International Journal for Housing Science and Its Applications
Publish September 9, 2025. Volume 47, Issue 1 Pages 292-306

https://doi.org/10.70517/ijhsa47124

The Teaching Value of Traditional Culture in Chinese Language and Literature Based on Big Data Analysis and Its Application in Vocational Education

Wei Li[1,*]
[1] School of International Communication, Hunan Mass Media Vocational and Technical College, Changsha, Hunan, 410100, China
Corresponding authors: (e-mail: 18008463868@163.com).

Abstract This paper combines N-gram models, LDA topic models, TF-IDF algorithms, and clustering analysis methods, and based on the LIA-BiLSTM text sentiment analysis model, collects, organizes, and analyzes a large amount of Chinese language and literature texts. This paper uses big data technology to deeply explore the educational value of traditional culture in Chinese language and literature and discusses its application path in the field of vocational education. Research indicates that the improved LIA-BiLSTM model developed in this study outperforms other models, with an AUC value of 0.9779, approaching 1. Traditional cultural elements in Chinese language and literature help enhance students' sense of cultural confidence, belonging, achievement, and responsibility. By constructing an integrated educational model of "course integration—cultural immersion—practical experience—evaluation feedback," this approach not only effectively preserves China's excellent traditional culture but also infuses vocational education with new cultural significance.

Index Terms N-gram model, LDA topic model, TF-IDF algorithm, cluster analysis, text sentiment analysis

I. Introduction

China plays a significant role in economic globalization, and Chinese elements and Chinese culture hold an important position in global culture. With the development of cross-cultural exchange and the tourism industry, Chinese-related cultural and creative products and traditional culture are being disseminated overseas. Chinese society is increasingly valuing traditional culture, and television programs related to traditional culture are also gradually increasing, which will inevitably influence contemporary people's social values [1]-[4]. In the field of education, the Ministry of Education has issued guidelines on integrating traditional culture into education, aiming to achieve a deep integration of traditional culture with modern education. Traditional culture is increasingly being recognized by educators [5], [6]. Therefore, teachers need to correctly understand the value of traditional culture in teaching and make effective adjustments in their teaching practices to bring traditional culture into the classroom, elevating the classroom to a new level and fully leveraging the value of contemporary classroom teaching [7]-[9].

The number of countries worldwide where Chinese is studied as a foreign language is steadily increasing, with over 150 million learners. However, most learners only grasp the basic meanings of Chinese and lack a deep understanding of its cultural connotations [10], [11]. Chinese language and literature serve as a vital carrier of traditional Chinese culture and a key manifestation of the nation's cultural spirit. ZHANG [12] notes that literature is one of the primary means of transmitting traditional culture, and Chinese language and literature play a pivotal role in the inheritance and development of traditional Chinese culture. Therefore, exploring the educational value of China's excellent traditional culture embedded in language and literature not only has positive significance for understanding, promoting, and inheriting traditional culture and building modern civilization, but also enriches language teaching content and injects fresh vitality into teaching.

Additionally, Cui [13] clarified that traditional Chinese culture has multiple effects on students, shaping their moral character, tempering their willpower, and fostering their thinking patterns. Yu [14] emphasizes that professional talent must possess a craftsman spirit, and Chinese traditional culture serves as a vital foundation for cultivating this spirit in vocational education. Nie [15] shares the value and application of traditional culture in undergraduate vocational education, highlighting it as an important educational resource for fostering students' sense of social responsibility and professional ethics, shaping their cultural literacy and humanistic spirit, and strengthening their international perspective and cross-cultural communication skills. In career development, students' mental health is a crucial factor influencing successful entrepreneurship and employment. Zheng[16] incorporated traditional culture into Chinese language and literature education, enhancing students' literary literacy while also improving their mental health levels. These studies not only highlight the educational value of traditional culture in language and literature but also emphasize its application value in vocational education. A deeper analysis reveals that traditional culture in Chinese language and literature plays a significant role in education.

In today's society, the rapid development of big data technology has profoundly impacted various fields. In the field of literature, Sun[17] utilized tools such as the LDA model, GBDT algorithm, SO-PMI algorithm, and vector space model under big data analysis to examine changes in the thematic classification, emotional orientation, and stylistic types of Chinese literary works. Yuan [18] used artificial intelligence and big data technology to analyze the character traits in "Dream of the Red Chamber," promoting an objective understanding of the humanistic spirit and social concepts in literary works. Hang [19] used big data analysis to explore the meaning of ancient literary texts, including the basic creative intentions and generative factors of the authors, to understand the historical and cultural context of the time and the cultural elements of the present.

In the field of traditional culture, big data analyzes ancient texts, historical documents, and folk legends to conduct data mining and analysis on music, customs, rituals, and festivals of different ethnic groups, thereby identifying common and unique cultural characteristics. Based on this, by extracting keywords, core concepts and ideological systems within traditional culture are analyzed, and through comparative analysis and summary, the essence of traditional culture is conveyed to modern people [20]-[24]. Additionally, through big data analysis, one can analyze the connections and exchanges between cultures from massive cultural data, discover cultural interpretation and integration methods, and thereby provide inspiration for innovation, driving contemporary cultural innovation and dissemination [25].

This paper focuses on using big data technology to deeply explore the educational value of traditional culture in Chinese language and literature. It proposes a text sentiment analysis model, LIA-BiLSTM, based on LERT and an interactive attention mechanism. Through text preprocessing, keyword extraction, co-occurrence analysis, clustering analysis, and sentiment analysis, it quantifies traditional cultural elements in the text. Based on this, it deeply explores the educational value of traditional culture in Chinese language and literature, thereby proposing feasible strategies for integrating China's excellent traditional culture into vocational education.

II. Quantitative analysis of traditional cultural elements in Chinese language and literature based on big data

This chapter uses big data analysis technology to conduct quantitative analysis of traditional cultural elements in Chinese language and literature, and combines the LIA-BiLSTM text sentiment analysis model to explore the educational value of traditional culture in Chinese language and literature.

II. A.Framework for quantitative analysis of traditional cultural elements based on text mining

In order to comprehensively and systematically quantify and analyze traditional cultural elements in Chinese language and literature, this paper uses R language to construct a quantitative analysis framework for Chinese language and literature texts based on text mining. The analysis framework is divided into four modules: First, a specialized dictionary for Chinese language and literature is constructed using Chinese language and literature texts. Second, all Chinese language and literature texts undergo a series of processes such as word segmentation and high-dimensional data reduction to provide support for subsequent research. Third, high-frequency vocabulary extraction and analysis. Fourth, LDA topic model construction.

II. A. 1) Text preprocessing

First, use the quanteda package to build a corpus for Chinese language and literature texts. Second, load the Sogou Professional Dictionary, along with the Harbin Institute of Technology stopword list and policy feature stopwords as stopword dictionaries. Next, use jiebaR for Chinese word segmentation, adding the constructed dictionaries during segmentation and performing stopword and invalid character processing. It is important to note that during text processing, words that are meaningless for the analysis of Chinese language and literature texts but appear with high frequency may be identified at any time. Therefore, such words should be added to the stopword list and the text preprocessing should be repeated. This process is typically repeated multiple times until no invalid information is found among the high-frequency words.

II. A. 2) N-gram model

In Chinese language and literature, there are some meaningful high-frequency words composed of 2–3 words. If these are directly processed using word segmentation, they will be split into two separate words, resulting in significant loss of textual information. The N-gram model [26] is a commonly used statistical language model in natural language processing and computational linguistics, which predicts the probability of the next word appearing based on the previous N-1 words in a given text. Therefore, this paper utilizes the N-gram model to construct a specialized dictionary for Chinese language and literature based on Chinese language and literature texts.

II. A. 3) TF-IDF Algorithm

Most Chinese language and literature texts are long texts with a high level of information noise, which can easily generate high-dimensional sparse matrices during preprocessing. This not only slows down the processing speed of text data but also causes a lot of information interference. To address this issue, the article uses the TF-IDF algorithm [27] for preliminary dimensionality reduction of the data. Specifically, the term frequency TF is first calculated:

$$TF(w_i,d)=\frac{Count(w_i,d)}{Count(d)} \tag{1}$$

Among them, TF represents the frequency of a word appearing in a document. $Count(w_i,d)$ represents the number of times the word w_i appears in document d, and $Count(d)$ represents the total number of words in document d.

Next, calculate the inverse document frequency IDF:

$$IDF(w_i,D)=\log\frac{N}{N(w_i)+1} \tag{2}$$

In this formula, N represents the total number of documents in the corpus, and $N(w_i)$ represents the number of documents containing the word w_i.

Finally, the formula for calculating $TF-IDF$ is:

$$TF-IDF(w_i,d,D)=TF(w_i,d)\times IDF(w_i,D) \tag{3}$$

The higher the $TF-IDF$ score, the greater the importance of the term in the document. Relative to the entire corpus, the more unique the term is, and the more common the term appears in many documents, the lower its $TF-IDF$ score will be. The article deletes terms with a TF value less than 0.0001 or a $TF-IDF$ value less than 0.0001.

II. A. 4) LDA topic model

LDA[28] is a probabilistic graphical model used for topic modeling in the field of machine learning. Its purpose is to discover the latent topics in a given set of documents, determine the distribution of each topic in each document, and determine the distribution of words in each topic. The LDA topic model can be used to better explore the connotations of Chinese language and literature and traditional cultural elements. The article uses Gibbs sampling to construct the LDA model. Specifically, the hyperparameters α is set to 0.1, β is set to 50/K, and the number of iterations is set to 3000.

II. A. 5) Theme Strength and Vocabulary Theme Relevance

After obtaining the latent themes of the text using LDA, the theme strength can be calculated to represent the relative weight of that theme in the entire Chinese language and literature text. The specific calculation formula is as follows:

$$P_j=\frac{\sum_{d=1}^{M}\theta_{dj}}{M} \tag{4}$$

Among them, P_j represents the strength of the j th topic, M is the number of documents, and θ_{dj} represents the probability of the j th topic in the d th document.

At the same time, the vocabulary topic relevance definition is used to show the relevance of each vocabulary to each topic:

$$r(w,j\mid\lambda)=\lambda\log(\varphi_{jw})+(1-\lambda)\log(\frac{\varphi_{jw}}{P_w}) \tag{5}$$

Among these, $r(w,j\mid\lambda)$ represents the relevance of word w to the j th topic. φ_{jw} represents the probability that word w belongs to topic j, and p_w represents the empirical distribution of word w obtained from the corpus. λ is a tunable parameter ranging from 0 to 1. The closer λ is to 1, the higher the relevance of words that appear more frequently under that topic. The closer λ is to 0, the more relevant unique words under that topic are to the topic. Following previous research, λ is set to 0.6 in this paper.

II. A. 6) Cluster analysis of thematic content

The themes are not independent but are interrelated and mutually inclusive. Multidimensional scaling analysis can be used to distribute the themes across two dimensions. Based on the relative positions of the themes, they can be clustered to form a distribution of themes in Chinese language and literature. This article uses the LDAvis visualization package in R to visualize the themes and cluster them based on the distances between them and their characteristic vocabulary.

II. B. Text Sentiment Analysis Model Based on LIA-BiLSTM

This section proposes a text sentiment analysis model called LIA-BiLSTM based on LERT and interactive attention mechanisms. The model mainly consists of a data preprocessing layer, an LERT word embedding layer, a bidirectional LSTM feature extraction layer, an interactive attention layer, and an output layer. The structure of the LIA-BiLSTM model is shown in Figure 1.

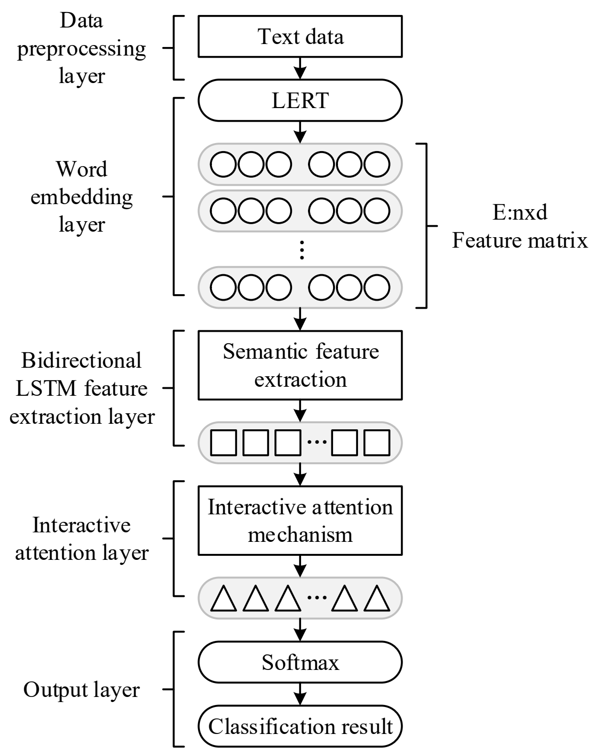

Figure 1: Structure of LIA-BiLSTM model

II. B. 1) Data Preprocessing Layer

The primary task of the data preprocessing layer is to clean raw text data, preparing it for text vector representation using the LERT pre-trained language model. To ensure more accurate data cleaning, this study integrated multiple commonly used stopword lists and special symbol lists to construct a comprehensive stopword and special symbol library. Finally, regular expressions and the jieba word segmentation tool were used to remove data noise and stopwords from the raw text data, resulting in a text dataset suitable for use in the LERT word embedding layer.

II. B. 2) LERT Word Embedding Layer

To obtain richer semantic features from Chinese language and literature text data, this study adopted a linguistically enhanced pre-training model called LERT[29]. This model adds multiple linguistic tasks to the BERT model architecture and optimizes each linguistic task through the LIP linguistic information pre-training strategy.

In the Pre-Training Architecture module, LERT segments the text data and adds special markers [CLS] and [SEP] at the beginning and end of each sentence. Each token is converted into a word vector with position information, H_i^{l-1} using the word embedding vector E_i and position encoding $PosEnc(i)$. The calculation methods for $PosEnc(i)$ and H_i^{l-1} are shown in Equations (6) and (7):

$$PosEnc(i)=\begin{cases}\sin\left(\dfrac{i}{10000^{\frac{j}{d}}}\right), \text{If j is an even number}\\ \cos\left(\dfrac{i}{10000^{\frac{j-1}{d}}}\right), \text{If j is an odd number}\end{cases} \tag{6}$$

$$H_i^{l-1}=E_i+PosEnc(i) \tag{7}$$

In formula (6), i denotes the position of the word in the sequence, j denotes the j th dimension of the vector, and d denotes the total dimension of the vector. In equation (7), E_i denotes the word embedding vector, $PosEnc(i)$ denotes the position encoding, and l denotes the number of layers in the Transformer network, where l is 1, indicating the initial embedding layer.

The word vector H_i^0 is processed through the multi-head attention mechanism and feedforward neural network of each layer of the Transformer network to compute H_i^l, as shown in Equations (8) to (10):

$$M_i=\tau\left(\frac{(H_i^{l-1}W_Q)(H_i^{l-1}W_K)^T}{\sqrt{d_k}}(H_i^{l-1}W_V)W_O\right) \tag{8}$$

$$Z_i=\sigma(W_1M_i+b_1)W_2+b_2 \tag{9}$$

$$H_i^l=\gamma\left(\frac{(H_i^{l-1}+Z_i)-\mu_i}{\sqrt{\delta_i^2+\varepsilon}}\right)+\beta \tag{10}$$

In equation (8), W_Q,W_K,W_V,W_O represent the query, key, value, and output matrices, respectively, while d_k and M_i denote the scaling factor and the output of the multi-head attention mechanism, τ denotes the softmax activation function, and l is the number of layers in the transformer network, where l is greater than 1, indicating a multi-layer network. In equation (9), W_1,W_2,b_1,b_2 are the weights and biases of the feedforward neural network, σ is the GELU activation function, and Z_i is the output of the feedforward neural network. In formula (10), γ and β are trainable scaling and translation parameters, ε is a constant to prevent instability, and μ and δ are mean and variance parameters.

In the Linguistic Analysis module, LERT utilizes its unique part-of-speech (POS), named entity recognition (NER), and dependency parsing (DEP) to annotate text data, using the LTP pre-training strategy to optimize the relationships between linguistic tasks, and training with annotated data to obtain vectors P_i,N_i,D_i as shown in formula (11):

$$\begin{cases}P_i=\sum_{j=1}^{K_{pos}}\delta_{j,pos_i}E_j^{(pos)}\\ N_i=\sum_{j=1}^{K_{pos}}\delta_{j,ner_i}E_j^{(ner)}\\ D_i=\sum_{j=1}^{K_{dep}}\delta_{j,dep_i}E_j^{(dep)}\end{cases} \tag{11}$$

Among these, P_i,N_i,D_i correspond to the annotation vectors for the three linguistic tasks, K is the total number of part-of-speech labels, E_j corresponds to the embedding matrices for the three linguistic tasks, and δ is an indicator variable. When j equals the index of the corresponding part-of-speech label (pos_i,nr_i, or dep_i), δ takes the value 1, otherwise δ takes the value 0.

The output H_i^l of the Transformer network is then fused with the outputs P_i,N_i,D_i from the linguistic tasks to obtain the dynamic vector E, as shown in formula (12):

$$E=H_i^l+\eta P_i+\theta N_i+\varphi D_i \tag{12}$$

Among them, η,θ,φ represent the relevant parameters for the three linguistic tasks of part-of-speech tagging, named entity recognition, and dependency parsing, respectively.

II. B. 3) Bidirectional LSTM Feature Extraction Layer

This study employs a bidirectional LSTM [30] structure as the feature extraction layer for the algorithm in this section to extract semantic features from text data. The bidirectional LSTM combines forward and backward LSTM units to more comprehensively capture contextual information in sequences, thereby enhancing the ability to extract text semantic features. The feature extraction structure of the bidirectional LSTM is shown in Figure 2.

In Figure 2, the bidirectional LSTM feature extraction structure consists of two LSTM layers: one layer processes sequence data from left to right, and the other layer processes sequence data from right to left. The hidden state h_t at each time step contains information from both the forward and backward directions of the sequence, $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$, as shown in Formula (13). The symbol [;] denotes vector concatenation, and $\overrightarrow{h}_t, \overleftarrow{h}_t$ are calculated according to formula (14). In this way, the bidirectional feature extraction layer can effectively integrate information from both the front and back contexts of the sequence, thereby more comprehensively extracting the semantic features H of the text data, as shown in formula (15):

$$h_t = [\overrightarrow{h}_t ; \overleftarrow{h}_t] \tag{13}$$

$$h_t = o_t \odot \tanh(C_t) \tag{14}$$

$$H = \{h_1, h_2, \ldots, h_T\} \tag{15}$$

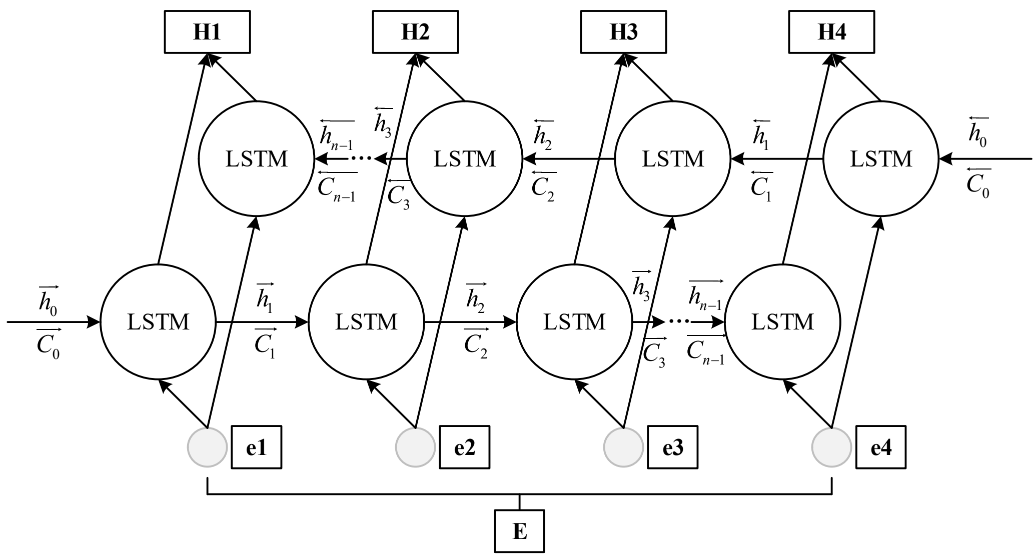

Figure 2: Bidirectional LSTM feature extraction structure

II. B. 4) Interaction Attention Layer

To enhance the model's ability to extract key semantic features from text data, this study introduces an interaction attention mechanism layer on top of the bidirectional LSTM feature extraction layer.

The interaction attention mechanism first receives the output from the LSTM, $H = \{h_1, h_2 \ldots h_T\}$, which not only combines the hidden state at the current time step h_t but also incorporates hidden states from other time steps h_j, thereby capturing the complex relationships between different time steps and features in the sequence. It then generates attention scores u_t through linear transformation and a nonlinear activation function, as shown in formula (16). These attention scores are mapped to scalars through another linear transformation and calculated using formula (17) to obtain attention weights α_t. The attention weights α_t are weighted and fused with the hidden state h_t of the LSTM output to generate the global key feature vector G, as shown in Formula (18):

$$u_t = \tanh(W \times h_t + \sum_{j=1}^{T} A_{t,j} \times h_j) \tag{16}$$

$$\alpha_t = \frac{\exp(u_t \times v)}{\sum_{t=1}^{T} \exp(u_t \times v)} \tag{17}$$

$$G = \sum_{t=1}^{T} \alpha_t \times h_t \tag{18}$$

Among them, $\tanh$ is a nonlinear activation function, $A_{i,j}$ is the association matrix between time step t and j, W is the initialized weight matrix, and v is the weight vector.

II. B. 5) Output layer

The output layer of the LIA-BiLSTM sentiment analysis model primarily consists of a fully connected layer and a normalization module. First, the interaction attention layer inputs the extracted key semantic feature vectors G of the text data into the fully connected layer, which then converts these high-dimensional feature vectors into an output form o suitable for classification tasks, as shown in Formula (19). Then, the normalization function softmax is applied to o to obtain the sentiment classification probability distribution g, as shown in Formula (20):

$$o = W \times G + b \tag{19}$$

$$g = softmax(o_i) = \frac{e^{o_i}}{\sum_{j=1}^{k} e^{o_j}} \tag{20}$$

In this context, W is the weight matrix, b is the bias vector, o_i is the i th input vector, and k is the vector dimension.

g enables the LIA-BiSTM model to provide a specific prediction probability for each sentiment category, thereby yielding the sentiment analysis results for the text data.

During the training process of the LIA-BiSTM sentiment analysis model, the loss between the model's prediction output g and the actual label y is calculated using the cross-entropy loss function. Then, the parameters of the fully connected layer are updated through the gradient descent strategy to continuously adjust the model parameters and improve the model's performance. The calculation method of the cross-entropy loss function is shown in Formula (21):

$$L(g, y) = -\frac{1}{N}\sum_{i=1}^{N}[g_i \log y_i + (1-g_i)\log(1-y_i)] \tag{21}$$

In this case, N is the sample size, y is the actual label, g is the model's predicted output, and i is the index.

III. Analysis of the educational value of traditional culture in Chinese language and literature

This chapter uses big data analysis technology to conduct a quantitative analysis of traditional cultural elements in Chinese language and literature texts, and based on the analysis results, summarizes the teaching value of traditional culture in vocational education.

III. A. Experimental Results and Analysis

III. A. 1) Data Collection and Preprocessing

This study selected classic literary works spanning various genres, including ancient Chinese poetry, prose, novels, and opera scripts, as its data sources. It also incorporated research papers and related interpretive materials by modern scholars on these works. Using web crawling technology, we collected raw text data from authoritative literary databases and academic journal platforms, and performed preprocessing operations such as noise removal, word segmentation, and annotation to ensure data quality and usability. For example, when processing ancient Chinese texts, we used specialized ancient text recognition tools to address issues such as traditional characters and variant characters. For special expressions in literary works, such as allusions and metaphors, we established corresponding knowledge graphs for semantic association annotation.

III. A. 2) Application Analysis of the LIA-BiLSTM Model

In this experiment, the selected Chinese language and literature works were divided into five datasets—poetry, prose, fiction, quatrains, and miscellaneous drama—based on genre, and named datasets A to E, respectively, to validate the effectiveness of the proposed LIA-BiLSTM model.

(1) Comparative experiments across datasets

First, through comparative experiments, we compared the performance of traditional machine learning methods such as SVM, KNN, and RF, deep learning methods such as GCN, CNN, and CSS, and the LIA-BiLSTM sentiment analysis model proposed in this paper across multiple sentiment analysis datasets. The comparison results of accuracy and F1 scores for different models across datasets A–E are shown in Table 1.

In terms of accuracy, most traditional machine learning methods performed well on dataset A, but overall, they still lagged behind deep learning methods. Among deep learning methods, the Bi-LSTM model demonstrated good accuracy on datasets C and D, but failed to achieve ideal results on datasets B and E. This suggests that the Bi-LSTM model faces challenges in processing complex sentiment text data and has limitations in analyzing local sentiment information. The CSS model, BiGRU-CNN model, BERT-BiLSTM model, and LIA-BiLSTM sentiment analysis model all demonstrate relatively balanced and high-precision model performance across multiple datasets. However, the LIA-BiLSTM sentiment analysis model used in this paper outperforms all other models in terms of accuracy on most datasets, particularly on Dataset C and Dataset D, achieving 96.89% and 95.72%, respectively.

In terms of F1 scores, traditional machine learning methods lag behind deep learning methods. Among deep learning methods, the BiLSTM model performs well on Datasets C and D but has lower F1 scores on other datasets. The CSS model and BiGRU-CNN model both achieved high F1 scores. The improved LIA-BiLSTM sentiment analysis model proposed in this paper performed exceptionally well on most datasets, consistently leading other models in F1 scores across all four datasets (A–D).

It can be observed that while traditional machine learning methods perform well on certain datasets, their performance on more complex datasets is unsatisfactory and cannot compete with deep learning methods. Mainstream deep learning models demonstrate superior performance across multiple datasets compared to traditional machine learning models, particularly achieving outstanding results on datasets involving complex text. The improved LIA-BiLSTM model proposed in this paper combines the LERT word embedding module, interaction attention mechanism, and Bi-LSTM feature extraction module to make corresponding improvements for different types of sentiment text. When evaluating the two metrics across various complex datasets, it achieves the best performance, demonstrating the effectiveness of the proposed improved sentiment analysis model.

Table 1: The comparison results of the accuracy and F1 scores of different models

	Dataset A		Dataset B		Dataset C		Dataset D		Dataset E	
Method	Accuracy /%	F1 /%	Accuracy /%	F1 /%	Accuracy /%	F1 /%	Accuracy /%	F1 /%	Accuracy /%	F1 /%
SVM	90.52	90.45	77.02	76.85	81.89	82.15	89.51	83.41	85.99	86.33
KNN	89.47	89.99	76.81	76.17	75.04	74.94	76.33	65.35	76.79	76.51
RF	88.52	89.32	75.25	75.55	74.57	74.73	86.37	80.49	78.99	78.99
BiLSTM	86.69	86.64	78.14	78.52	86.91	87.15	89.17	89.37	87.11	87.14
BiGRU	86.09	85.96	78.83	78.89	86.81	87.22	89.54	89.42	86.76	87.32
CNN	86.94	86.93	78.43	77.66	89.54	89.25	90.25	89.79	89.03	89.23
GCN	86.16	86.40	77.64	77.55	84.11	84.59	89.78	90.34	89.21	89.21
FastText	85.42	85.21	76.88	77.74	86.91	86.86	87.74	88.12	87.51	87.67
CSS	93.29	93.56	91.34	91.56	95.88	96.18	93.96	94.19	94.83	94.77
BiGRU-CNN	93.05	93.11	90.63	90.29	96.12	96.09	94.33	94.62	94.17	93.94
BERT-BiLSTM	93.15	93.26	90.72	91.35	96.35	96.63	95.55	95.39	93.48	93.71
LIA-BiLSTM	93.49	93.38	92.59	92.06	96.89	96.85	95.72	95.64	93.75	93.87

(2) ROC curve

The ROC curve is a graphical representation used to evaluate the performance of a binary classification model. This paper uses dataset B as an example to demonstrate the model performance of the LIA-BiLSTM model as shown by the ROC curve. From the comparative experiments, it can be observed that among deep learning-based sentiment analysis models, the LIA-BiLSTM model proposed in this paper exhibits performance comparable to that of the CSS model and the BiGRU-CNN model. Therefore, this paper will further conduct a detailed comparison of the model performance of these three models using the ROC curve comparison method.

The core idea of the ROC curve is to evaluate model performance using two metrics: true positive rate (TPR), whose formula is shown in Equation (22):

$$TPR = \frac{TP}{TP + FN} \tag{22}$$

Where TP is the number of samples correctly predicted as positive, and FN is the number of positive samples incorrectly predicted as negative.

Another metric is the false positive rate (FPR), which is given by equation (23):

$$FPR = \frac{FP}{FP + TN} \tag{23}$$

Among these, FP represents the number of negative samples incorrectly predicted as positive, and TN represents the number of samples correctly predicted as negative.

When the model's ROC curve approaches the upper-left corner, i.e., FPR approaches 0 and TPR approaches 1, this indicates that the false positive rate is very low and the true positive rate is very high, signifying excellent classification performance of the model.

Another important metric for evaluating the ROC curve is the AUC (Area Under the Curve), which represents the model's classification capability through the area under the ROC curve. Typically, the AUC value ranges between 0 and 1. When the AUC value is 1, it indicates that the model's classification capability is nearly perfect. When the value is 0.5, it indicates that the classifier has no discriminative ability, equivalent to random guessing. When the value is less than 0.5, it indicates that the model's classification performance is poor.

The comparison of the ROC curves for the LIA-BiLSTM model, CSS model, and BiGRU-CNN model is shown in Figure 3. All three models are close to the top-left corner of the figure, indicating that their classification performance is good. From the shape of the curves, the LIA-BiLSTM model's curve maintains a high TPR and low FPR across the entire range, with an AUC value of 0.9779, which is close to 1. This indicates that the model can effectively distinguish between positive and negative classes, making highly accurate predictions for positive class samples while minimizing misclassifications of negative class samples. This means that the LIA-BiLSTM model maintains high classification performance across all classification thresholds. It also demonstrates the advantages of the LIA-BiLSTM model over the CSS model and BiGRU-CNN model.

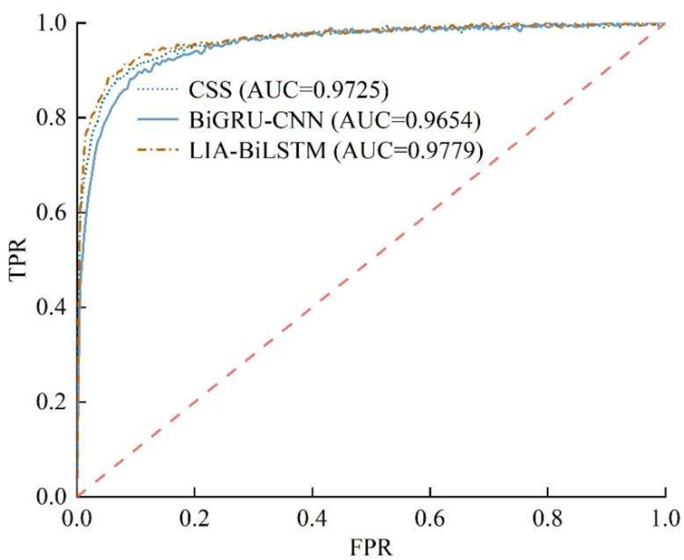

Figure 3: ROC curve

III. A. 3) Quantitative Analysis of Traditional Cultural Elements

(1) Keyword extraction and organization

Since some specialized vocabulary in Chinese language and literature texts cannot be recognized by tools, it is necessary to manually define some special vocabulary in the context of the Chinese language to ensure the rigor of keyword extraction. At the same time, vocabulary with low relevance to the research topic is included in the filter word list to improve the accuracy of data screening. After further extraction, a word frequency list within the scope of the research topic is obtained, serving as the effective keyword word frequency list for this study, as shown in Table 2. It can be seen that these 20 keywords span the three major literary genres of poetry, fiction, and prose, balancing formal elements, thematic content, and cultural core. Corresponding keyword examples include: rhythm, homeland, and literary tradition. These keywords can serve as a keyword framework for Chinese language and literature course design, helping students construct a systematic knowledge map.

Table 2: List of valid keywords for Chinese language and literature texts (Top 20)

Serial number	Key words	Word frequency (times per million words)	Serial number	Key words	Word frequency (times per million words)
1	Poetry	858	11	Frontier	434
2	Novels	726	12	Pastoral	416
3	Prose	684	13	Love	393
4	Artistic conception	597	14	Family and country	384
5	Imagery	565	15	Loyalty and filial piety	372
6	Allusion	536	16	Chivalry	355
7	Rhetoric	518	17	Allegory	342
8	Rhythm & Rhyme	492	18	Textual criticism	326
9	Antithesis	475	19	Composition techniques	315
10	Landscape painting	453	20	Cultural lineage	302

(2) Co-occurrence analysis

The generated valid keywords were organized and processed using ROSTCM software to identify co-occurring keywords, and a valid co-occurrence matrix was generated as shown in Table 3. The data on the diagonal represents the joint frequency of the keyword appearing in all Chinese language and literature texts, while the data off the diagonal represents the joint frequency of two identical keywords appearing together in the same Chinese language and literature text. The higher the joint frequency value between two keywords, the closer the thematic connection within Chinese language and literature. Analysis reveals that Chinese language and literature centers around core keywords such as “poetry,” “prose,” and “novel,” fully reflecting the three major literary genres of Chinese language and literature. Keywords reflecting formal aesthetics, such as “aesthetic realm,” “aesthetic image,” and “rhetoric,” rank high in both frequency and co-occurrence frequency. Additionally, keywords related to natural and humanistic imagery, such as “mountains and rivers,” “rural landscapes,” and “frontier regions,” also have high frequencies, indicating a certain degree of attention to formal norms and aesthetic techniques in Chinese language and literature. Additionally, keywords representing ethical value systems in traditional culture, such as “love,” “loyalty and filial piety,” “chivalry,” and “nation and family,” have high word frequencies and co-occurrence frequencies, further highlighting the educational function of traditional cultural elements in Chinese language and literature.

Table 3: Partial list of co-word matrices

Vocabulary	1	2	3	4	5	6	7	8	9	…	19	20
Poetry	858	113	126	241	253	148	177	78	222	…	132	222
Novels	117	726	84	155	119	72	133	47	132	…	0	116
Prose	45	29	684	807	0	0	84	0	0	…	0	0
Artistic conception	63	39	0	597	0	120	0	49	107	…	0	0
Imagery	70	32	0	78	565	0	51	0	0	…	0	0
Allusion	37	18	0	116	0	536	0	66	141	…	0	0
Rhetoric	45	33	60	0	47	0	518	0	72	…	0	75
Rhythm & Rhyme	20	12	0	49	0	69	0	492	74	…	0	0
Antithesis	55	32	0	101	0	137	70	69	475	…	0	0
Landscape painting	33	18	0	0	0	0	0	0	0	…	0	0
Frontier	48	35	0	70	47	0	49	0	72	…	0	0
Pastoral	53	45	39	67	68	56	90	0	79	…	0	0
Love	32	18	39	0	98	0	0	0	0	…	0	0
Family and country	43	31	44	0	50	0	105	0	0	…	108	0
Loyalty and filial piety	29	22	0	0	0	0	0	0	0	…	0	0
Chivalry	33	21	45	82	0	100	0	0	74	…	0	0
Allegory	49	38	92	49	95	0	86	0	0	…	0	88
Textual criticism	43	23	0	77	0	83	54	0	94	…	0	0
Composition techniques	2	0	0	0	0	0	0	0	0	…	315	0
Cultural lineage	38	19	0	0	0	0	51	0	0	…	0	302

(3) Cluster Analysis

Cluster analysis, also known as cluster analysis, is a commonly used multivariate statistical analysis method in classification problems. This method categorizes data based on the strength of associations between keywords, using keywords that partially represent the research topic as the center, and divides the data into meaningful or research-worthy clusters. After cluster analysis, objects within the same cluster theoretically exhibit high cluster similarity, while the cluster correlation between different object groups is relatively low. The clusters obtained should exhibit high intra-cluster similarity and low inter-cluster similarity.

Based on the characteristics of the policy texts in this study, the data was imported into SPSS for analysis. The "sum of squared deviations" clustering method was used, with "Euclidean distance" selected as the standard for interval division. In terms of the selection of measurement standardization methods for variable conversion values, the method based on variable z-scores was chosen. The final clustering results were output in the form of a dendrogram, as shown in Figure 4.

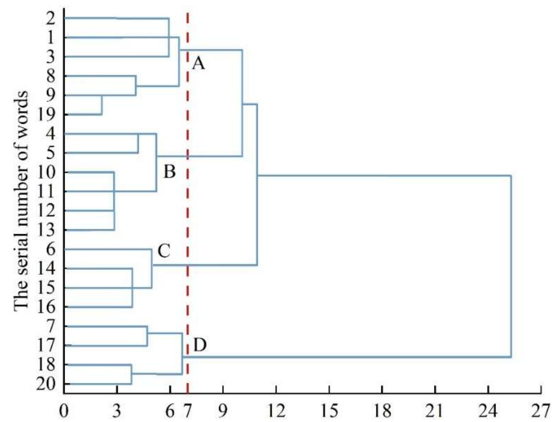

Figure 4: Thematic co-word clustering tree diagram

Clustering dendrograms can accurately map the clustering process, providing detailed information on how individual samples are merged during the clustering process. The dendrogram divides and organizes all keywords by adjusting different distances to generate different thematic clusters. The closer the distance, the more frequently the keywords in the thematic cluster appear in the same article, and the closer the relationship between the research themes represented by the clusters. Density is used to reflect the strength of connections within a thematic cluster, while centripetal force is used to reflect the strength of connections between different thematic clusters. The average number of co-occurrences of keywords within a cluster is calculated as the density of that thematic cluster, while the average number of co-occurrences of a thematic cluster with other thematic clusters is calculated as the centripetal force. As shown in Table 4, the spatial centripetal force and density of various thematic clusters are calculated as follows.

After comprehensive consideration, this study sets the clustering distance to 7, which divides all keywords into four thematic clusters, labeled A, B, C, and D. Keywords such as "poetry," "novel," "prose," "rhythm," "antithesis," and 'composition' form thematic cluster A, while keywords such as "landscape," "pastoral," "frontier," "mood," "Imagery," and "Love" form thematic cluster B. Keywords such as "Loyalty and Filial Piety," "Nation and Family," "Chivalry," and 'Allusions' are grouped into thematic cluster C due to their high interconnectivity. The remaining keywords, including "Rhetoric," "Satire," "Erudition," and "Literary Tradition," form thematic cluster D.

Table 4: Clustering results of subject words

Theme group number	The name of the theme group	Centripetal degree value	Density value
A	Literary genres and formal aesthetics	0.00	224.59
B	Natural imagery and aesthetic conception	203.85	315.28
C	Ethical narratives and social values	127.56	417.05
D	Language art and cultural analysis	0.00	684.31

(4) Sentiment Analysis

Using the proposed LIA-BiLSTM model to conduct sentiment analysis on Chinese language and literature texts, the results indicate that Chinese language and literature works contain both positive and optimistic spirits as well as profound reflections on the hardships of life. This diverse expression of emotions helps cultivate students' healthy personalities and good psychological qualities. At the same time, it also reflects a strong sense of social responsibility, with a patriotic spirit and the ideal of self-cultivation, family management, national governance, and world peace running throughout. Love is an eternal theme in Chinese language and literature.

III. B. Exploring the educational value of traditional culture in vocational education

Based on a quantitative analysis of traditional cultural elements in Chinese language and literature, this paper explores the educational value of traditional culture in vocational education as follows:

(1) It helps to enhance cultural identity and a sense of belonging.

In vocational education, when students delve into the excellent traditional Chinese culture contained in Chinese language and literature, they discover that the values and national spirit embedded within are closely connected to their own lives. When studying Qu Yuan's “Li Sao,” students can feel the poet's love and loyalty to his motherland, and they are deeply moved by his spirit of “even if I die nine times, I will have no regrets.” This understanding of patriotism in traditional Chinese culture makes students realize that they are the inheritors and guardians of Chinese culture, thereby enhancing their sense of identity with their own culture. Traditional festival cultures, such as the reunion during the Spring Festival, ancestor worship during the Qingming Festival, and moon appreciation during the Mid-Autumn Festival, reflect the Chinese nation's reverence for family ties, ancestors, and nature. When studying related literary works and folk customs, students gain a deeper understanding of the essence of these festivals, fostering a strong sense of belonging to their ethnic culture. This sense of belonging is a crucial foundation for cultural confidence.

(2) It is conducive to enhancing cultural appreciation and cultural confidence.

There are many types of literary works in Chinese traditional culture, each with its own style. From the simplicity of “The Book of Songs” to the romanticism of “The Songs of Chu,” from the splendor of Han fu to the majesty of Tang poetry, from the elegance of Song ci to the popularity of Yuan qu, to the realism of Ming and Qing novels. In the process of learning these works, students continuously improve their appreciation skills. Take Tang poetry as an example. Li Bai's poetry is bold and elegant, such as in “Bring Wine,” which shows the poet's confidence and open-mindedness: “I was born with talent that will surely be useful, and even if I spend all my money, it will come back again.” Du Fu's poetry is melancholic and full of twists and turns. In “Climbing High,” the poet's profound insight into the hardships of life is reflected in the lines, “I am a guest in this sad autumn, thousands of miles away, and I am sickly and alone on this platform.” Through analyzing the imagery, rhythm, and emotional expression of these poems, students cultivate their aesthetic perception, thereby enhancing their cultural appreciation skills. This enhancement further strengthens their belief in the value of China's excellent traditional culture, making them firmly believe that it possesses unique artistic charm and intellectual depth.

(3) It helps to stimulate cultural innovation and creativity.

Chinese excellent traditional culture has opened up a vast space for innovation for students. Many modern writers draw inspiration from Chinese excellent traditional culture for their creative work. For example, Jin Yong's wuxia novels incorporate a large number of elements from Chinese excellent traditional culture. From the chivalrous spirit of Confucianism, the concept of non-action in Daoism, to the compassionate ideology of Buddhism, all are reflected in his works. Additionally, Jin Yong innovates in narrative structure and plot design, making his stories thrilling and gripping. This innovation is based on a deep exploration of the essence of Chinese excellent traditional culture, showcasing the charm of its integration with modern creative techniques, and providing students with innovative ideas.

(4) It helps enhance cultural heritage awareness and dissemination capabilities

When students come into contact with the classics of Chinese traditional culture, they deeply realize the preciousness of this cultural heritage, which gives them a sense of responsibility to pass it on. For example, when learning about the protection and restoration of ancient books, students will learn how ancient book restorers carefully restore ancient books that have undergone the vicissitudes of time, so that precious books such as the “Yongle Encyclopedia” can be preserved, which can inspire students' awareness of passing on Chinese traditional culture at a deeper level. With the increasing frequency of international cultural exchanges, students utilize their professional knowledge to promote Chinese traditional culture worldwide through translation, teaching Chinese as a foreign language, and other means, enabling more people to understand and appreciate Chinese traditional culture. This enhances students' ability to disseminate Chinese traditional culture, and such improvements in preservation and dissemination capabilities are an outward manifestation of cultural confidence.

IV. Strategies for applying traditional culture in vocational education

In the new era, integrating the excellent traditional Chinese culture from Chinese language and literature into vocational education requires vocational colleges and universities to actively explore effective strategies. This can be approached from the following five aspects.

IV. A. Strengthen teacher training and integrate resources

Integrating China's excellent traditional culture into vocational education requires effective teacher training. First, vocational colleges can collaborate with cultural institutions to provide teachers with abundant practical resources, enabling them to gain a deeper understanding of the essence and value of China's excellent traditional culture through hands-on experience, thereby enhancing their ability to incorporate such culture into their teaching. Second, vocational colleges can implement a “Master Teacher Program” to attract outstanding professional talent, continuously enrich existing teaching resources, and provide students with higher-quality course content. Finally, vocational colleges can establish special funds to support teachers' participation in traditional culture seminars and other activities, thereby continuously enriching their knowledge reserves and improving their teaching standards. This will help to improve overall teaching quality and meet students' learning needs.

IV. B. Develop school-based courses on traditional culture and optimize course content

The system of China's outstanding traditional culture is vast, encompassing diverse content such as classical literary works, folk culture, and philosophical thought. In the process of curriculum development, vocational colleges should establish a dual-track model combining “subject integration” and “school-based specialized courses.” Specifically, teachers can incorporate classical literary content into Chinese language course instruction, while practical content such as skill inheritance, folk culture experiences, and philosophical discussions should be developed into school-based traditional culture courses. In terms of course content selection, teachers should adhere to the principle of “balancing value guidance with cognitive development,” selecting materials that cultivate students' moral character while also introducing traditional cultural topics with analytical value. In terms of implementation, teachers should base their choices on the characteristics of their students, selecting teaching resources that combine cultural depth, intellectual rigor, and practical value. Additionally, they should align with the school's unique characteristics and educational objectives to scientifically construct a teaching framework, carefully organize classroom activities, and ensure that traditional cultural knowledge is systematically and hierarchically integrated into the curriculum.

IV. C. Innovating teaching methods to stimulate students' interest in learning

Vocational schools need to create a more authentic learning environment for students, continuously innovate teaching methods, and guide students to learn about China's excellent traditional culture through various means to enhance their interest in learning.

In actual teaching, teachers should actively transform traditional teaching concepts, update existing teaching methods, and create efficient classrooms to allow students to subtly feel the charm of China's excellent traditional culture. For example, teachers can invite masters of traditional culture to give knowledge lectures, using on-site demonstrations and scenario reenactments to enhance students' learning experiences and deepen their understanding of China's excellent traditional culture. At the same time, teachers can also use classic cases and free discussions to enhance interaction between teachers and students and continuously improve students' spiritual realm. Furthermore, vocational colleges can create an “Internet Plus Chinese Excellent Traditional Culture” teaching model for students, utilizing multiple channels to achieve online and offline knowledge dissemination. Additionally, employing modern teaching methods such as virtual reality and smart classrooms to construct immersive and intelligent teaching environments can effectively improve students' learning outcomes and promote the coordinated development of their knowledge and skills.

IV. D. Integrate into campus culture construction and create a positive cultural atmosphere

Vocational colleges and universities should fully integrate China's excellent traditional culture into campus cultural development. At the material cultural level, vocational colleges and universities can create traditional culture-themed squares, set up cultural landscape features, and build digital cultural exhibition halls to make the campus environment a vivid carrier of China's excellent traditional culture. At the spiritual cultural level, vocational colleges and universities should fully leverage the communication advantages of traditional media such as campus radio, school newspapers, and bulletin boards, as well as new media platforms, to regularly conduct traditional culture-themed promotional activities and create a strong cultural atmosphere.

In terms of cultural inheritance, vocational colleges should establish a two-way mechanism of “bringing in and going out.” On the one hand, they should regularly invite intangible cultural heritage inheritors, folk craft masters, and cultural scholars to campus through master workshops, skill performances, and cultural lectures, allowing students to experience the charm of China's excellent traditional culture up close. On the other hand, organize students to participate in local folk activities, cultural festivals, and other practical experiences to deepen their cultural understanding. At the same time, focus on building traditional culture practice bases that integrate teaching, experience, and innovation, and develop traditional culture course modules with vocational education characteristics, such as traditional craft training, opera performance, and tea ceremony studies, to organically combine cultural inheritance with professional skill cultivation. Furthermore, vocational colleges should emphasize the active role of teachers and students, encouraging the establishment of traditional culture-related clubs. Through organizing cultural festivals, skill competitions, and creative design activities, enthusiasm for participating in campus cultural development can be stimulated. By constructing a cultural education ecosystem characterized by “environmental immersion—master guidance—practical experience—innovative development,” the deep integration of Chinese traditional culture with modern vocational education can be achieved, laying a solid foundation for cultivating high-quality technical and skilled talent with cultural confidence.

IV. E. Improve evaluation and incentive mechanisms to stimulate internal motivation

In the process of integrating China's excellent traditional culture into vocational education, establishing a scientific and comprehensive evaluation mechanism is a critical step in improving teaching quality. Vocational colleges and universities should move beyond a single knowledge-based assessment model and establish a comprehensive evaluation standard that encompasses three dimensions: “cognition, skills, and literacy.” In the knowledge dimension, the focus should be on assessing students' understanding of the essence of China's excellent traditional culture. In the skills dimension, the emphasis should be on evaluating their mastery of traditional techniques. In the literacy dimension, the evaluation should emphasize checking students' cultural identity and the internalization of cultural values.

Additionally, vocational colleges should establish a systematic incentive mechanism. For example, they could implement a credit bank system, incorporating traditional culture course learning, participation in practical activities, and cultural project development into the scope of credit recognition, and awarding credits to outstanding performers. Special reward funds should be established to recognize individuals and teams that excel in cultural inheritance and innovation, such as the “Intangible Cultural Heritage Inheritance Newcomer Award” and the “Traditional Culture Innovation Award.” Additionally, growth portfolios should be established to document the development trajectory of students' cultural literacy throughout their academic journey.

In terms of teacher development, vocational colleges should incorporate the effectiveness of integrating Chinese excellent traditional culture into teaching into the teacher evaluation system. Through incentive measures such as the “Teaching Innovation Award” and the “Cultural Heritage Contribution Award,” they should promote teachers' professional development. Teachers should be encouraged to participate in traditional culture training programs, and support should be provided for teaching reform research to establish a virtuous cycle mechanism of “evaluation-driven teaching and award-driven excellence.” Through the dual drivers of evaluation mechanisms and incentive mechanisms, the educational effectiveness of integrating Chinese excellent traditional culture into vocational education should be continuously enhanced.

V. Conclusion

This paper utilizes text mining technology and the text sentiment analysis model LIA-BiLSTM to quantify traditional cultural elements in Chinese language and literature texts. Based on the exploration of the educational value of traditional culture, it proposes effective strategies for integrating traditional culture into vocational education.

The accuracy of the LIA-BiLSTM sentiment analysis model outperforms all other models on most datasets, reaching a maximum of 96.89%, and the F1 score remains ahead of other models on most datasets. Additionally, compared to the CSS model and BiGRU-CNN model, the LIA-BiLSTM model achieves a higher AUC value of 0.9779, closer to 1, demonstrating superior classification performance.

Quantitative analysis of traditional cultural elements indicates that the diverse emotional expressions in Chinese language and literature works, which blend optimistic spirit with profound reflection on the hardships of life, help cultivate students' healthy personalities and good psychological qualities, and also reflect a strong sense of social responsibility. The ideals of patriotism, self-cultivation, family harmony, national governance, and world peace run throughout. This demonstrates the educational value of traditional culture in Chinese language and literature in enhancing students' cultural confidence, strengthening their sense of cultural identity, sense of accomplishment, and sense of responsibility. Based on this, this paper proposes the following five strategies for integrating traditional culture into vocational education:

(1) Strengthen teacher training and achieve resource integration.
(2) Develop school-based traditional culture courses and optimize course content.
(3) Innovate teaching methods to stimulate students' interest in learning.
(4) Integrate traditional culture into campus cultural development to create a positive cultural atmosphere.

(5) Improve evaluation and incentive mechanisms to stimulate the intrinsic motivation of vocational education.

References

[1] YANG, J. (2023). Research on the Transmission Path of Traditional Culture under the Background of Cross-cultural Communication—Taking the Short Video of Li Ziqi as an Example. DEStech Transactions on Social Science, Education and Human Science, (isss).

[2] Xianyue, L. (2019). Research progress and trend of integration of Chinese cultural industry and tourism industry: Based on catespace analysis. Economic geography, 39(12), 212-220.

[3] Jing, X. (2024). An Exploration of Traditional Chinese Cultural Programs: The Construction of Rituals, Narrative Strategies, and Experiential Insights. Contemporary Social Sciences, 9(4).

[4] Song, L. (2021). Construction of socialist core values from the perspective of Chinese traditional culture. International Journal of Frontiers in Sociology, 3(12), 69-76.

[5] Zhang, L. (2022). Chinese traditional culture education: Implementing the child's position and perspective in the elementary school textbook morality and law. ECNU Review of Education, 5(4), 702-719.

[6] Gan, L. (2023). Exploring the pathway of integrating chinese fine traditional culture into pre-school education. Frontiers in Educational Research, 6(3).

[7] Zhao, L. (2025). The Educational Function of Traditional Chinese Culture and Its Realization Path. Curriculum Learning and Exploration, 2(6).

[8] Sun, J., Zhang, Y., & Li, D. (2024). A practical study of integrating Chinese traditional culture into the research curriculum mode. Journal of Humanities, Arts and Social Science, 8(7).

[9] Tang, Y. (2024). Inheritance and Innovation of Traditional Culture in Primary School Chinese Curriculum. Journal of Modern Education and Culture, 1(1).

[10] Ganassin, S. (2019). Teaching and learning about Chinese culture: pupils' and teachers' experiences of Chinese community schooling in the UK. Language and Intercultural Communication, 19(2), 167-183.

[11] Chan, J., Woore, R., Molway, L., & Mutton, T. (2022). Learning and teaching Chinese as a foreign language: A scoping review. Review of Education, 10(3), e3370.

[12] ZHANG, Y. (2023). The Function of Inheriting Traditional Culture in Chinese Language and Literature. DEStech Transactions on Social Science, Education and Human Science, (isss).

[13] Cui, Y. (2019, August). Thought on traditional culture and college Chinese teaching. In 1st International Symposium on Innovation and Education, Law and Social Sciences (IELSS 2019) (pp. 277-281). Atlantis Press.

[14] Yu, J. (2020). On the inheritance of traditional culture and the cultivation of artisan spirit in vocational education. Advances in Educational Technology and Psychology, 5(4), 33-36.

[15] Nie, K. (2023). The Value and Application of Traditional Culture in Undergraduate Vocational Education. Transactions on Comparative Education, 5(8), 98-106.

[16] Zheng, Y. (2022). The influence of traditional culture integration into chinese language and literature teaching on the improvement of mental health of college students. Journal of Environmental and Public Health, 2022(1), 9528503.

[17] Sun, Y. (2024). Exploring the trend of stylistic evolution of chinese modern and contemporary literary works based on big data analysis. Journal of Combinatorial Mathematics and Combinatorial Computing, 124, 477-487.

[18] Yuan, Y. (2022). Analysis of fictional characters in the context of artificial intelligence and big data: taking A Dream of Red Mansions for example. Digital Scholarship in the Humanities, 37(1), 289-297.

[19] Hang, J. (2021, August). Big Data Analysis of the Formative Factors of Text Meaning in Ancient Chinese Literature of Higher Education. In EAI International Conference, BigIoT-EDU (pp. 165-174). Cham: Springer International Publishing.

[20] Li, J. (2020, March). The Influence of Chinese Traditional Culture in the Process of Teaching and Education Based on Big Data Analysis. In International Conference on Application of Intelligent Systems in Multi-modal Information Analytics (pp. 683-688). Cham: Springer International Publishing.

[21] Yuan, M. (2022, April). A study on the application of big data technology in the excavation of intangible cultural resources. In International Conference on Big Data Engineering and Technology (pp. 49-59). Cham: Springer International Publishing.

[22] Martinho, T. D. (2018). Researching culture through big data: Computational engineering and the human and social sciences. Social sciences, 7(12), 264.

[23] He, P. (2022). Chinese Character Culture Communication and Comparative Analysis Based on Big Data Visualization. Wireless Communications and Mobile Computing, 2022(1), 6371538.

[24] Zhang, Q. (2021, April). Aesthetic differences between Chinese and western cultures based on big data analysis. In Journal of Physics: Conference Series (Vol. 1852, No. 2, p. 022022). IOP Publishing.

[25] Li, J. (2021, October). The Potential of Big Data Technology in Promoting the Dissemination of Tibetan Culture in the Information Age. In International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy (pp. 489-496). Cham: Springer International Publishing.

[26] Chao Zhan, Chunyang Wang, Bibo Lu, Wei Yang, Xian Zhang & Gaige Wang. (2025). NGSTGAN: N-Gram Swin Transformer and Multi-Attention U-Net Discriminator for Efficient Multi-Spectral Remote Sensing Image Super-Resolution. Remote Sensing,17(12),2079-2079.

[27] Hainan Wang. (2024). Automatic question-answering modeling in English by integrating TF-IDF and segmentation algorithms. Systems and Soft Computing,6,200087-.

[28] Jingze Li. (2025). Analysis of International Research Hotspots on the Application of Chatbots in the Field of Psychological Counseling—Text Mining Based on LDA Topic Model Analysis. Journal of Artificial Intelligence Practice,8(2),

[29] Shouqiang Sun, Ziming Zeng, Qingqing Li & Tingting Li. (2025). A fine-grained scene retrieval model for narrative images based on multi-view feature fusion and SC-fused re-ranking. Knowledge-Based Systems,311,113126-113126.

[30] Rubaya Khatun & Arup Sarkar. (2025). Classification of triple extraction and RDF generation using boosted BERT fused attention convolutional Bi-directional LSTM with optimization. Knowledge-Based Systems,324,113756-113756.