recognition results.

Modal	Feature selection	Method selection	Recognition accuracy
Voice	LBPH	CNN	66.8%
	Integration (ours)	SAE+CNN (ours)	73.5%
Image	MFCC	LSTM	58.4%
	Depth (ours)	DBM+LSTM (ours)	61.7%
Voice+image	Multimodalities (ours)	Integration (ours)	75.3%

algorithm compared with other algorithms for the speech channel and the video image channel. The comparison experiments are conducted using the Berlin Database of Emotional Speech (EMO-DB) for the speech channel and the fer2013 data set for the video image channel. In the comparison of image modality, the recognition accuracy of this method is only slightly lower than that of VGGNet+Focal Loss method, which also achieves better recognition results. It can be seen that the improved CNN and LSTM proposed in this paper can perform effective recognition in unmoral mode. The accuracy of the proposed method in this paper reaches 91.3% and 76.4% in sound and image recognition,

Table 5: CHEAVD2.0 test set all kinds of emotion recognition accuracy statistical table.

Classification	Number of samples	Number of identification	Recognition accuracy
Natural	400	285	71.3%
Angry	252	174	69.0%
Happy	236	188	79.7%
Sad	132	112	84.8%
Scared	293	216	73.7%
Surprised	51	32	62.7%
Disgust	42	27	64.3%

Table 6: Confusion matrix for the actual test.

	Natural	Angry	Happy	Sad	Scared	Surprised	Disgust
Natural	96.1%	0	2.4%	1.3%	0	0	0.2%
Angry	1.8%	95.8%	0	0	2.4%	0	0
Happy	0	0	97.4%	0	0	2.6%	0
Sad	0	1.7%	0	96.5%	0	0	1.8%
Scared	2.1%	0	0.5%	0	96.2%	0	0
Surprised	0	0	0	0	1.6%	95.4%	0
Disgust	0	2.5%	0	2.2%	0	0	95.4%

respectively, which is substantially ahead of the presently available methods.

In this paper, the recognition effect of multichannel fusion is verified with the test set of CHEAVD2.0. As shown in Table 4, the recognition accuracy of image channel can be improved after using SAE, the recognition accuracy of language channel can be improved after DBM fusion of features, and the recognition accuracy of multimodal fusion is higher. This shows that the multimodal fusion recognition strategy can achieve better recognition results.

Its recognition accuracy in various types of emotions is shown in Table 5, which shows that it can achieve good results in the recognition of natural, happy, angry, and sad emotions, and fewer samples are assigned to wrong emotion types, among which more samples are assigned to natural emotion types by mistake. The largest number of samples was misclassified as natural and angry. The overall recognition accuracy reaches 72%, which is an improvement compared with the traditional unmoral recognition accuracy.

3.3.3. Practical Application. After validating our method on the dataset, we apply the method to actual enterprise management. In-enterprise monitoring is used to obtain staff social data in real time and to identify their emotions. A 10-fold cross-validation is used, and the experimental results are averaged from the 10 cross-validation results. In order to evaluate the effectiveness of the proposed method, a baseline is provided for each dataset, and the results of the baseline are the results of the CNN network extracted from the expression features directly used for expression classification. When using the CNN network for expression classification, a fully-connected layer with 6-dimensional or 7-dimensional (depending on the number of expression categories in the dataset) values is added at the end of the CNN network, and cross-entropy is used as the loss function for expression classification. Table 6 shows the confusion matrix of various expressions recognized in the practical application of this paper. The values in the diagonal line of the table correspond to the recognition accuracy of each expression, and the other values are the expression recognition error rate. The confusion matrix shows that the recognition rate of all expressions exceeds 95%, which has a high accuracy rate.

4. Conclusions

About enterprise management society, many scholars have conducted empirical studies based on theoretical exploration. Research on the relationship between business management and social data has gradually become clearer, and several articles have emerged that directly investigate the impact of social data and technological innovation on business management. These articles conclude that the effective use of social data enhances the positive effect of corporate performance in higher-technology firms, while lower-technology firms counteract the positive effect of social data on market value and thus reduce corporate performance. To this end, we propose a sentiment recognition method that makes full use of social data to assist business management. To improve the accuracy and real-time of emotion recognition, video image modality is implemented based on local binary histogram method, sparse autoencoder, improved convolutional neural network; speech modality is implemented based on the improved depth-constrained Boltzmann machine and improved long-short time memory network; more detailed features of images are obtained using SAE, and deeper expression of voice features are obtained using DBM. The experimental results show that the fusion recognition strategy improves the recognition accuracy and has good recognition results in both the dataset and the actual testing process. By applying our method to enterprise management, we can enable enterprises to make full use of social data, which has good effect on the improvement of enterprise performance and also provides motivation for enterprise management to carry out technological innovation. In the future, we plan to carry out a knowledge graph-based framework for social data sensing and fusion for enterprise management.

Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The author declares that he has no conflict of interest.

References

[1] Y. Hong and M. L. Andersen, "The relationship between corporate social responsibility and earnings management: an exploratory study," *Journal of Business Ethics*, vol. 104, no. 4, pp. 461–471, 2011.

[2] J. McGuire, S. Dow, and B. Ibrahim, "All in the family? Social performance and corporate governance in the family firm," *Journal of Business Research*, vol. 65, no. 11, pp. 1643–1650, 2012.

[3] M. Hansel and K. Hammond, "Seven corporate governance lessons from David Jones," *Governance Directions*, vol. 66, no. 3, pp. 166–168, 2014.

[4] R. Russo, "Risk management in taxation," in *Risk management and corporate governance*, Edward Elgar Publishing, 2010.

[5] E. Gras-Gil, M. P. Manzano, and J. H. Fernández, "Investigating the relationship between corporate social responsibility and earnings management: evidence from Spain," *BRQ Business Research Quarterly*, vol. 19, no. 4, pp. 289–299, 2016.

[6] P. Heikkurinen and J. Mäkinen, "Synthesising corporate responsibility on organisational and societal levels of analysis: An integrative perspective," *Journal of Business Ethics*, vol. 149, no. 3, pp. 589–607, 2018.

[7] S. A. Hosseini and S. Haghighat, "The Relationship between Corporate Governance and Community Engagement in Listed Companies of Tehran Stock Exchange," *Journal of accounting and social interests*, vol. 6, no. 4, pp. 103–128, 2016.

[8] J. S. Baek, "Corporate finance for advancement in emerging markets," *Emerging Markets Finance and Trade*, vol. 51, no. -sup3, pp. 1-2, 2015.

[9] A. Marcia, W. Maroun, and C. Callaghan, "Value relevance and corporate responsibility reporting in the South African context: An alternate view post King-III," *South African Journal of Economic and Management Sciences*, vol. 18, no. 4, pp. 500–518, 2015.

[10] W. J. Baumol, "Education for innovation: Entrepreneurial breakthroughs versus corporate incremental improvements," *Innovation policy and the economy*, vol. 5, pp. 33–56, 2005.

[11] A. Lioui and P. Maio, "Interest rate risk and the cross section of stock returns," *Journal of Financial and Quantitative Analysis*, vol. 49, no. 2, pp. 483–511, 2014.

[12] S. S. Aneel, U. T. Haroon, and I. Niazi, *Corridors of Knowledge for Peace and Development*, Sustainable Development Policy Institute, 2019.

[13] T. M. Fischer and A. A. Sawczyn, "The relationship between corporate social performance and corporate financial performance and the role of innovation: evidence from German listed firms," *Journal of Management Control*, vol. 24, no. 1, pp. 27–52, 2013.

[14] M. Awais, N. Badruddin, and M. Drieberg, "A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability," *Sensors*, vol. 17, no. 9, p. 1991, 2017.

[15] W. Qi, S. E. Ovur, Z. Li, A. Marzullo, and R. Song, "Multi-sensor guided hand gesture recognition for a teleoperated robot using a recurrent neural network," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 6039–6045, 2021.

[16] D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, "An emotion recognition model based on facial recognition in virtual learning environment," *Procedia Computer Science*, vol. 125, pp. 2–10, 2018.

[17] A. Mehrabian and J. A. Russell, *An approach to environmental psychology*, the MIT Press, 1974.

[18] L. Chen, X. Mao, Y. Xue, and L. L. Cheng, "Speech emotion recognition: features and classification models," *Digital Signal Processing*, vol. 22, no. 6, pp. 1154–1160, 2012.

[19] T. Chen, S. Ju, F. Ren, M. Fan, and Y. Gu, "EEG emotion recognition model based on the LIBSVM classifier," *Measurement*, vol. 164, p. 108047, 2020.

[20] H. Eilbirt and I. R. Parket, "The practice of business," *Business Horizons*, vol. 16, no. 4, pp. 5–14, 1973.

[21] H. Yin and Y. Sun, "Analysis of China's Film Industry in 2020," *Journal of Chinese Film Studies*, vol. 1, no. 2, pp. 295–328, 2021.

[22] C. Lin, Y. Ma, and D. Su, "Corporate governance and firm efficiency: evidence from China's publicly listed firms," *Managerial and Decision Economics*, vol. 30, no. 3, pp. 193–209, 2009.

[23] A. Joseph, *Schumpeter: The economics and sociology of capitalism*, Princeton University Press, 2020.

[24] T. W. Dunfee and T. Donaldson, "Contractarian business ethics: current status and next steps," *Business Ethics Quarterly*, vol. 5, no. 2, pp. 173–186, 1995.

[25] S. Bhagat and B. Black, *The uncertain relationship between board composition and firm performance*, The Business Lawyer, 1999.

[26] B. Peters, *Innovation and firm performance: An empirical investigation for German firms*, Springer Science & Business Media, 2008.

[27] J. J. Chrisman, J. H. Chua, and P. Sharma, "Trends and directions in the development of a strategic management theory of the family firm," *Entrepreneurship theory and practice*, vol. 29, no. 5, pp. 555–575, 2005.

[28] R. Agarwal, G. Gao, C. DesRoches, and A. K. Jha, "Research commentary the digital transformation of healthcare: current status and the road ahead," *Information Systems Research*, vol. 21, no. 4, pp. 796–809, 2010.

[29] X. Xie, H. Wang, and H. Jiao, "Non-R&D innovation and firms' new product performance: the joint moderating effect of R&D intensity and network embeddedness," *R&D Management*, vol. 49, no. 5, pp. 748–761, 2019.

[30] H. Chun and S. B. Mun, "Determinants of R&D cooperation in small and medium-sized enterprises," *Small Business Economics*, vol. 39, no. 2, pp. 419–436, 2012.

[31] S. Michailova, "Contextualizing in international business research: why do we need more of it and how can we be better at it?," *Scandinavian Journal of Management*, vol. 27, no. 1, pp. 129–139, 2011.

[32] H. Huang, "Shareholder derivative litigation in China: Empirical findings and comparative analysis," *Banking & Finance Law Review*, vol. 27, no. 4, p. 619, 2012.

[33] K. J. Wang and Y. D. Lestari, "Firm competencies on market entry success: Evidence from a high-tech industry in an emerging market," *Journal of Business Research*, vol. 66, no. 12, pp. 2444–2450, 2013.

[34] R. Hu, Q. Liang, C. Pray, J. Huang, and Y. Jin, "Privatization, public R&D policy, and private R&D investment in China's agriculture," *Journal of Agricultural and Resource Economics*, pp. 416–432, 2011.

[35] G. E. Bronnikov, S. O. Vinogradova, V. S. Mezentseva, and E. V. Samoĭlova, "Reasons causing a lag period in the oxidative phosphorylation process. Isn't ATP an internal

uncoupler of ATP synthetase?," *Biofizika*, vol. 44, no. 3, pp. 465–473, 1999.

[36] E. Bell and A. Bryman, "The ethics of management research: an exploratory content analysis," *British Journal of Management*, vol. 18, no. 1, pp. 63–77, 2007.

[37] W. Wu, F. Peng, Y. G. Shan, and L. Zhang, "Litigation risk and firm performance: The effect of internal and external corporate governance," *Corporate governance: an international review*, vol. 28, no. 4, pp. 210–239, 2020.

[38] Y. C. Chen, M. Hung, and Y. Wang, "The effect of mandatory CSR disclosure on firm profitability and social externalities: Evidence from China," *Journal of accounting and economics*, vol. 65, no. 1, pp. 169–190, 2018.

[39] X. Ji, J. Ding, X. Xie et al., "Pollution status and human exposure of decabromodiphenyl ether (BDE-209) in China," *ACS omega*, vol. 2, no. 7, pp. 3333–3348, 2017.

[40] H. K. Baker, N. Pandey, S. Kumar, and A. Haldar, "A bibliometric analysis of board diversity: current status, development, and future research directions," *Journal of Business Research*, vol. 108, pp. 232–246, 2020.

[41] Y. Li, H. Zhang, Y. Liu, and Q. Huang, "Impact of Embedded Global Value Chain on Technical Complexity of Industry Export—An Empirical Study Based on China's Equipment Manufacturing Industry Panel," *Sustainability*, vol. 12, no. 7, p. 2694, 2020.

[42] Y. Zhang, U. Khan, S. Lee, and M. Salik, "The influence of management innovation and technological innovation on organization performance. A mediating role of sustainability," *Sustainability*, vol. 11, no. 2, p. 495, 2019.

[43] H. Jiao, J. Zhou, T. Gao, and X. Liu, "The more interactions the better? The moderating effect of the interaction between local producers and users of knowledge on the relationship between R&D investment and regional innovation systems," *Technological Forecasting and Social Change*, vol. 110, pp. 13–20, 2016.

[44] S. M. Abdullah, S. Y. Ameen, M. A. Sadeeq, and S. Zeebaree, "Multimodal emotion recognition using deep learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 2, pp. 52–58, 2021.

[45] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1D & 2D CNN LSTM networks," *Biomedical signal processing and control*, vol. 47, pp. 312–323, 2019.

[46] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1301–1309, 2017.

[47] D. Wang, J. Su, and H. Yu, "Feature extraction and analysis of natural language processing for deep learning English language," *IEEE Access*, vol. 8, pp. 46335–46345, 2020.

[48] Y. Zhang, X. Shi, H. Zhang, Y. Cao, and V. Terzija, "Review on deep learning applications in frequency analysis and control of modern power system," *International Journal of Electrical Power & Energy Systems*, vol. 136, article 107744, 2022.

[49] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 211–223, 2012.

[50] G. Yolcu, I. Oztel, S. Kazan, C. Oz, and F. Bunyak, "Deep learning-based face analysis system for monitoring customer interest," *Journal of ambient intelligence and humanized computing*, vol. 11, no. 1, pp. 237–248, 2020.

[51] B. Chen, Q. Cao, M. Hou, Z. Zhang, G. Lu, and D. Zhang, "Multimodal emotion recognition with temporal and semantic consistency," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3592–3603, 2021.

[52] F. Zhang, T. Zhang, Q. Mao, and C. Xu, "Geometry guided pose-invariant facial expression recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 4445–4460, 2020.

[53] P. Lorette and J. M. Dewaele, "The relationship between bi/multilingualism, nativeness, proficiency and multimodal emotion recognition ability," *International Journal of Bilingualism*, vol. 23, no. 6, pp. 1502–1516, 2019.

[54] S. Zhang, X. Pan, Y. Cui, X. Zhao, and L. Liu, "Learning affective video features for facial expression recognition via hybrid deep learning," *IEEE Access*, vol. 7, pp. 32297–32304, 2019.

[55] P. Xu, A. Madotto, C. S. Wu, J. H. Park, and P. Fung, "Emo2-vec: Learning generalized emotion representation by multi-task training," 2018, https://arxiv.org/abs/1809.04505.

[56] A. Zadeh, M. Chen, S. Poria, E. Cambria, and L. P. Morency, "Tensor fusion network for multimodal sentiment analysis," 2017, arXiv preprint arXiv: 1707.07250.

[57] H. Li, J. Zhu, C. Ma, J. Zhang, and C. Zong, "Read, watch, listen, and summarize: multi-modal summarization for asynchronous text, image, audio and video," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 5, pp. 996–1009, 2019.

[58] C. H. Chan, J. Kittler, N. Poh, T. Ahonen, and M. Pietikäinen, "(Multiscale) local phase quantisation histogram discriminant analysis with score normalisation for robust face recognition," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pp. 633–640, Kyoto, Japan, 2009.

[59] J. G. Chen, N. Wadhwa, Y. J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk, "Modal identification of simple structures with high-speed video using motion magnification," *Journal of Sound and Vibration*, vol. 345, pp. 58–71, 2015.

[60] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: a survey of approaches and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2724–2743, 2017.

[61] Q. Guo, F. Zhuang, C. Qin et al., "A survey on knowledge graph-based recommender systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 8, pp. 3549–3568, 2022.

[62] O. Yildirim, U. B. Baloglu, R. S. Tan, E. J. Ciaccio, and U. R. Acharya, "A new approach for arrhythmia classification using deep coded features and LSTM networks," *Computer Methods and Programs in Biomedicine*, vol. 176, pp. 121–133, 2019.

[63] S. Muzaffar and A. Afshari, "Short-term load forecasts using LSTM networks," *Energy Procedia*, vol. 158, pp. 2922–2927, 2019.

[64] N. Reimers and I. Gurevych, "Optimal hyperparameters for deep lstm-networks for sequence labeling tasks," 2017, arXiv preprint arXiv: 1707.06799.

[65] X. Zhu, X. Ao, Z. Qin et al., "Intelligent financial fraud detection practices in post-pandemic era," *The Innovation*, vol. 2, no. 4, p. 100176, 2021.

[66] Z. Xu and W. Zhou, "A data technology oriented to information fusion to build an intelligent accounting computerized model," *Scientific Programming*, vol. 2021, Article ID 6031324, 12 pages, 2021.