IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 September 2024, accepted 17 September 2024, date of publication 23 September 2024, date of current version 2 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3465556

RESEARCH ARTICLE

Research on the Application of Mediation Model Based on Deep Learning in Dispute Resolution

WU DAN[1], ZUO XIANGBIN[2], AND DING HUANHUAN[3,4]
[1]Faculty of Finance and Management, Guangzhou Sontan Polytechnic College, Guangzhou, Guangdong 511370, China
[2]Faculty of Law, The National University of Malaysia (UKM), Bangi, Selangor 43600, Malaysia
[3]School of Management, Guangzhou Xinhua University, Guangzhou, Guangdong 510520, China
[4]UCSI Graduate Business School, UCSI University, Cheras, Kuala Lumpur 56000, Malaysia

Corresponding author: Zuo Xiangbin (p127253@siswa.ukm.edu.my)

This work was supported by the 2021 Scientific Research Innovation Team, Guangzhou Sontan Polytechnic College "School-Enterprise-Village Integration Rural Revitalization Scientific Research Innovation Team," under Grant GSZKYCX202104.

ABSTRACT This study explores the application and practical effects of a mediation model based on deep learning in dispute resolution. Traditional dispute resolution methods are limited by human resources and time costs, whereas deep learning technology provides new possibilities for dispute resolution with its ability to automatically extract data features and perform efficient learning and prediction. This study constructs an Attention-based LSTM (Long Short-Term Memory) model to achieve automated analysis and processing of dispute cases, aiming to improve the efficiency and accuracy of dispute mediation. Experiments show that the constructed Attention-based LSTM model can achieve 92.5% accuracy in the verification set when dealing with the dispute mediation task, and has high performance in the evaluation indexes such as recall, precision and F1 value. The model can not only accurately classify dispute cases as mediation success or mediation failure, but also capture key events and emotional tendencies in the text, providing valuable reference for the mediation process. In addition, the influence of different parameter settings on the performance of Attention-based LSTM is also discussed, and the advantages of Attention-based LSTM in dealing with long texts and complex semantic relationships are further verified through comparative experiments with other text classification models. The introduction of attention mechanism enables the model to focus on the key information in the text, thus improving the understanding ability of complex and long-length texts. This is particularly important when dealing with social conflicts and disputes, because these events often involve a lot of background information and detailed descriptions. The main contribution of our research is to develop a novel mediation model based on deep learning, which significantly improves the efficiency and accuracy of dispute classification. By automating the classification process, we aim to reduce the workload of mediators and reduce the possibility of human error in event classification.The mediation model based on deep learning shows great potential and application prospect in dispute settlement. This study not only provides a new automatic and intelligent means for dispute mediation, but also opens up a new direction for the application of deep learning in the legal field.

INDEX TERMS Dispute resolution, mediation model, deep learning, attention-based LSTM.

I. INTRODUCTION

Dispute settlement is an indispensable part of social operation, which is related to social stability, fairness and justice and the protection of individual rights and interests. Traditional dispute settlement methods, such as manual mediation, arbitration or litigation, can effectively solve disputes to some extent, but are limited by human resources, time cost and efficiency. With the rapid development of AI technology, especially the remarkable achievements of deep learning in many fields, it is possible to use intelligent methods to assist or replace traditional dispute resolution methods [1].

The current challenges in the field of dispute resolution include, but are not limited to, the surge in the number

The associate editor coordinating the review of this manuscript and approving it for publication was Hongwei Du.

of cases, the increase in complexity and the need for efficient and fair solutions. Traditional methods, such as manual review and classification, are not only time-consuming and labor-intensive, but also easily influenced by subjective judgment, resulting in slow processing speed and poor consistency. In addition, with the changes in the socio-economic situation, especially in the post-epidemic period and the changes in the global economic structure, the complexity and challenges of dispute resolution are also increasing. By developing a mediation model based on deep learning, this study aims to make use of the powerful data processing ability and pattern recognition advantages of machine learning to improve the efficiency and accuracy of dispute resolution. The model specially designed the attention mechanism to better capture and understand the key information in the text, which is very important for understanding and solving complex controversial events. By automating the classification and prediction process, our research not only aims to reduce the workload of mediators, but also hopes to improve the quality of dispute resolution by providing more objective and consistent solutions.

Deep learning, a subset of machine learning, is capable of autonomously deriving features from data and facilitating proficient learning and prediction by emulating human brain processes. Within the realm of natural language processing (NLP), deep learning has proven its robust capability for text analysis and comprehension, thereby presenting significant promise in managing disputes that involve extensive textual information [2], [3]. Through the deep learning model, the case can be analyzed more quickly, mediation suggestions can be provided, and even mediation results can be predicted, thus greatly improving the efficiency and accuracy of dispute handling. The purpose of this study is to explore the application of mediation model based on deep learning in dispute settlement and evaluate its actual effect. Through the construction and training of deep learning model, the automatic analysis and handling of dispute cases can be realized, which provides a new method and perspective for dispute mediation. This can not only reduce the workload of mediators, improve the efficiency of mediation, but also help to achieve a more just and accurate dispute settlement. Therefore, this study has important theoretical and practical significance.

II. RELATED WORK

With the rapid development of artificial intelligence technology, deep learning has been widely used in the legal field. From text classification and contract automation to legal consultation system and legal risk assessment, deep learning technology is helping legal practitioners to improve their work efficiency and enhance their decision-making ability. Recent advancements in the domain of deep learning have been notably prominent, particularly within NLP. Model architectures such as Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM), and the Transformer model have seen extensive use across diverse applications including text categorization, sentiment evaluation, and language translation, yielding outstanding results. SiMaLSTM-SNP model combines Siamese network and membrane computing, and proposes a novel semantic relevance learning model [4]. In our research, we adopt a similar structure to enhance the model's ability to capture complex semantic relations in legal texts. The embedded model of temporal knowledge map specializes in dealing with knowledge maps with long time intervals and irregular time stamps [5], which is similar to the challenges faced in dealing with historical legal cases. This study draws lessons from its method of processing time information to better understand and predict the development trend of the case. DPAL-BERT is a lighter question answering model [6], and the exploration of DPAL-BERT in improving computational efficiency is instructive for us to optimize the running time and resource consumption of the model. Stacked noise-reducing self-encoder -OCEAN, a personalized recommendation model, enhances the accuracy of recommendation through noise-reducing self-encoder. In our research, similar noise reduction techniques are also used to improve the robustness of the model to noise data [7]. These advanced frameworks possess the inherent capability to adeptly discern and derive meaningful characteristics from expansive text datasets, furnishing robust assistance for endeavors centered on textual analysis.

Text classification is the basis of the application of deep learning in the legal field, including case classification and legal provisions classification. By training the deep neural network model, legal texts can be automatically identified and classified, and the efficiency and accuracy of information retrieval can be improved. Contract automation mainly includes contract generation, contract review and contract analysis. The deep learning model can automatically extract key clauses in contracts, identify potential risks and provide decision support for legal practitioners. For example, using BiLSTM-CRF for contract risk assessment can effectively identify the key information in the contract text [8]. The legal consultation system aims to provide legal consultation services for users. Deep learning model can understand users' legal problems, provide relevant legal provisions and explanations, and recommend solutions. For example, the legal consultation system based on Large Language Model (LLMs) can provide various legal services for different user groups [9]. Legal risk assessment aims to assess the potential risks in legal activities. Deep learning model can analyze historical case data, predict the trend of cases and provide decision-making basis for lawyers. For example, using BERT model to predict the outcome of a case can improve the accuracy of legal risk assessment.

In the field of dispute resolution and law, some studies have begun to explore the application of deep learning. For example, some studies use deep learning model to classify and retrieve legal documents, which improves the efficiency of legal work [10]. There is also research on using deep learning technology to automatically abstract and extract keywords from legal cases, so that legal practitioners can understand and analyze the case more quickly. However, although deep learning is widely used in NLP and legal fields, there are still relatively few studies on applying deep learning to dispute mediation. At present, most dispute mediation still relies on manual experience and intuition, lacking data-driven support and verification [11]. Therefore, this study aims to fill this gap, and provide a more scientific, efficient and accurate method for dispute resolution by constructing a mediation model based on deep learning.

In addition, some researches focus on how to use machine learning technology to predict the outcome of legal cases [12]. By analyzing the historical case data, these studies build a prediction model in order to provide valuable reference in the process of case handling. However, these studies mainly focus on the prediction of litigation results, while the modeling and optimization of mediation process are relatively less involved.

To sum up, although deep learning and machine learning have been applied in the field of law and dispute resolution, the research on mediation model based on deep learning is still a new and challenging topic. Legal texts have been formalized, and there are many domain knowledge and concepts in deep learning. How to use legal knowledge is of great significance. The deep learning model needs to strictly follow the rules clearly defined in the law for reasoning. Deep learning model is usually regarded as a "black box" and it is difficult to explain its decision-making process. In the legal field, interpretability is as important as performance. The purpose of this study is to model and optimize the process of dispute mediation through deep learning technology, so as to improve the efficiency and success rate of mediation and bring new breakthroughs and innovations to the field of dispute settlement.

III. RESEARCH METHOD

A. A DATA COLLECTION AND PRETREATMENT

In order to build a mediation model based on deep learning, it is necessary to collect a large number of dispute handling cases as training data. These cases cover different types of disputes, such as civil disputes, commercial disputes, etc., ensuring the diversity and representativeness of cases. Through cooperation with many courts, mediation agencies and law firms, we have obtained rich data on dispute handling cases.

Taking a typical civil dispute case as an example, the collected data includes statements of both parties to the dispute, evidence materials, records of mediation process and final mediation results. These data exist in various forms such as text, pictures and audio, and a series of data cleaning and prepossessing are carried out for unified processing and analysis.Firstly, clean the text data, remove irrelevant information, such as advertisements and watermarks, and correct typos and grammatical errors in the text. Use professional tools to convert and extract the picture and audio data, and convert them into a format that the model can handle.Next, label the data. Annotation is a very important step in machine learning, which tells the model which information is important and which information can be ignored. In our research, the mediation results are mainly marked and divided into two categories: successful mediation and mediation failure. At the same time, the key events and emotional tendencies in the mediation process are marked so that the model can better understand the dynamic process of mediation. Table 1 shows the results of data annotation in civil dispute cases.

Finally, the data is preprocessed so that the model can learn and extract features better. To reduce the sparsity of text data and improve the generalization ability of the model, some operations such as word segmentation, stop words removal and stem extraction are carried out. The image and audio data are normalized and reduced in dimension to extract features that are beneficial to model training.

B. CONSTRUCTION OF DEEP LEARNING MODEL

In order to deal with the task of dispute mediation, this study chooses a deep learning model based on attention mechanism, which combines LSTM and Attention-based LSTM to adapt to the complex semantic relations and long-distance dependence in dispute texts [13]. The LSTM model based on attention mechanism combines the advantages of LSTM network and attention mechanism. LSTM network can capture long-distance dependence and is suitable for processing sequence data such as text. The attention mechanism can make the model focus on the most relevant part of the input sequence and improve the analytical accuracy. This combination makes the model especially suitable for dealing with complex legal texts rich in technical terms, so this model is chosen instead of other deep learning models in this paper.

The schematic of the deep learning architecture employed in this research is depicted in Figure 1. Initially, the input layer is fed with an embedded representation of the contested text. Subsequently, an LSTM layer processes this input to discern temporal nuances within the text. To account for long-range dependencies, the LSTM's output undergoes a weighting by a self-attention mechanism. This weighed output is then channeled through a fully connected layer, which projects it into the ultimate prediction dimensionality. The culmination of this process occurs at the output layer, where a softmax function is employed to execute a multi-classification forecast, differentiating between successful mediation and unsuccessful mediation outcomes.

The LSTM model is a distinct type of RNN. It introduces a gating system and memory cell, effectively addressing the challenges of gradient vanishing or exploding often encountered by RNNs when processing extensive sequences. The LSTM unit comprises a forget gate, an input gate, and an output gate, which collectively regulate the preservation and flow of information (Figure 2). In the context of this research, the LSTM layer serves to distill temporal attributes from disputed

TABLE 1. Data labeling result.

data type	Data content	Mediation result labeling	Key event labeling	Emotional tendency labeling
state	Plaintiff: The defendant failed to pay the goods as agreed in the contract.	Mediation succeeded.	Payment dispute	negative side
	Defendant: The goods provided by the plaintiff have quality problems.		Quality problem of goods	negative side
Evidence material	A copy of the contract showing the terms of payment.		Evidence of payment for goods	No emotion
	Photos of the goods showing the damage.		Evidence of cargo damage	No emotion
Mediation process record	Mediator: Are both parties willing to mediate?		Start mediation	neutral
	Plaintiff: Agree to mediation, but demand full payment.		Plaintiff makes a request	frontage
	Defendant: agreed to partial payment, but asked the plaintiff to solve the quality problem.		The defendant made a request	frontage
	Mediator: Can the two sides reach a consensus on payment and quality?		Seek consensus	neutral
	Plaintiff: Partial payment is acceptable, but the defendant needs to bear the maintenance costs.		Plaintiff compromise	frontage
	Defendant: agreed to bear part of the maintenance costs.		The defendant compromised.	frontage
Mediation result	The two parties reached a mediation agreement, and the defendant paid part of the purchase price and borne part of the maintenance costs.	Mediation succeeded.	come to terms	frontage

texts and to assimilate the contextual nuances present within the texts [14], [15], [16].

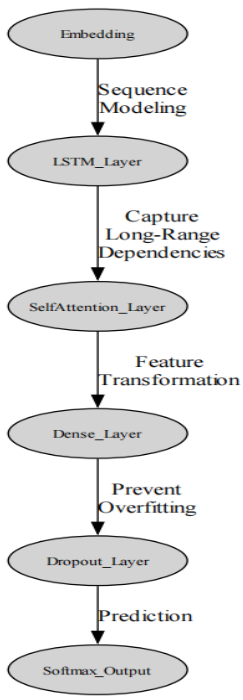

FIGURE 1. Attention-based LSTM structure.

The forgetting gate determines which information to discard from the previous state. It looks at the previous state h_{t-1} and the current input x_t, and outputs a number between 0 and 1 for each element of the previous state, indicating the proportion of information to be retained.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

where σ is the sigmoid function, W_f is the weight matrix of forgetting gate, and b_f is the bias term.

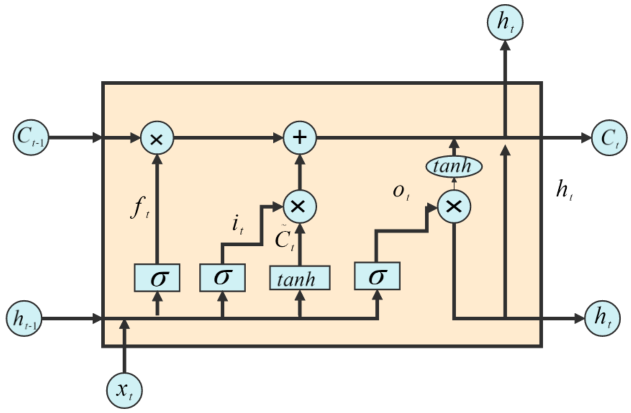

FIGURE 2. LSTM model structure.

The input gate plays a crucial role in selecting the fresh data that should be incorporated into the cell state. This selection process unfolds in two stages: initially, a sigmoid activation layer identifies the specific values slated for updating; subsequently, a tanh activation layer generates an innovative candidate value vector, which is subsequently integrated into the state [17].

$$i_t = \sigma \left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{2}$$

$$\tilde{C}_t = \tanh \left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{3}$$

where W_i is the weight matrix of the input gate, b_i is the bias term of the input gate, and W_C is the weight matrix of the candidate state, b_C is the bias term of the candidate state.

Next, the cell state C_t is updated by combining the results of the forgetting gate and the input gate. Multiply the result of the forgetting gate to "forget" some old states, and add the result of the input gate to "remember" new information.

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t \tag{4}$$

Here $\otimes$ stands for element-level multiplication.

The output gate is responsible for ascertaining the subsequent hidden state's value. This determination unfolds in a two-step process: initially, a sigmoid activation layer discerns which segments of the state value are designated for output. Following this, the cell state undergoes processing through the tanh function, wherein its value is confined to a range between -1 and 1, and then it is multiplied by the sigmoid gate's output, culminating in the derivation of the ultimate hidden state.

$$o_t = \sigma \left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{5}$$

$$h_t = o_t \otimes \tanh \left(C_t\right) \tag{6}$$

where W_o is the weight matrix of the output gate and b_o is the bias term of the output gate.

Self-attention mechanism is the core component of Transformer model, which allows the model to refer to other words in the sequence while processing one word, thus capturing long-distance dependencies (Figure 3). By calculating the correlation scores between words, the self-attention mechanism can dynamically weight different parts of the sequence, so that the model can focus on the most important information for the current word prediction. In the designed model, the self-attention layer is placed on the LSTM layer to further refine and enhance the characteristic representation of LSTM output [18], [19].

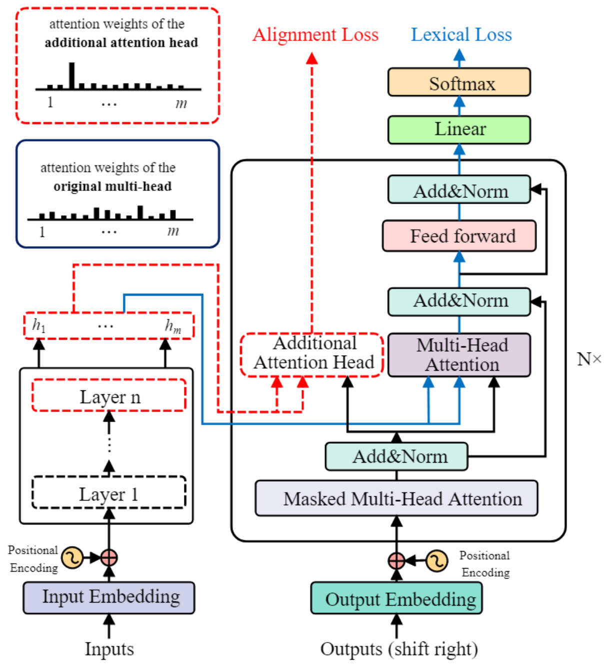

FIGURE 3. Improved transformer model.

For the self-attention mechanism, Query, Key and Value vectors need to be calculated for each input word. These are obtained from the input embedding by linear transformation.

$$Q = XW_Q, K = XW_K, V = XW_V \tag{7}$$

where X is the input embedding matrix and W_Q, W_K, W_V is the learned weight matrix.

Calculate the dot product between the query and the key, and then apply a scaling factor to get the attention score.

$$AttentionScores = \frac{QK^T}{\sqrt{d_k}} \tag{8}$$

where d_k is the dimension of the key vector. This scaling factor helps to prevent the dot product result from being too large, causing the softmax function to enter the saturation region.

This output captures the relevance of each word in the input sequence to other words, and makes a weighted representation accordingly. In our model, this self-attention output will be further processed to enhance the feature representation extracted by LSTM layer, so as to capture the long-distance dependencies in the text more accurately [20], [21].

The objective of model training is to diminish the discrepancy between the forecasted outcomes and the true labels. In this investigation, the cross-entropy loss function is employed as the objective criterion [22], and its computational formula is articulated as follows:

$$L = -\sum_{i=1}^{N} y_i \log(p_i) + (1 - y_i)\log(1 - p_i) \tag{9}$$

where N is the number of samples, y_i is the real label (0 or 1) of the i sample, and p_i is the probability that the model predicts that the i sample is positive.

In order to optimize the model parameters, Adam optimization algorithm is adopted in this study. Adam algorithm combines the ideas of Momentum and RMSprop, and adjusts the learning rate by calculating the first and second moments of the gradient, thus achieving faster and more stable convergence [23], [24]. The update rules of Adam optimizer are as follows:

Calculate gradient:

$$g_t = \nabla_\theta J(\theta_{t-1}) \tag{10}$$

Update the first moment estimate:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{11}$$

Update the second moment estimate:

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{12}$$

To correct the deviation of m_t, v_t:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{13}$$

Update parameters:

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\hat{v}_t} + \varepsilon}\hat{m}_t \tag{14}$$

where η is the learning rate, β_1, β_2 is a hyperparameter, set to 0.9 and 0.999, and ε is a very small number to prevent the divisor from being 0.

It is found that the number of hidden units in LSTM and the weight decay rate in attention mechanism have significant effects on the model performance. Appropriately increasing the number of hidden units can improve the processing ability of the model for complex legal texts, while a reasonable weight attenuation rate can help prevent over-fitting (Table 2).

TABLE 2. Parameter setting comparison.

parameter	Before optimization	After the optimization	Performance improvement
Number of hidden units	64	128	clear
Weight attenuation rate	0.001	0.0005	medium

C. EXPERIMENTAL SETUP AND EVALUATION INDEX

The experimental environment of this study is configured as follows: Ubuntu 18.04 is used as the operating system, TensorFlow is used as the deep learning framework, and the computing resources are equipped with servers of NVIDIA GPU to speed up the model training and reasoning process.

To construct a mediation model grounded in deep learning, an extensive compilation of dispute resolution cases is amassed for use as training data. This comprehensive dataset encompasses the arguments presented by both disputing parties, evidentiary documents, detailed records of the mediation proceedings, and the ultimate outcomes of the mediation. The sources of these data are varied, encapsulating text, images, and audio formats. Following a meticulous cleansing process, the refined dataset is apportioned into distinct sets designated for training, validation, and testing, with an approximate distribution of 70%:15%:15%, respectively. The training set serves the purpose of educating the model, the validation set aids in fine-tuning the model's parameters and hyperparameters, while the test set is reserved for gauging the final efficacy of the model.

The data set is provided by several judicial institutions and contains all kinds of legal documents and mediation cases in the past five years. A total of 5,000 mediation cases were collected, and each case contained complete mediation process documents and metadata. The text was cleaned, irrelevant information (header and footer, case number, etc.) was eliminated, and the terminology was unified. At the same time, key information is extracted from the text, such as case type, parties involved, mediation results, etc., and converted into a format that can be used by the model. In order to prevent some types of cases from affecting the model training too much, the proportion of various cases is balanced by random sampling to ensure that the model can learn the characteristics of different types of cases. The data set features are summarized in Table 3:

TABLE 3. Data set characteristics.

features	describe
scale	Contains 5,000 mediation cases
type	Covering contract disputes, civil disputes and other types
notes	Each case contains keywords of professional lawyers
distribution	After treatment, all types of cases are evenly distributed

In pursuit of model training, the batch gradient descent technique is employed, wherein each batch comprises a specific quantity of samples. Determining an optimal learning rate, batch size, and number of training iterations is paramount to ensuring that the model thoroughly assimilates the data's characteristics and attains commendable performance. In parallel, vigilant monitoring of the model's performance on the validation set is conducted, allowing for premature cessation of training when necessary to mitigate the risk of overfitting.

The accuracy, recall, precision and F1 score are selected as evaluation indicators, because these indicators can comprehensively reflect the performance of the model in dispute mediation tasks [25], [26]. Accuracy measures the proportion of correct prediction of the model, which is directly related to the overall performance of the model. The recall rate focuses on the ability of the model to correctly identify all relevant cases, which is very important to ensure that no important information is missed. Precision measures how many cases with positive model predictions are truly relevant, which helps to reduce false positives and improve the efficiency of mediation. F1 score is the harmonic average of recall and precision, which provides a single indicator to balance these two indicators, especially for data sets with unbalanced categories.

In the context of dispute mediation, these indicators are closely related to our goals. The goal of the research is to develop a model that can handle disputes accurately and efficiently, which means that a model that can identify all relevant cases to the maximum extent (high recall rate) and ensure the accuracy of prediction (high precision) is needed. As a comprehensive index, F1 score helps us find the best balance between recall and precision, and ensures the effectiveness and reliability of the model in practical application.

To thoroughly assess the model's efficacy, a series of evaluative metrics have been delineated:

Accuracy: This metric signifies the fraction of samples that have been accurately categorized relative to the total sample count.

Recall: This ratio contrasts the true positive instances (TP) with the aggregate of actual positive instances (TP + FN), serving as a barometer for the model's proficiency in identifying genuine positive cases.

Precision: This measure juxtaposes the true positive instances (TP) against the sum of all predicted positive cases (TP + FP), thereby gauging the model's precision in predicting positive outcomes.

F1 Score: This metric computes the harmonic mean between precision and recall, offering an integrated assessment of the model's overall performance.

IV. EXPERIMENTAL RESULT

The change of loss function is monitored during the training process of Attention-based LSTM. Using the cross entropy loss function as the optimization goal, with the increase of the number of training rounds, the losses on the training set and the verification set gradually decrease. The initial loss of the training set is 2.3, which gradually decreases with the increase of the number of training rounds, and then decreases to 0.8 in the 20th round, and then slows down, and stabilizes at about 0.5 in the 50th round. The initial loss of the verification set was 2.1, which gradually decreased with the training, reached the lowest point of 0.7 in the 30th round, and then increased slightly, so the early stop was implemented in the 35th round. As shown in Figure 4.

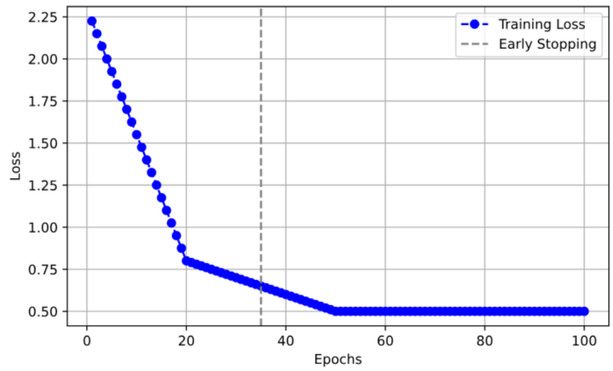

FIGURE 4. Variation of loss function.

The performance of Attention-based LSTM is evaluated on the verification set. The experimental results show that the model achieves high accuracy in the verification set. At the same time, the recall rate and precision of the model have reached a high level, which shows that the model has high sensitivity and specificity in identifying mediation success and mediation failure cases. As a comprehensive evaluation index, F1 value has reached a satisfactory level, which proves that the model has a good performance in dispute settlement. Table 4 shows the performance evaluation results of the verification set.

TABLE 4. Verification set performance evaluation results.

Evaluation index	Results (%)
Accuracy	92.5
Recall	90.0
Precision	88.0
F1 Score	89.0

Attention-based LSTM achieves 92.5% accuracy in the verification set, which means that the model can correctly classify most dispute handling cases. High accuracy is one of the important indexes to evaluate the performance of the model, which reflects the overall judgment ability of the model. The recall rate of the model on the verification set is 90.0%, which shows that the model can successfully identify 90.0% cases of successful mediation or failed mediation. The high recall rate means that the model has high sensitivity in identifying positive samples (for example, cases of successful mediation), that is, fewer false positives.

The precision of the model on the verification set is 88.0%, which means that when the model predicts a case as mediation success or mediation failure, there is an 88.0% probability that this prediction is correct. High precision shows that the model has high specificity in making predictions, that is, less false positives. On the verification set, the F1 value of the model reaches 89.0%, which is a fairly high level. A high F1 value indicates that the model maintains a high precision while maintaining a high recall rate, that is, the model is sensitive and specific in identifying cases of mediation success and mediation failure.

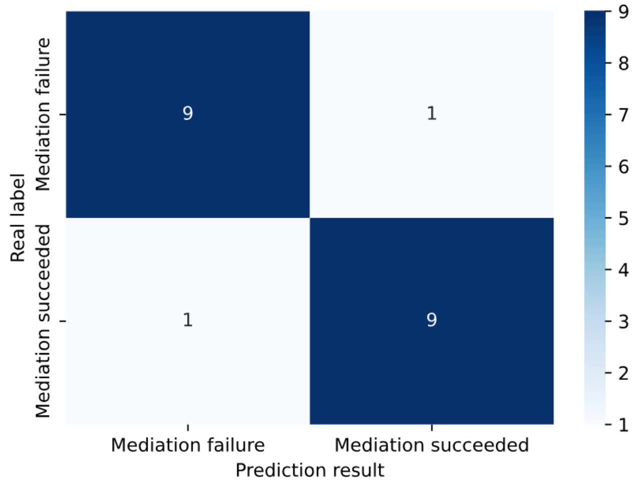

FIGURE 5. **Attention-based LSTM classification performance confusion matrix.**

From Figure 5, we can see the classification performance of the model in two categories: mediation success (labeled as 1) and mediation failure (labeled as 0). Diagonal elements of the matrix represent the number of samples correctly classified by the model, while non-diagonal elements represent the number of samples wrongly classified by the model.

In most cases, the model can correctly classify dispute cases as mediation success or mediation failure. Specifically, for successful mediation cases, the model correctly classifies the vast majority, with only a few misjudgments; Similarly, the model also shows a high classification accuracy for cases with mediation failure. Attention-based LSTM shows high performance in the task of classifying dispute mediation cases. The model can accurately identify the cases of mediation success and mediation failure, and there are few cases of misclassification. This fully shows that the model has high reliability and effectiveness in dealing with this kind of problems, and provides strong support for the automation and intelligence of dispute mediation.

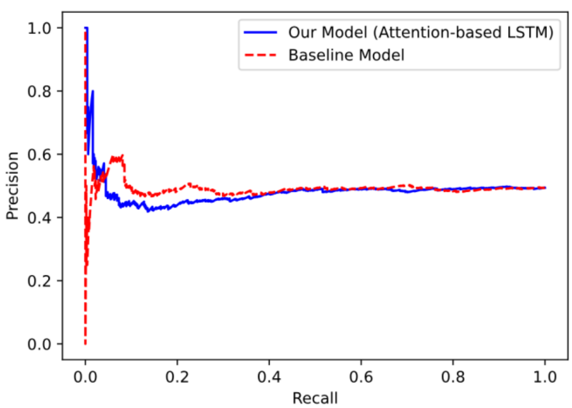

FIGURE 6. **P-R curve.**

As can be seen from Figure 6 above, the blue curve representing Attention-based LSTM keeps its accuracy at a relatively high level in most recall rates. This shows that the model has high accuracy in identifying positive samples, that is, the model can effectively select real positive samples from all samples, rather than misjudging negative samples or wrongly identifying negative samples as positive samples. Compared with the baseline model represented by the red dotted line, our model usually has higher accuracy at the same recall rate. This means that under the same ability to find positive samples, our model misjudges negative samples as positive samples less and shows stronger discrimination ability. Especially in areas with high recall rate, our model can still maintain a relatively high accuracy, which shows that the model can still maintain a good accuracy even when trying to identify more positive samples. This is very important, because in many practical applications, it is often hoped to find as many positive samples as possible while maintaining a low rate of misjudgment.

Fig. 7 clearly shows the uneven distribution of attention weights in text sequences. The color of some areas is obviously dark, which shows that the model pays more attention to these parts of the text. This uneven weight distribution reflects the selectivity and focus of the model in dealing with text. In some areas of the heat map, we can see that the weights are obviously concentrated in specific sequence positions. This may mean that the text information of these positions is very important to the decision-making process of the model. For example, in some NLP tasks, specific keywords or phrases may play a key role in understanding the meaning of the whole text, so the model will give them higher attention weight. Although the weight distribution is uneven, the change of weight between adjacent sequence positions is relatively continuous, not abrupt. This shows that the model can smoothly shift its focus when processing text, so as to understand the whole text sequence more comprehensively.

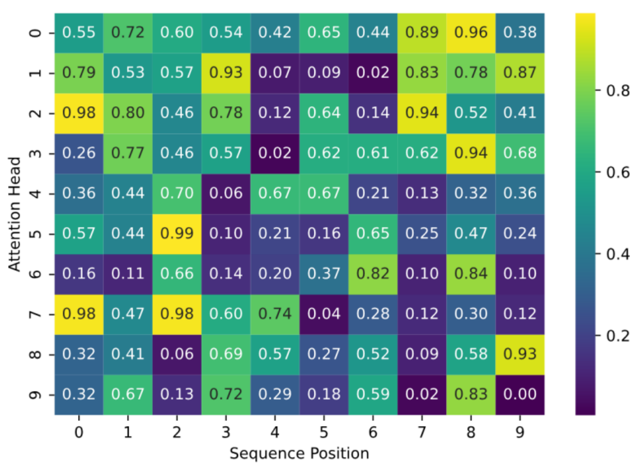

FIGURE 7. **Attention weight heat map.**

Comparing the performance under different parameter settings can help to find the best parameter combination to optimize the performance of the model. Table 5 shows the effects of different learning rates (0.001 and 0.01), batch sizes

(32, 64 and 128), hidden layer dimensions (64, 128 and 256) and discard rates (0.1, 0.2 and 0.3) on the performance of Attention-based LSTM. Performance evaluation indicators include verification set accuracy, test set accuracy and test set F1 score.

TABLE 5. Performance of under different parameter settings.

serial number	Learning rate	Batch size	Hidden layer dimension	Discarding rate	Accuracy of verification set	Accuracy of test set	F1 score of test set
1	0.001	32	128	0.2	82.3%	80.5%	0.79
2	0.001	64	128	0.2	83.1%	81.2%	0.80
3	0.001	128	128	0.2	82.8%	80.8%	0.79
4	0.01	32	128	0.2	79.5%	78.1%	0.77
5	0.01	64	128	0.2	80.2%	78.9%	0.78
6	0.01	128	128	0.2	79.8%	78.4%	0.77
7	0.001	64	64	0.2	81.5%	79.6%	0.78
8	0.001	64	256	0.2	83.7%	81.8%	0.81
9	0.001	64	128	0.1	83.4%	81.5%	0.81
10	0.001	64	128	0.3	82.6%	80.7%	0.79

The lower learning rate (0.001) is better than the higher learning rate (0.01) in most cases. This may be because the lower learning rate allows Attention-based LSTM to adjust its weight more carefully and avoid skipping the minimum value in the optimization process. When the learning rate is 0.001, the performance of Attention-based LSTM with batch size of 64 is generally better than that of models with batch sizes of 32 and 128. This shows that a moderate batch size is helpful for the model to find a better gradient descent direction in the training process, which is neither too noisy nor too smooth.

Attention-based LSTM with hidden layer dimension of 256 has the best performance, followed by 128, and 64 is relatively poor. This shows that increasing the dimension of hidden layer can provide more capacity to capture the complexity of data, but it may also increase the risk of over-fitting. However, in this study, higher dimensions seem to bring performance improvement.

Attention-based LSTM with a discard rate of 0.2 performs best in most cases, while the performance of lower discard rate (0.1) and higher discard rate (0.3) is slightly inferior. Discarding rate is a regularization technique to prevent over-fitting. Moderate discard rate is helpful for Attention-based LSTM to maintain good performance on both training set and test set. The best parameter combination tends to use lower learning rate, moderate batch size, higher hidden layer dimension and moderate discard rate.

For a more exhaustive assessment of the Attention-based LSTM's capabilities, this study also draws comparisons with an array of widely adopted text classification methodologies. These encompass traditional machine learning paradigms such as logistic regression and the support vector machine (SVM), as well as alternative deep learning frameworks, including CNN and conventional LSTM models. Empirical findings underscore that the Attention-based LSTM consistently outshines its counterparts across all evaluated metrics, particularly in the realm of managing extensive texts and intricate semantic entanglements. The comparative outcomes are illustrated in Figure 8.

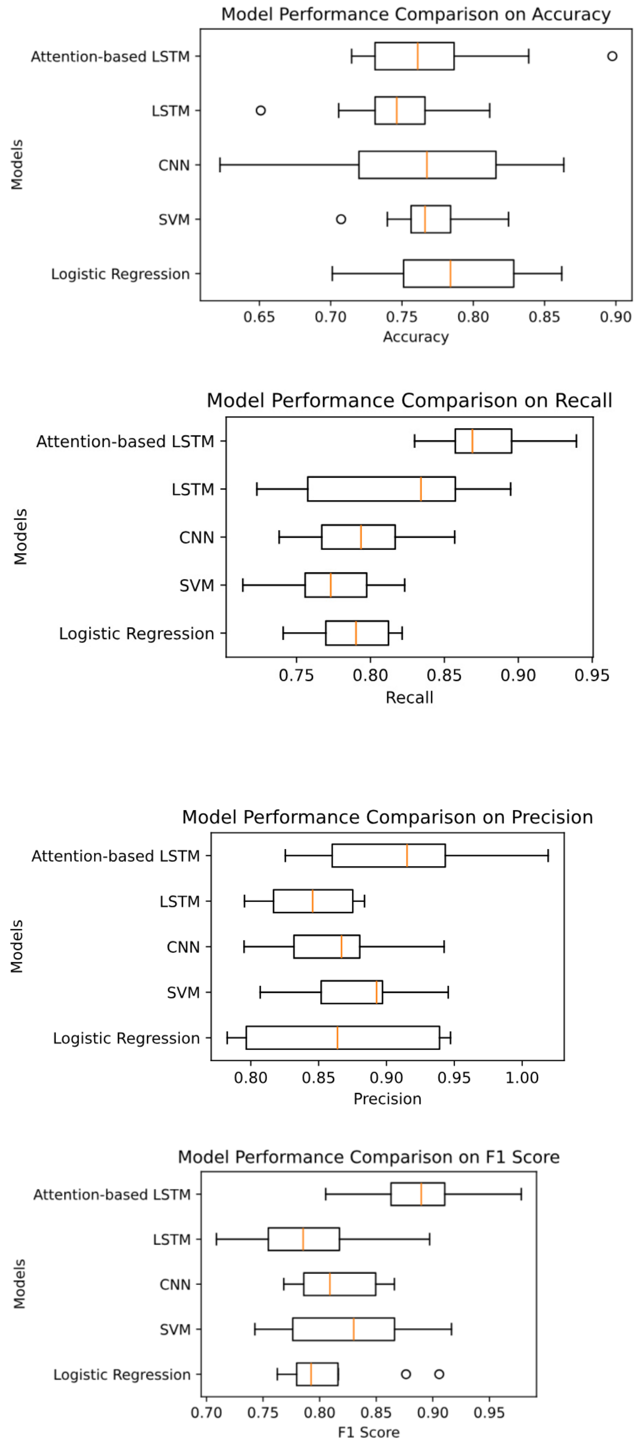

FIGURE 8. Performance distribution of different models.

In terms of accuracy, Attention-based LSTM shows a relatively higher performance level, and its box is above other models as a whole, which shows that its accuracy score is generally high and its fluctuation range is relatively small, showing better stability and superiority. In other models, the performance of LSTM and CNN is similar, and they are located in the second echelon, while the performance of logistic regression and support vector machine is relatively low.

In terms of recall rate, Attention-based LSTM is also the best, in which the range of digits and quartiles are significantly higher than other models, indicating that it can maintain a high recall rate in different experiments. The recall distribution of other models is relatively scattered and lower than that of Attention-based LSTM on the whole.

For precision, the box chart shows that Attention-based LSTM still dominates, and its performance distribution is concentrated at a high level. In contrast, although the precision of other models also has certain performance, there is a clear gap with Attention-based LSTM as a whole.

Finally, on the comprehensive index of F1 value, Attention-based LSTM shows its superiority again, and the box position is obviously higher than other models, which shows that it can still maintain excellent performance after considering the accuracy and recall. In other models, the distribution of F1 values of LSTM and CNN is similar, while the F1 values of logistic regression and support vector machine are relatively low.

Attention-based LSTM model is superior to other contrast models in all evaluation indexes, mainly because the attention mechanism enables the model to better capture the key information and complex semantic relations in the text, especially for long texts, which can effectively focus on the parts that are more important for classification decisions. The LSTM structure itself has the ability to process sequence data, and combined with the attention mechanism, it can better remember the long-term dependencies in the text, thus improving the accuracy and recall rate of classification. Attention-based LSTM shows small performance fluctuation in different experiments, which shows that it has strong generalization ability and stability.

Compared with other baseline models, the LSTM model based on attention mechanism shows significant advantages in text classification tasks, especially when dealing with long texts and complex semantic relationships. Its advantages mainly come from the effectiveness of attention mechanism, the sequential processing ability of long-term and short-term memory and the high stability of the model. These findings provide valuable reference for model selection and optimization in future text classification tasks.

V. DISCUSSION

During the training process, the change of loss function was monitored, and it was found that with the increase of training rounds, the loss in training set and verification set decreased gradually and finally stabilized at a low level. This shows that the model can effectively learn and adapt to the characteristics in the data set [27]. The accuracy, recall, precision and F1 value on the verification set have reached a high level, which proves that the model has high reliability and effectiveness in identifying mediation success and mediation failure cases. Especially, the increase of F1 value reflects that the model maintains a high precision while maintaining a high recall rate.

Attention-based LSTM model can automatically capture the key information in text sequence and give it higher weight by introducing attention mechanism. This helps the model to understand the meaning of text more accurately, thus improving the classification performance. Compared with other text classification models, Attention-based LSTM performs well in all evaluation indexes, especially when dealing with long texts and complex semantic relationships. This proves that this model has unique advantages in dealing with complex tasks such as case classification of dispute mediation [22].

By comparing the performance under different parameter settings, it is found that a lower learning rate (0.001) and a moderate batch size (64) are helpful for the model to find a better gradient descent direction in the training process, thus obtaining better performance. Higher hidden layer dimension (256) and moderate discard rate (0.2) can provide more capacity to capture the complexity of data and prevent overfitting [23]. The combination of these parameters makes the model keep good performance in both training set and test set. And the text information in some areas is very important to the decision-making process of the model, so it is given higher weight. This uneven weight distribution reflects the effectiveness and efficiency of the model in text processing.

Traditional models may lose contextual information when dealing with long texts, while LSTM model based on attention mechanism can selectively focus on different parts of the text through attention weight, thus effectively capturing long-distance dependence. When dealing with a multi-page contract dispute document, the model can accurately identify key clauses, no matter where they appear in the document. Legal texts often contain complex semantic relations, and the LSTM model based on attention mechanism can strengthen the relationship between related sentences through attention mechanism and reveal the implied logic and causality. In civil dispute cases, the model successfully identified the relevance between the two parties' dispute points, and accurately extracted the key evidence supporting mediation decision.

Legal texts usually contain technical terms and implied semantic relations, which requires the model to have a high degree of understanding. The LSTM model based on attention mechanism can better capture the long-distance dependence and complex semantic information by introducing attention weight into the hidden state, so as to understand the subtle differences in the text. Controversial cases involve a wide range of topics and forms, which requires the model to have good generalization ability. Our model can learn the language patterns of different types of cases by training on a large number of diversified legal documents, and can adapt to various new and unprecedented types of cases. Through these designs, the model can not only understand the complex legal language, but also adapt to the diversity of dispute cases, thus effectively supporting the dispute resolution process.

To sum up, the Attention-based LSTM model shows excellent performance and advantages in the task of classifying dispute mediation cases. By introducing attention mechanism and optimizing parameter settings, the model can accurately capture the key information in the text and effectively deal with long texts and complex semantic relationships. This provides strong support for the automation and intelligence of dispute mediation, and is expected to play an important role in more similar natural language processing tasks.

In the future, the research in the legal field will be carried out in many directions, including using deep learning technology to analyze contracts to automatically identify key terms and risks, training models to predict case results to assist legal decision-making, exploring the application of models in other legal fields such as intellectual property rights and labor law, and making necessary customized adjustments. At the same time, the research will cover multimodal learning to integrate various data types in legal documents, improve the interpretability and transparency of the model to meet the special needs of the legal field, and pay attention to ethical and privacy issues to ensure the protection of personal privacy and compliance with ethical standards while using big data.

In these fields, the model may encounter new challenges, such as the diversity and complexity of data, the professionalism of legal knowledge, and higher requirements for model interpretation [28]. Therefore, future research will require interdisciplinary cooperation, combining the knowledge of legal experts and the technology of data scientists to jointly promote the further development of deep learning in the legal field.

VI. CONCLUSION

In this study, the mediation model based on deep learning is successfully applied to dispute resolution. By constructing and training the deep learning model, the automatic analysis and resolution of dispute cases are realized, which provides a new method and perspective for dispute resolution. The research results show that Attention-based LSTM has obvious advantages in dealing with text classification tasks, which can maintain high recall rate and high accuracy, and effectively identify cases of successful mediation and failed mediation. In addition, by optimizing the parameters of Attention-based LSTM, it is found that lower learning rate, moderate batch size, higher hidden layer dimension and moderate discard rate are helpful for Attention-based LSTM to find a better gradient descent direction in the training process, so as to obtain better performance on the test set. Although Attention-based LSTM has achieved high performance, there are still some misclassification cases, mainly because the text expressions of some dispute cases are vague, and some cases involve complicated legal issues and facts, which are beyond the current understanding ability of Attention-based LSTM. In order to solve these problems, the future work will continue to optimize the structure and training methods of Attention-based LSTM, introduce more legal knowledge bases and case databases to enrich the background knowledge of the model, and explore more feature extraction methods and model fusion strategies to further improve the accuracy and reliability of the model. The mediation model based on deep learning in this study shows good application potential and value in dispute handling, which can not only improve the efficiency and accuracy of dispute handling, but also help to achieve more just and accurate dispute handling, and has important theoretical and practical significance.

REFERENCES

[1] W. Wang and A. Feng, "Self-information loss compensation learning for machine-generated text detection," *Math. Problems Eng.*, vol. 2021, pp. 1–7, Feb. 2021.

[2] S. Zheng, S. Chen, P. Qi, H. Zhou, and X. Yang, "Spectrum sensing based on deep learning classification for cognitive radios," *China Commun.*, vol. 17, no. 2, pp. 138–148, Feb. 2020.

[3] L. Wu, Y. Rao, H. Yu, Y. Wang, and N. Ambreen, "A multi-semantics classification method based on deep learning for incredible messages on social media," *Chin. J. Electron.*, vol. 28, no. 4, pp. 754–763, Jul. 2019.

[4] X. Gu, X. Chen, P. Lu, X. Lan, X. Li, and Y. Du, "SiMaLSTM-SNP: Novel semantic relatedness learning model preserving both Siamese networks and membrane computing," *J. Supercomput.*, vol. 80, no. 3, pp. 3382–3411, Feb. 2024.

[5] H. Huang, L. Xie, M. Liu, J. Lin, and H. Shen, "An embedding model for temporal knowledge graphs with long and irregular intervals," *Knowl.-Based Syst.*, vol. 296, Jul. 2024, Art. no. 111893.

[6] D. Alexandrov, A. Zakharova, and N. Butakov, "Does noise really matter? Investigation into the influence of noisy labels on BERT-based question answering system," *Int. J. Semantic Comput.*, vol. 18, no. 1, pp. 77–96, Mar. 2024.

[7] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1017–1027, Apr. 2017.

[8] L. Shen, C. Ma, J. Luo, and B. Hong, "An automatic classification pipeline for the complex synaptic structure based on deep learning," *J. Syst. Sci. Complex.*, vol. 35, no. 4, pp. 1398–1414, Aug. 2022.

[9] Q. Cheng, Y. Ke, and A. Abdelmouty, "Negative emotion diffusion and intervention countermeasures of social networks based on deep learning," *J. Intell. Fuzzy Syst.*, vol. 39, no. 4, pp. 4935–4945, Oct. 2020.

[10] C. Wang, X. Yang, and L. Ding, "Imbalanced sentiment classification based on sequence generative adversarial nets," *J. Intell. Fuzzy Syst.*, vol. 39, no. 5, pp. 7909–7919, Nov. 2020.

[11] X. He, L. Deng, R. Rose, M. Huang, I. Trancoso, and C. Zhang, "Introduction to the special issue on deep learning for multi-modal intelligence across speech, language, vision, and heterogeneous signals," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 3, pp. 474–477, Mar. 2020.

[12] J. Guo, "Deep learning approach to text analysis for human emotion detection from big data," *J. Intell. Syst.*, vol. 31, no. 1, pp. 113–126, Jan. 2022.

[13] O. Sharif and M. M. Hoque, "Tackling cyber-aggression: Identification and fine-grained categorization of aggressive texts on social media using weighted ensemble of transformers," *Neurocomputing*, vol. 490, pp. 462–481, Jun. 2022.

[14] R. Ghasemi, S. A. Ashrafi Asli, and S. Momtazi, "Deep Persian sentiment analysis: Cross-lingual training for low-resource languages," *J. Inf. Sci.*, vol. 48, no. 4, pp. 449–462, Aug. 2022.

[15] F. Assef, M. T. Steiner, P. J. S. Neto, and D. G. de Barros Franco, "Classification algorithms in financial application: Credit risk analysis on legal entities," *IEEE Latin America Trans.*, vol. 17, no. 10, pp. 1733–1740, Jun. 2019.

[16] R. L. Thom and A. J. Cronin, "Legal and regulatory challenges for emerging regenerative medicine solutions for diabetes," *Transplantation*, vol. 108, pp. 1072–1079, Sep. 2023.

[17] X. Guo, H. Zhang, L. Ye, and S. Li, ''Learning users' intention of legal consultation through pattern-oriented tensor decomposition with bi-LSTM,'' *Wireless Commun. Mobile Comput.*, vol. 2019, pp. 1–16, Mar. 2019.
[18] L. Liu and S.-B. Tsai, ''Intelligent recognition and teaching of English fuzzy texts based on fuzzy computing and big data,'' *Wireless Commun. Mobile Comput.*, vol. 2021, no. 1, pp. 1–10, Jan. 2021.
[19] J. Lee, H. Kim, T. Kang, and S. Yook, ''Scanner model classification with characteristic brightness variations,'' *J. Forensic Sci.*, vol. 67, no. 5, pp. 2055–2061, Sep. 2022.
[20] X. Yuan, ''Emotional tendency of online legal course review texts based on SVM algorithm and network data acquisition,'' *J. Intell. Fuzzy Syst.*, vol. 37, no. 5, pp. 6253–6263, Nov. 2019.
[21] Z. Bai, G. Sun, H. Zang, M. Zhang, P. Shen, Y. Liu, and Z. Wei, ''Identification technology of grid monitoring alarm event based on natural language processing and deep learning in China,'' *Energies*, vol. 12, no. 17, p. 3258, Aug. 2019.
[22] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, ''Sentiment analysis based on deep learning: A comparative study,'' *Electronics*, vol. 9, no. 3, p. 483, Mar. 2020.
[23] Z. Cen and J. Wang, ''Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer,'' *Energy*, vol. 169, pp. 160–171, Feb. 2019.
[24] F. Lv, W. Wang, Y. Wei, Y. Sun, J. Huang, and B. Wang, ''Detecting fraudulent bank account based on convolutional neural network with heterogeneous data,'' *Math. Problems Eng.*, vol. 2019, no. 1, pp. 1–11, Jan. 2019.
[25] J. Wang, Z. Chen, Y. Qin, D. He, and F. Lin, ''Multi-aspect co-attentional collaborative filtering for extreme multi-label text classification,'' *Knowl.-Based Syst.*, vol. 260, Jan. 2023, Art. no. 110110.
[26] B. Parlak and A. K. Uysal, ''A novel filter feature selection method for text classification: Extensive feature selector,'' *J. Inf. Sci.*, vol. 49, no. 1, pp. 59–78, Feb. 2023.
[27] S. Malik and S. Jain, ''Deep convolutional neural network for knowledge-infused text classification,'' *New Gener. Comput.*, vol. 42, no. 1, pp. 157–176, Mar. 2024.
[28] N. K. Dahlan, M. A. S. Azman, R. Rajamanickam, and M. Z. M. Zahir, ''Online mediation: Issues, applications and challenges,'' *Asian J. Res. Educ. Social Sci.*, vol. 5, no. 3, pp. 81–92, 2023.

WU DAN was born in China, in 1983. She received the master's degree in civil and commercial law from Hunan Normal University. She is currently pursuing the degree with the Department of Law, National University of Malaysia. She is also an Associate Professor of law at Guangzhou Sontan Polytechnic College. She has studied in the fields of ADR and labor law.

ZUO XIANGBIN was born in Xinxiang, Henan, China, in 1988. He received the master's degree from South China University of Technology, China. He is currently pursuing the degree with the Faculty of Law, The National University of Malaysia, Malaysia. His research interests include commercial law and international economic law.

DING HUANHUAN was born in Xuchang, Henan, China, in 1986. She received the bachelor's degree in computer science and technology and the master's degree from Guangdong University of Finance and Economics, China. She is currently pursuing the Ph.D. degree with the UCSI Graduate Business School, UCSI University, Malaysia. She is also a Lecturer with the School of Management, Guangzhou Xinhua University. Her research interests include business and management.