Multi-Task BERT–BiLSTM–CNN–MHA Framework for Legal Sentiment Analysis and Judgment Prediction

Hongqi Liu
School of Marxism, Henan Vocational Institute of Arts, Zhengzhou, 451464, China
E-mail:13903860568@163.com

Keywords: legal text, sentiment analysis, legal judgment prediction, BERT-BiLSTM model, multi-task learning framework

Received: June 3, 2025

This paper proposes a comprehensive multi-task learning (MTL) framework that integrates BERT, Bidirectional Long Short-Term Memory (BiLSTM), Convolutional Neural Network (CNN), and Multi-Head Attention (MHA) to jointly perform legal text sentiment analysis (SA) and predict judicial judgments. The architecture utilizes a shared BERT–BiLSTM encoder to generate contextualized representations, which are then extracted by a CNN module to capture fine-grained sentiment features. Meanwhile, an MHA module fuses information and guides the multi-task output for decision-related subtasks. To ensure effective joint optimization, a weighted loss function is designed to balance task-specific objectives and prevent dominance by any single task. Experiments are conducted on the public CAIL2018 dataset using ten-fold cross-validation to guarantee fairness and reproducibility. In this framework, BERT first encodes legal documents into deep semantic embeddings, which BiLSTM then processes to capture sequential dependencies. The CNN subnetwork extracts localized emotional features from the BiLSTM outputs, achieving 98.0% accuracy and an F1 score of 0.96 on sentiment classification (CAIL-big subset). Simultaneously, the MHA module leverages the shared encoder outputs to highlight and weight relevant legal clauses, supporting downstream tasks such as legal article recommendation and sentencing estimation. The recommendation task achieves 94.0% top-1 accuracy, and sentencing regression achieves a mean squared error (MSE) of 0.028. Compared with traditional baselines such as Word2Vec–BiLSTM and single-task CNN models, the proposed BERT–BiLSTM–CNN–MHA framework delivers over 5% improvement in both sentiment analysis and judgment prediction tasks. Its modular design, deep semantic representation, and robust empirical results validate its effectiveness and practical value for deployment in intelligent legal decision-support and judicial assistance systems.

Povzetek: Prispevek obravnava pravno informatiko. Uvede večopravilni BERT–BiLSTM–CNN–MHA za analizo sentimenta in napoved sodb na CAIL2018; SM arhitektura deli koder, CNN zajame lokalne emocije, MHA usmerja pravne podnaloge.

1 Introduction

With the continuous advancement of natural language processing technology, intelligent applications in the legal field have gradually attracted widespread attention [1-2]. Legal texts are highly professional, have complex grammatical structures, and have unique terminology [3-4], making it difficult for traditional SA (sentiment analysis) and judgment prediction methods to handle these characteristics effectively. In traditional SA, these methods are based on bag-of-words models, which shallow models ignore, resulting in biased analysis results. Judgment prediction fails to utilize the emotional information in court statement texts fully, and in complex cases, the model lacks intelligent weighing of legal terms, morality, and legal responsibility. As the demand for legal text processing increases, effectively improving the accuracy of SA and intelligently predicting judgment results has become a key issue that needs to be addressed in the current legal field.

This study aims to address the problems existing in traditional legal text SA and legal judgment prediction methods, specifically the lack of polysemous word processing and new word understanding, as well as the lack of intelligent trade-offs in SA for legal judgments. This paper adopts BERT-BiLSTM as the basic shared model, combines it with CNN (Convolutional Neural Networks) for sentiment feature extraction, and uses MHA (Multi-Head Attention) to handle legal judgment-related tasks. During the research process, the BERT-BiLSTM model was first constructed to generate context-related text representations, followed by the use of CNN to extract sentiment information, and finally MHA was used to recommend legal provisions, predict convictions, and predict sentences. The experiment

employed a joint training MTL (multi-task learning) framework. The multi-task learning (MTL) framework shares the underlying encoder, allowing sentiment

analysis and judgment prediction tasks to share semantic representations in the feature extraction stage and achieve information complementarity. Sentiment analysis helps reveal the attitude tendencies of court statements and assists in making more accurate judgments in complex contexts. Compared with a single task, the MTL framework can capture legal semantics more comprehensively and improve overall performance. This showed that the experimental model achieved significant performance improvements in legal text SA and legal judgment prediction, outperforming the traditional Word2Vec model in both accuracy and F1 value. The experimental results show the potential of this method in the legal field, which can enable more intelligent legal judgment prediction and informed moral trade-offs.

Research question: This paper explores whether the emotion-aware multi-task learning framework can significantly improve the sentencing prediction and sentiment classification performance of legal texts. It also verifies the effect of introducing courtroom statement emotion features in judicial judgment prediction on the accuracy and comprehensive index (F1) of sentencing results.

Expected contributions: (1) Propose the emotion-aware MTL hypothesis - for the first time, jointly model the legal text sentiment analysis and sentencing prediction tasks in the same framework to verify the gain of emotion features on judgment prediction; (2) Construct a modular BERT-BiLSTM-CNN-MHA framework - design three components: shared encoder, multi-channel convolution, and multi-head attention to achieve efficient fusion of deep context and fine-grained emotion; (3) Provide rigorous empirical evidence based on CAIL2018 - through ten-fold cross-validation, quantify the performance improvement of emotion-aware MTL compared with Word2Vec-BiLSTM, single CNN and other SOTA.

2.Related Work

In the field of sentiment analysis (SA) of legal texts, numerous researchers have explored pre-trained language models and deep learning techniques, yielding a range of empirical results. Royyan et al. employed the Word2Vec model to conduct SA on legal public policies, achieving a classification accuracy of 78.99% [5]. Anand et al. combined 1D-CNN with LSTM to capture the semantic features of legal documents, reporting promising Rouge scores [6]. The GloVe (Global Vectors for Word Representation) model has also been widely adopted in SA matching tasks, contributing to improvements in overall classification accuracy [7]. In addition, scholars such as Bello and Khan have applied the classic BERT (Bidirectional Encoder Representations from Transformers) model to legal texts, enhancing contextual understanding of semantic meaning [8–9]. While Word2Vec and GloVe-based models laid the foundation for sentiment representation in legal documents, their performance remained suboptimal due to limitations in modeling deep contextual semantics.

With the advancement of deep learning, its application in legal text SA has expanded. For instance, Abimbola et al. utilized a CNN-LSTM hybrid model for sentiment classification in legal texts, achieving an accuracy of 98.05% [10]. Rajapaksha et al. proposed the Sigmalaw PBSA (Party-Based Sentiment Analysis) model for legal opinion documents, improving task-specific performance [11]. More recent studies have leveraged architectures such as CNN-LSTM [12–13], LSTM [14–15], GRU [16], RoBERTa [17], and Conv-BiLSTM-Frog Leap [18], demonstrating notable improvements in sentiment classification accuracy. These models effectively enhance the representation of syntactic and sequential features but often fail to fully extract and integrate deep semantic structures and fine-grained emotional features in legal language.

Parallel to SA research, recent studies have clarified that judicial judgment prediction typically includes three subtasks: legal article recommendation, conviction prediction, and sentence prediction [19]. For instance, Sengupta et al. applied the SVM (Support Vector Machine) model to predict applicable legal statutes from case reports, obtaining an F1-score of 0.75 [20]. Alghazzawi et al. employed a CNN-LSTM model for court decision prediction, reaching an accuracy of 92.05% [21]. Similarly, Esan and colleagues employed traditional machine learning models, such as Random Forest (RF), achieving an accuracy of around 72% in judicial decision systems [22]. While these models demonstrated particular effectiveness, they primarily relied on legal provision text and neglected the integration of sentiment signals from court narratives. Consequently, they could not model the nuanced interplay between legal reasoning and moral sentiment, which is crucial for intelligent and human-aligned legal decision-making.

A summary of the key studies and their performance comparisons is provided in Table 1.

Table 1: Summary table of the study

References	Models	Datasets	Achieved Metrics	References	Models	Datasets	Achieved Metrics
[5]	Word2Vec	Self-built dataset	Classification	[14]	LSTM	Self-built	Accuracy

			accuracy (78.99%)			dataset	(91.9)
[6]	1DCNN 和 LSTM	Self-built dataset	Rouge-1 (0.436)	[15]	LSTM	Twitter	Accuracy (94.7%)
[7]	GloVe-HeBiLSTM	emotions-dataset-for-NLP dataset	Accuracy (93.70%)	[16]	TS-GRU	IMDB	Accuracy (90.85%)
[8]	BERT-CNN	Self-built dataset	Accuracy (93%)	[17]	RoBERTa	Self-built dataset	Accuracy (92.35%)
[9]	BERT	Self-built dataset	F1 score (81.49%)	[18]	Conv-BiLSTM-Frog Leap	Twitter	Accuracy (97.5%)
[10]	CNN-LSTM	Self-built dataset	Accuracy (98.05%)	[19]	-		
[11]	Sigmalaw PBSA	SigmaLaw-ABSA dataset	Accuracy (0.7086)	[20]	SVM	Self-built dataset	F1 (0.75)
[12]	CNN-LSTM	Self-built dataset	Accuracy (92.5%)	[21]	CNN-LSTM	Self-built dataset	Accuracy (92.05%)
[13]	CNN-LSTM	Reddit dataset	F1 (0.84)	[22]	RF	Self-built dataset	Accuracy (72%)

Notably, recent works published in Informatica have further advanced sentiment analysis (SA) and aspect-based sentiment analysis (ABSA) techniques. For instance, an enhanced RoBERTa-based framework has demonstrated superior performance on aspect-level SA benchmarks [37]. A comprehensive review has summarized various sentiment analysis methods on social media platforms [38], and specialized models have been developed for ChatGPT tweet sentiment classification using classical machine learning algorithms [39]. These studies underscore the growing importance of deep contextual understanding in sentiment-driven applications.

As summarized in Table 1, previous research has achieved noteble results in isolated tasks, such as semantic analysis, sentiment classification, and judgment prediction of legal texts, using models like Word2Vec, GloVe, CNN–LSTM, and RoBERTa. However, these approaches primarily rely on static or shallow word vector representations, limiting their ability to capture the nuanced semantic relationships inherent in legal discourse. Moreover, the association between sentiment cues and legal decision-making has

often been overlooked. Existing single-task learning setups fail to realize inter-task information sharing and collaborative optimization, which may lead to conflicts or bottlenecks when handling multiple interconnected subtasks.

To address these limitations, this paper proposes a multi-task learning (MTL) framework built upon a shared BERT–BiLSTM encoder, a CNN-based sentiment feature extractor, and a multi-head attention (MHA) module. This design enables the integration of deep contextual embeddings with fine-grained emotional features while a weighted joint loss function balances the training objectives across tasks. The proposed approach enhances the model's capacity to comprehend the complex semantics and sentiment tendencies of legal language. Furthermore, it enables the collaborative improvement of subtasks such as legal article recommendation, conviction prediction, and sentencing estimation, thereby effectively overcoming the performance and interpretability limitations observed in prior single-task and shallow modeling approaches.

3 Legal text SA modeling and legal decision prediction model

3.1 BERT-BiLSTM basic shared model

3.1.1 BERT model

BERT is a pre-trained language model based on the Transformer architecture, which understands text through bidirectional contextual information [23-24]. During the pre-training phase, BERT employs the MLM (Masked Language Modeling) task, where it randomly masks some words in the input and infers these masked words based on the surrounding context to capture more comprehensive semantic information. BERT utilizes the NSP (Next Sentence Prediction) task to learn the relationship between sentences, enabling it to exhibit strong migration capabilities across multiple natural language processing tasks.

In this paper, BERT is used as a basic shared model to undertake the task of deep semantic understanding of legal texts. The input of the BERT model consists of three parts: word vector representation, sentence representation, and position representation. The input representation of BERT is shown in formula (1).

$$B_i=B1_i+B2_i+B3_i \quad (1)$$

$B1_i$ represents word vector representation, $B2_i$ represents sentence representation, and $B3_i$ represents position representation.

BERT's encoder is based on a multi-layer Transformer structure and utilizes a self-attention mechanism for bidirectional modeling [25-26]. Based on this structure, BERT can effectively capture the complex dependencies between words. In each Transformer block, the updated calculation formula for the input word representation is shown in formula (2).

$$C_i^{new}=LN(C_i+A(C_i,C_i,C_i)) \quad (2)$$

LN represents the layer normalization operation, and C_i represents the input word representation.

In BERT, the pre-training process consists of two main tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). The goal of MLM is to maximize the likelihood probability of formula (3).

$$L_M=-\sum_{i=1}^{n} logP(x_i|x\backslash i) \quad (3)$$

$x¥i$ represents the part of the input without word x_i, and $P(x_i|x¥i)$ represents the conditional probability based on the context.

The NSP task is used to learn the relationship between sentences, and the objective function is shown in formula (4).

$$L_N=-\sum_{i=1}^{n} logP(y1_i|S_1,S_2) \quad (4)$$

S_1 and S_2 represent two sentences.

For the input legal text, after passing through the BERT model, the context representation of each word is generated, and then it is passed to the subsequent CNN module for SA and to the MHA module for legal judgment prediction. During the fine-tuning process, all BERT parameters can be optimized according to the task loss allowing the model to yield better results on subsequent tasks.

3.1.2 BiLSTM model

This paper introduces the BiLSTM (Bidirectional Long Short-Term Memory) model [27-28] to enhance further the contextual representation generated by BERT. In the forget gate, the formula is shown in formula (5) [29-30].

$$e_t=\sigma(W_e\cdot[h_{t-1},z_t]+\alpha_e \quad (5)$$

z_t represents the input vector, W_e represents the weight matrix of the forget gate, α_e represents the bias vector.

The input gate is as shown in formula (6).

$$j_t=\sigma(W_j\cdot[h_{t-1},z_t]+\alpha_j \quad (6)$$

The candidate state is shown in formula (7).

$$\widehat{D}_t=tanh(W_D\cdot[h_{t-1},z_t]+\alpha_D \quad (7)$$

*tan*h represents the hyperbolic tangent activation function.

The output gate is shown in formula (8).

$$q_t=\sigma(W_q\cdot[h_{t-1},z_t]+\alpha_q \quad (8)$$

The cell state is shown in formula (9).

$$D_t=e_t\cdot D_{t-1}+j_t\cdot\widehat{D}_t \quad (9)$$

The hidden state is shown in formula (10).

$$h_t = q_t \cdot tanh(D_t) \quad (10)$$

h_t represents the hidden state.

The BiLSTM model performs a merge operation after calculating the hidden states using the forward and backward LSTMs at each time step. The bidirectional hidden state representation is shown in formula (11).

$$\hat{h}_t = [h_t^1, h_t^2] \quad (11)$$

3.1.3 BERT-BiLSTM hybrid model

In this study, the hybrid model of BERT and BiLSTM combines the word embedding generated by BERT pre-training with the bidirectional temporal modeling capability of BiLSTM to form a powerful feature extractor and context modeling module. The hybrid steps are as follows:

1) The input legal text is encoded using the BERT model to generate context-dependent representations of each word. BERT captures long-range dependencies in the text through the self-attention mechanism, particularly in professional contexts and complex syntactic structures within the legal field, providing rich contextual information.

2) The embedded representation generated by BERT is used as the input of BiLSTM further to model the time series characteristics of the text, capture the dependencies from left to right and from right to left, and ensure that the context and sequential information in the text are fully understood.

The BERT-BiLSTM hybrid model is shown in Figure 1.

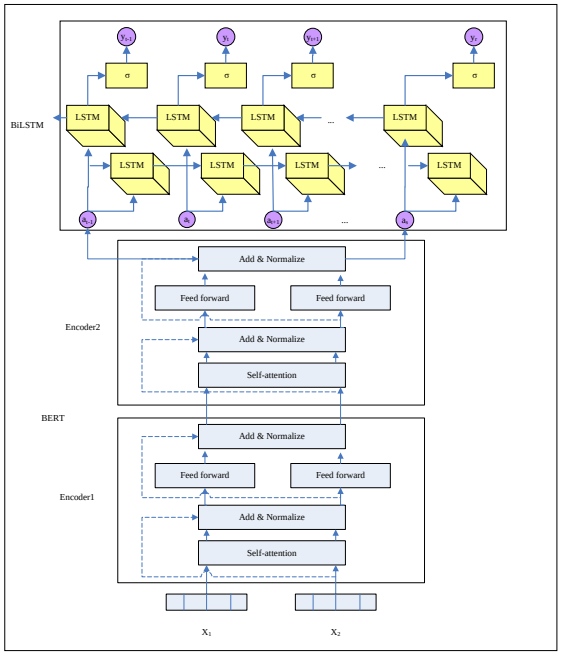

Figure 1: BERT-BiLSTM hybrid model

In Figure 1, it can be seen that BERT and BiLSTM are connected in series. The output of the BERT model passes through a linear transformation layer, adjusted to match the BiLSTM input dimension, and then sent to the BiLSTM network for processing. BiLSTM further integrates the context information provided by BERT, captures more fine-grained semantic and structural information, and outputs a bidirectional context representation of each word. The context vector output by BiLSTM is combined with the word vector output by BERT through weighted summation to form a global semantic representation as input for SA and legal judgment prediction.

3.2 CNN Sentiment Feature Extraction and Classification

This paper introduces the CNN model [31-32] to further extract sentiment features based on the text representation processed by the BERT-BiLSTM model. CNN extracts local sentiment features from word vectors to help the model capture the sentiment tendency information in the text.

In the model architecture design of this study, CNN is placed after BERT-BiLSTM for sentiment feature extraction. The reason is that the BERT-BiLSTM module can fully capture the deep semantics and temporal information of the legal text, providing rich and semantically complete input for subsequent sentiment recognition. Moreover, compared with using CNN as a preprocessing or parallel path, local feature extraction based on BERT-BiLSTM output helps to enhance further the local recognition capability of the sentiment dimension on the premise of modeling global semantics, realize the complementarity of context and sentiment features, and improve the accuracy of sentiment classification and the emotion perception ability of judgment prediction.

To effectively extract sentiment features, this paper employs a CNN architecture comprising multiple convolutional layers and pooling layers. Each convolution operation captures the sentiment information in the text from a different perspective. In the convolutional layer, multiple filters are used for processing. The calculation of the convolution operation is shown in formula (12).

$$o_i^{(k)} = g(W_k \cdot H_i + \beta_k) \quad (12)$$

$o_i^{(k)}$ represents the convolution feature, H_i represents the context vector output by BiLSTM and the word vector output by BERT. g represents the ReLU nonlinear activation function, W_k represents the convolution kernel, and β_k represents the bias term.

The formula for the pooling operation is shown in formula (13).

$$r_k = \max_{i=1,\ldots,n-w+1} (o_i^{(k)}) \quad (13)$$

r_k represents the pooled output of the convolution kernel, and w represents the size of the convolution kernel.

To further enhance the diversity of features, this model employs multi-channel convolution operations, which convolve the input text with convolution kernels of varying sizes to capture emotional features at different granularities. The emotional features extracted by each convolution kernel can form a separate channel after pooling, and the feature channels can be concatenated. The multi-channel output after pooling is shown in formula (14).

$$R=[r_1,r_2,...,r_m] \quad (14)$$

m represents the number of convolution kernels, and r_m represents the output of each convolution kernel.

After extracting the emotional features, this paper can further process them using a fully connected layer. The function of the fully connected layer is to map local features to the global space and generate emotional representations with a higher level of abstraction. The formula of the fully connected layer is shown in formula (15).

$$y_2=\text{softmax}(W_\gamma \cdot R+\delta_\gamma) \quad (15)$$

Among them, W_γ and δ_γ represent the weight matrix and bias term of the fully connected layer respectively.

3.3 MHA Judgment Prediction

3.3.1 Feature fusion

In legal judgment prediction, there are three subtasks, including the recommendation of relevant laws, conviction prediction, and sentence prediction. The emotional factors in legal texts have a significant impact on various types of cases, including criminal cases. This paper uses the results after CNN processing as part of the MHA input. The study first merges the context representation generated by BERT-BiLSTM with the emotional features extracted by CNN. The comprehensive representation obtained by splicing the two features is shown in formula (16).

$$X_{in}=concat(H,y2) \quad (16)$$

H represents the context vector output by BiLSTM and the word vector output by BERT.

3.3.2 MHA mechanism

The MHA mechanism captures multi-scale relationships in text by computing multiple different attention heads in parallel [33-34]. Each attention head focuses on different parts of the input data according to different attention weights, capturing different levels of temporal dependencies. The multi-head attention mechanism uses a linear transformation to output queries, keys, and values. The expression of the query is shown in formula (17). The expressions of the keys and values are shown in formulas (18) and (19), respectively.

$$Q_t=X_{in}W_Q \quad (17)$$

$$K_t=X_{in}W_K \quad (18)$$

$$V_t=X_{in}W_V \quad (19)$$

W_Q represents the query weight matrix, W_K and W_V represent the key and value weight matrices respectively.

The calculation formula of attention weight is shown in formula (20).

$$A_{head}=\text{softmax}(\frac{QK^T}{\sqrt{\zeta_{head}}}) \quad (20)$$

ζ_{head} represents the dimension of each attention head.

The weighted output of each head is shown in formula (21).

$$U_{head}=A_{head}V \quad (21)$$

In multi-head attention, the weighted sum of multiple attention heads is calculated in parallel, and the final output representation is obtained through linear transformation, as shown in formula (22).

$$U=Concat(U_{head_1},U_{head_2},...,U_{head_\eta})W_R \quad (22)$$

Concat represents the concatenation operation, η represents the number of independent attention heads, and W_R represents the output weight matrix.

3.3.3 Task-specific output layer design

The features generated by MHA are fed into three different task-specific output layers; each subtask includes an independent, fully connected layer to generate the final prediction result.

1) Recommending relevant laws

The experiment recommends relevant laws based on the intention expressed in the legal text. The design of the fully connected layer is shown in formula (23).

$$y_3=softmax(W_{law}U+\theta_{law}) \quad (23)$$

2) Conviction prediction

The design of the fully connected layer for the conviction prediction task is shown in formula (24).

$$y_4=softmax(W_{con}U+\theta_{con}) \quad (24)$$

(3) Sentence prediction

The design of the fully connected layer for the conviction prediction task is shown in formula (25).

$y_5=softmax(W_{sen}U+\theta_{sen})$(25)

4.MTL Framework

4.1 Construction of MTL Framework

This study utilizes the BERT-BiLSTM shared model as its basis, introduces feature extraction and prediction modules for CNN and MHA, and constructs an MTL framework to simultaneously address the SA of legal texts and legal judgment prediction tasks. The framework is constructed as follows: First, the BERT-BiLSTM model is used to extract the contextual representation of the text, providing shared basic features for SA and legal judgment prediction.

The MTL framework is shown in Figure 2.

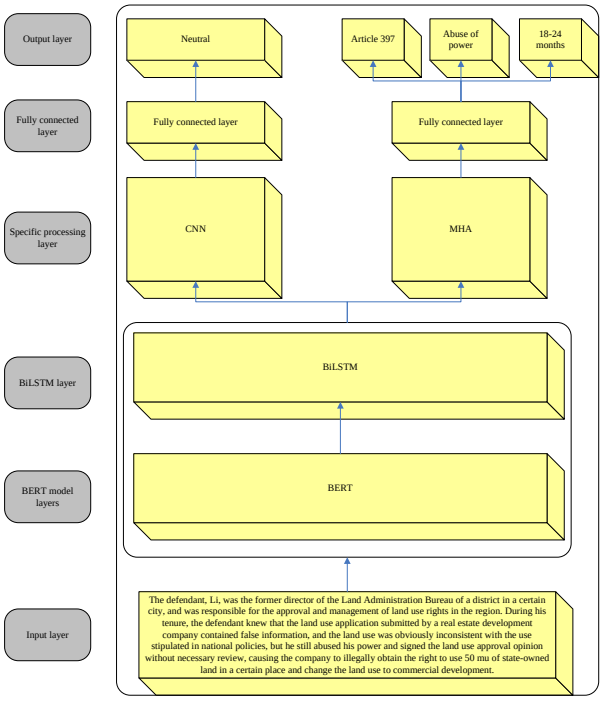

Figure 2: MTL framework

In the sentiment analysis (SA) of legal texts, Convolutional Neural Networks (CNNs) effectively extract and classify sentiment features by stacking convolutional and pooling layers to capture local semantic patterns. In the proposed multi-task learning (MTL) framework, the subtasks include legal text sentiment analysis, relevant law recommendation, and conviction and sentence prediction. The model architecture is composed of shared layers and task-specific output layers.

In the shared layer, a BERT–BiLSTM structure is employed to generate deep contextualized representations. BERT provides rich pre-trained embeddings, while BiLSTM enhances temporal dependency modeling. On top of this foundation, CNN is utilized to extract fine-grained sentiment features, and a Multi-Head Attention (MHA) module captures inter-task dependencies, highlighting salient information relevant to each subtask.

For the task-specific output layers, a fully connected layer is applied to produce sentiment classification results for the SA task. For legal article recommendation, a softmax-based classifier generates predictions of applicable laws. Conviction and sentence prediction tasks are handled by separate output layers, designed respectively for classifying conviction categories and regressing sentence duration.

During training, task-specific loss functions are combined using weighted summation, and joint optimization is performed to improve the overall accuracy across all subtasks. This approach ensures coordinated learning, effectively balancing the optimization between SA and legal judgment tasks. By leveraging the MTL framework, the model can perform multiple tasks within a unified network architecture, facilitating the sharing of semantic and emotional information across tasks and significantly enhancing performance in each task.

4.2 Joint training and optimization

4.2.1 Loss function design

The loss function of legal text SA is shown in the formula (26).

$L_s=-\sum_{i=1}^{N} y_s^{(i)} log(\hat{y}_s^{(i)})$(26)

$y_s^{(i)}$ represents the true value of the sentiment label, and $\hat{y}_s^{(i)}$ represents the sentiment probability value predicted by the model.

In the recommendation of relevant laws, convictions, and sentence predictions, each subtask uses a cross-entropy loss function, and the loss function design is the same as that of legal text SA.

In MTL, the loss functions of each task differ in numerical scale. In order to avoid a particular task dominating the optimization process of the model, the loss functions of each task are now weighted and merged. The comprehensive weighted loss function is shown in formula (27).

$L_{total}=\iota_1 L_s+\iota_2 L_{law}+\iota_3 L_{con}+\iota_4 L_{sen}$(27)

Among them, ι_1 to ι_4 represent the weight hyperparameters of the task loss function.

4.2.2 Model training

During the training process, the experimental data is fed into the multi-task learning (MTL) framework. The model first generates deep semantic representations through the shared BERT–BiLSTM layer, capturing contextual and sequential information from legal texts. These representations are then processed by the CNN module, which extracts fine-grained sentiment features, and the Multi-Head Attention (MHA) module, which captures long-range dependencies and decision-related signals. Finally, task-specific output layers generate

predictions for sentiment classification, legal article recommendation, conviction, and sentencing.

In each training iteration, the model computes the individual loss values for all subtasks, combines them using a weighted summation strategy, and performs backpropagation based on the aggregated loss. Through continuous gradient descent and parameter updates, the model achieves joint optimization across all tasks, aiming to simultaneously improve prediction accuracy for each task.

This paper adopts a joint training strategy, enabling the model to achieve mutual enhancement between the two core tasks: sentiment analysis and legal judgment prediction. This approach effectively avoids task conflicts and promotes overall performance gains. Notably, in the domain of legal judgment prediction, the model demonstrates a stronger ability to recognize the influence of emotional factors on judicial outcomes.

To prevent overfitting, the training process incorporates an early stopping mechanism: if the F1-score on the validation set fails to improve over five consecutive epochs, training is halted. Additionally, to address potential loss conflicts between tasks, the model dynamically monitors the gradient trends of each task's loss during training. If the gradient variation of a particular task exceeds a defined threshold, its corresponding loss weight (λ) is adaptively adjusted. This dynamic loss balancing strategy helps maintain optimization stability enhances convergence efficiency and ensures that no single task dominates the learning process.

The training hyperparameters are shown in Table 2.

Table 2: Hyperparameters

Parameters	Value	Parameters	Value
Learning rate	0.0001	Number of CNN convolution kernels	128
Constant	10^{-8}	Number of attention heads	8
Batch size	32	Number of LSTM units	512
SA weight	0.4	Conviction prediction weight	0.2
Law recommendation weight	0.2	Sentence prediction weight	0.2

4.2.3 Model optimization

In the joint training, the experiment uses the Adam optimizer to update parameters [35-36]. The updated formula of the Adam optimizer is shown in formula (28).

$$\lambda_{t+1}=\lambda_t-\frac{\mu\tilde{\zeta}_t}{\hat{v}_t+o} \quad (28)$$

μ represents the learning rate, and o represents a very small constant.

5 Experiment on SA of Legal Texts and Prediction of Legal Judgments

5.1 Experimental data

The experimental data in this paper come from the China Artificial Intelligence and Legal Challenge CAIL2018 public dataset, including CAIL-small and CAIL-big. In the dataset, each case covers fact description, relevant laws, and sentences, etc. The experiment uses a ten-fold cross-validation to divide the cases, and the mean is then taken as the final experimental result. The summary information of the dataset is shown in Table 3.

Table 3: Summary information of the dataset

Serial number	Project	CAIL-small	CAIL-big
1	Case	128368	1773099
2	Related laws and regulations	103	118
3	Criminal category	119	130
4	Sentence category	12	12

In Table 3, CAIL-small includes 128,368 cases, 103 related laws, 119 crime categories, and 12 sentence categories. CAIL-big includes 1,773,099 cases, 118 related laws, 130 crime categories, and 12 sentence categories.

5.2 Data preprocessing

This article conducts sentiment annotation on legal texts, relevant legal provisions, crime category annotation, and sentence category annotation. The sentiment annotation categories include positive sentiment, neutral sentiment, and negative sentiment.

This article classifies sentences into months. The minimum sentence for a single crime is 6 months, and the maximum is 15 years. The combined sentence for multiple crimes does not exceed 25 years. Life imprisonment is classified separately.

The experimentutilizes THULAC (THU Lexical Analyzer for Chinese) to segment legal texts, remove punctuation marks and stop words, and discard the text after it is deemed "sufficient to identify".

5.3 Experimental process

In this study, sentiment analysis (SA) is conducted using the BERT–BiLSTM–CNN model, while legal judgment prediction is performed using the BERT–BiLSTM–MHA model. The experimental evaluation consists of three parts: (1) the SA validation experiment, (2) the legal judgment prediction validation experiment, and (3) the ablation study.

For sentencing prediction, the original sentencing texts are first normalized by converting all durations into months. The minimum sentence length for a single offense is set to 6 months, and the maximum is capped at 15 years (i.e., 180 months). For cases involving multiple offenses, cumulative sentencing is limited to 25 years (i.e., 300 months). Cases involving life imprisonment are encoded as a separate category labeled "lifetime." Based on this standardization, each case is mapped into a predefined set of discrete intervals (e.g., 6–8 months, 8–12 months, 12–18 months, ..., 180–300 months, and life imprisonment) to align with the model's multi-class classification output requirements for judicial sentencing prediction.

In the SA experiment, the proposed model is compared with several baseline methods, including Word2Vec, GloVe, LSTM, BERT–GRU (a combination of BERT and Gated Recurrent Unit), and CNN–LSTM. For the legal judgment prediction task, comparison models include HARNN–MTL (Hierarchical Attention-based Recurrent Neural Network with Multi-Task Learning), CNN–MTL, LADAN (Law Article Distillation-based Attention Network), and HAN–MTL (Heterogeneous Graph Attention Network with Multi-Task Learning). These baselines are selected to represent a variety of shallow, sequential, and attention-based neural architectures, enabling a comprehensive evaluation of the proposed framework's effectiveness across tasks.

6 Results of SA of legal texts

6.1 Performance of sentiment classification of legal texts

The performance results of sentiment classification of legal texts are shown in Figure 3.

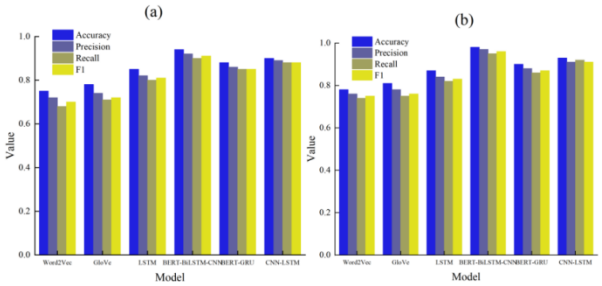

Figure 3 (a): Legal text sentiment classification performance under CAIL-small dataset

Figure 3 (b): Legal text sentiment classification performance under CAIL-big dataset

Figure 3: Legal text sentiment classification performance

In Figure 3(a), the accuracy of traditional models Word2Vec and GloVe is 0.75 and 0.78, respectively, with corresponding F1-scores of 0.70 and 0.72. These models rely primarily on static word embeddings, which are unable to capture dynamic contextual semantics, resulting in limited capability for sentiment classification in legal texts. Deep learning-based models, such as LSTM and CNN–LSTM, demonstrate improved performance. The LSTM model achieves an accuracy of 0.85 and an F1-score of 0.81, while the CNN–LSTM model further improves to 0.90 accuracy and 0.88 F1-score. This indicates that deep neural architectures are more effective in extracting semantic representations, particularly on the CAIL-small dataset. Among all models evaluated, the BERT–BiLSTM–CNN model achieves the best performance, with an accuracy of 0.94 and an F1-score of 0.91, significantly outperforming the baselines. Leveraging pre-trained contextual embeddings from BERT and a multi-layer feature extraction pipeline, the model effectively captures both global semantic and fine-grained sentiment features. Despite the limited size of the dataset, it exhibits strong generalization ability.

In Figure 3(b), similar trends are observed on the CAIL-big dataset. The Word2Vec and GloVe models achieve accuracies of 0.78 and 0.81, with F1-scores of 0.75 and 0.76, respectively. The BERT–BiLSTM–CNN model again outperforms all baselines, reaching an accuracy of 0.98 and an F1 score of 0.96. This superior performance can be attributed to the large-scale

pre-training of the BERT language model, which enables a more accurate representation of complex and domain-specific semantics in legal texts.

Additionally, integrating the model into a multi-task learning framework further enhances its classification capacity by enabling shared semantic and sentiment learning. Compared with results on small datasets, the use of large-scale data significantly reduces the risk of overfitting. It enhances the model's generalization capability, providing more robust and reliable support for sentiment classification in practical legal applications.

6.2 Ablation experiment

The ablation experiment results are shown in Table 4 under the CAIL-big dataset.

Table 4: Ablation experiment results

Model	Accuracy	Precision	Recall	F1
BERT	0.85	0.83	0.8	0.81
BiLSTM	0.89	0.86	0.84	0.85
CNN	0.88	0.87	0.83	0.85
BERT-BiLSTM	0.92	0.89	0.87	0.88
BERT-CNN	0.93	0.91	0.88	0.89
BERT-BiLSTM-CNN	0.98	0.97	0.95	0.96

In Table 4, the BERT-BiLSTM-CNN model achieves the best performance, with an accuracy of 0.98 and an F1 value of 0.96. After gradually removing the modules, the performance gradually decreases. After removing the BiLSTM module, the accuracy of BERT-CNN is 0.93, and the F1 value is 0.89. After removing the CNN module, the accuracy of BERT-BiLSTM is 0.92, and the F1 value is 0.88, indicating that CNN is more important than BiLSTM. When using only BERT, the accuracy drops to 0.85, and the F1 value is 0.81. When using only BiLSTM, the accuracy reaches 0.89, and the F1 value is 0.85.

In BERT-BiLSTM-CNN, BERT provides contextual semantic representation, BiLSTM captures long-distance dependencies in sentences, and CNN plays a key role in extracting and combining local features. This paper uses these modules in combination to achieve optimal performance in semantic understanding and SA.

7.Legal Judgment Prediction Results

7.1 Legal Judgment Confusion Matrix

In the CAIL-big dataset, the test set has 18765 cases of 6-8 years, 23120 cases of 8-12 years, 27483 cases of 12-18 years, 21982 cases of 18-24 years, 17675 cases of 24-36 years, 15894 cases of 36-48 years, 12764 cases of 48-60 years, 10985 cases of 60-96 years, 9432 cases of 96-120 years, 8543 cases of 120-180 years, 7214 cases of above 180 years, and 3452 cases of life imprisonment. The legal judgment confusion matrix is shown in Figure 4.

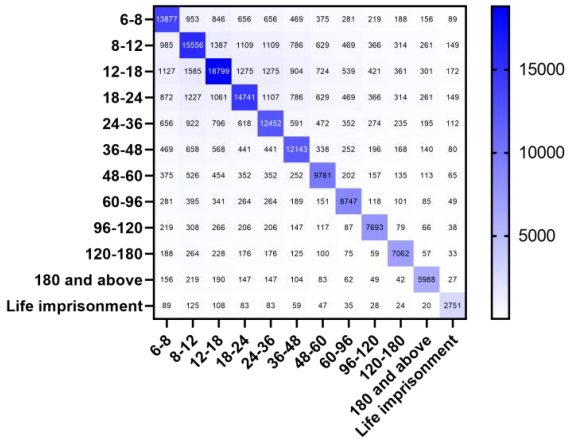

Figure 4: Legal judgment confusion matrix

In Figure 4, 13,877 cases were correctly classified as 6-8 years, 953 were misclassified as 8-12 years, and 846 were misclassified as 12-18 years. 15,556 cases were correctly classified as 8-12 years, and 18,799 cases were correctly classified as 12-18 years. In summary, different categories can be well classified, and there are very few that are incorrectly predicted as other categories.

7.2 Legal judgment prediction performance under CAIL-small dataset

The results of legal judgment prediction performance under the CAIL-small dataset are shown in Figure 5.

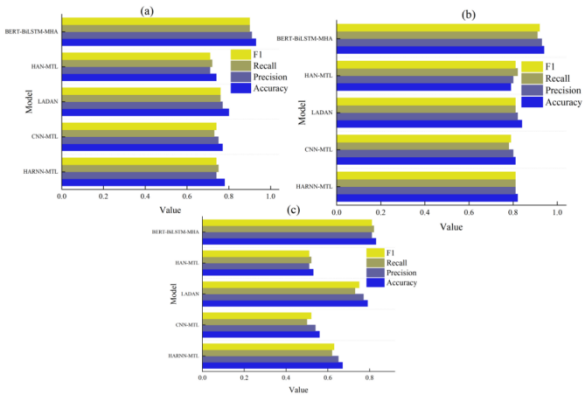

Figure 5 (a): Legal article recommendation prediction performance in CAIL-small dataset
Figure 5 (b): Legal article conviction prediction performance in CAIL-small dataset
Figure 5 (c): Legal article sentence prediction performance in CAIL-small dataset
Figure 5: Legal judgment prediction performance in CAIL-small dataset

In terms of legal article recommendation, the BERT-BiLSTM-MHA model performs best, with an accuracy of 0.93, a precision of 0.91, a recall of 0.90, and an F1 value of 0.90. LADAN has an accuracy of 0.80 and an F1 value of 0.76, while HAN-MTL has an accuracy of only 0.74 and an F1 value of 0.71. In the task of conviction prediction, the BERT-BiLSTM-MHA model exhibits a significant advantage, achieving an accuracy of 0.94, a precision of 0.93, a recall of 0.91, and an F1 score of 0.92. LADAN has an accuracy of 0.84 and an F1 value of 0.81, while HAN-MTL has an accuracy of 0.79 and an F1 value of 0.81.

In the sentence prediction task, BERT-BiLSTM-MHA achieved an accuracy of 0.83, a precision of 0.81, a recall of 0.82, and an F1 value of 0.81, which are much higher than other models. LADAN has an accuracy of 0.79 and an F1 value of 0.75. HARNN-MTL has an accuracy of 0.67 and an F1 value of 0.63. HAN-MTL and CNN-MTL have lower performance, with accuracy of only 0.53 and 0.56, respectively.

In summary, in the CAIL-small dataset, the BERT-BiLSTM-MHA model performs worse in prison sentence prediction in legal judgment prediction but performs better in conviction prediction and legal article recommendation. BERT-BiLSTM-MHA has a profound ability to capture semantic context. It achieves higher prediction accuracy through the context representation provided by BERT, the long-distance dependency captured by BiLSTM, and the important features focused on by MHA.

7.3 Legal judgment prediction performance under CAIL-big dataset

The legal judgment prediction performance results under the CAIL-big dataset are shown in Figure 6.

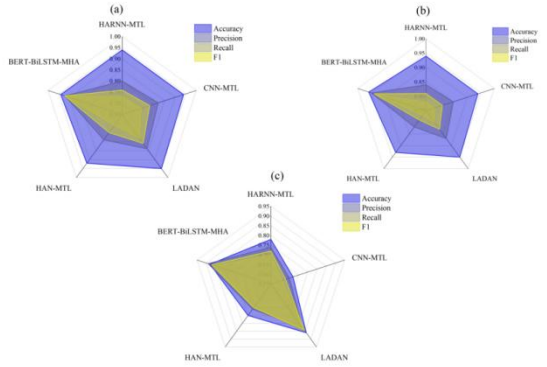

Figure 6 (a): Legal article recommendation prediction performance under CAIL-big dataset
Figure 6 (b): Legal article conviction prediction performance under CAIL-big dataset
Figure 6 (c): Legal article sentence prediction performance under CAIL-big dataset
Figure 6: Legal judgment prediction performance under CAIL-big dataset

In the legal article recommendation task of the CAIL-big dataset, BERT-BiLSTM-MHA achieved an accuracy of 0.94, a precision of 0.92, a recall of 0.92, and an F1 score of 0.92, leading to its overall performance. The LADAN model has a higher accuracy of 0.95 and an F1 value of 0.81, while the F1 value of HAN-MTL is only 0.75. In the conviction prediction task, BERT-BiLSTM-MHA performed best, with an accuracy of 0.96, a precision of 0.95, a recall of 0.93, and an F1 value of 0.94, which are significantly higher than other models. LADAN has an accuracy of 0.95 and an F1 value of 0.83. HAN-MTL has an accuracy of 0.93 and an F1 value of only 0.79, with the weakest performance.

In the prison sentence prediction task, BERT-BiLSTM-MHA has an accuracy of 0.88, a precision of 0.89, a recall of 0.85, and an F1 value of 0.87, ranking first. LADAN has an accuracy of 0.86 and an F1 value of only 0.84. HARNN-MTL and CNN-MTL perform poorly, with F1 values of only 0.72 and 0.63, respectively.

In summary, BERT-BiLSTM-MHA combines the contextual semantic information provided by BERT and the multi-head attention mechanism to capture the semantic associations between legal provisions more accurately. At the same time, it has a strong generalization ability for large datasets and avoids information omission. In the CAIL-big data set, the overall performance is better than that in the CAIL-small data set.

7.4 Significance test

In order to test the statistical significance of the performance difference between BERT-BiLSTM-CNN-MHA and each comparison model under ten-fold cross-validation, this paper designed the following experimental steps:

(1) For each model, ten-fold cross validation was completed on the CAIL-big dataset. The accuracy (one group for each of the three tasks of legal recommendation, conviction prediction, and sentence prediction) and F1 value of each fold were recorded. A total of 10 data sequences were obtained.

(2) For each task, a paired sample t test was used to compare the difference between BERT BiLSTM CNN MHA and each comparison model in the ten-fold results, with the significance level set to $\alpha=0.05$.

(3) The test results of each comparison group were summarized in the form of p-value and marked with "*" for $p<0.05$ (significant), "**" for $p<0.01$ (highly significant), otherwise no mark.

The t-test p-values of BERT-BiLSTM-CNN-MHA and the comparison models on the three tasks' Accuracy are shown in Table 5.

Table 5: T-test results

Model comparison	Legal recommendation p value	Conviction prediction p value	Sentence prediction p-value
HARNN-MTL	0.021*	0.018*	0.003**
CNN-MTL	0.015*	0.012*	0.0005**
LADAN	0.045*	0.034*	0.012*
HAN-MTL	0.008**	0.009**	0.001**

As shown in Table 5, the improvements of BERT-BiLSTM-CNN-MHA on the three tasks have reached statistically significant levels, especially in the sentence prediction task ($p<0.01$), which further verifies the robustness and superiority of this method.

7.5 Discussion on the accuracy of each category (sentence length)

In order to deeply evaluate the performance of the model on different sentence lengths, this paper statistically analyzes the accuracy and recall of the BERT-BiLSTM-MHA model for each sentence category under the CAIL-big dataset. The sentence categories are divided into 12 intervals: 6–8 months, 8–12 months, 12–18 months, 18–24 months, 24–36 months, 36–48 months, 48–60 months, 60–96 months, 96–120 months, 120–180 months, more than 180 months, and life imprisonment.

The prediction accuracy and recall results of each category are shown in Table 6.

Table 6: Prediction accuracy and recall of each category

Sentence type (unit: month)	Accuracy	Recall
6–8	0.86	0.84
8–12	0.88	0.85
12–18	0.90	0.87
18–24	0.89	0.86
24–36	0.91	0.88
36–48	0.92	0.89
48–60	0.90	0.87
60–96	0.87	0.84
96–120	0.85	0.82
120–180	0.83	0.81
180+	0.81	0.78
Life imprisonment	0.95	0.91

As shown in Table 6, the model performs most stably in medium-term sentences (i.e., 12–48 months), with generally high accuracy and recall. For long-term sentences (i.e., those exceeding 180 months), the model's predictive ability has declined, which is attributed to the relatively small number of samples or the lack of concentration of text features. Life imprisonment has obvious emotional characteristics due to its severe circumstances and significant language description. The model has the strongest ability to distinguish this type,

with an accuracy of 0.95. This analysis further verifies the adaptability and effectiveness of the emotion perception mechanism for different types of sentences in sentencing prediction.

7.6 Verification of emotional factors

To verify the importance of emotional factors in legal judgments, this paper conducts experiments using the CAIL-big dataset. The verification results of emotional factors are shown in Table 7.

Table 7: Verification results of emotional factors

	Type	Accuracy	Precision	Recall	F1
Legal recommendation	Emotion	0.94	0.92	0.92	0.92
	Emotion not considered	0.92	0.91	0.89	0.9
	Type	Accuracy	Precision	Recall	F1
Conviction prediction	Emotion	0.96	0.95	0.93	0.94
	Emotion not considered	0.93	0.92	0.92	0.92
	Type	Accuracy	Precision	Recall	F1
Sentence prediction	Emotion	0.88	0.89	0.85	0.87
	Emotion not considered	0.85	0.84	0.83	0.83

In Table 7, the BERT-BiLSTM-MHA model performs better than the model without considering emotional factors in terms of legal article recommendation, conviction prediction, and sentence prediction, which is more in line with actual needs and comprehensively balances morality and legal articles.

8 Further discussion

The model proposed in this paper demonstrates significantly superior performance compared to traditional methods in both legal text sentiment analysis (SA) and legal judgment prediction. Traditional models such as Word2Vec and GloVe are relatively simple in their representation capabilities. They fail to capture contextual dependencies, resulting in low accuracy fully and F1 scores in sentiment classification tasks. In contrast, the BERT-BiLSTM-CNN architecture exhibits excellent performance on both the CAIL-small and CAIL-big datasets. The BERT model provides rich semantic representations through pre-training. When combined with BiLSTM and CNN, it effectively captures long-range dependencies and localized features, substantially improving the accuracy of sentiment analysis.

The impact and significance of this study are primarily reflected in two aspects. First, the introduction of the BERT-BiLSTM-CNN model enhances the effectiveness of legal sentiment classification and legal judgment prediction. Its application on large-scale datasets highlights the model's capability to handle complex semantic structures in legal texts. Second, the research offers a novel perspective for automating legal text processing and provides valuable references for sentiment analysis tasks in other domains. Furthermore, the study employs a multi-task learning (MTL) framework to enhance the accuracy of legal judgment prediction—particularly in legal article recommendation and conviction prediction—yielding outstanding model performance. Overall, this work offers practical value in integrating legal technology with artificial intelligence, contributing to the improvement of efficiency and precision in legal information processing.

Empirically, the proposed BERT-BiLSTM-CNN-MHA model achieves an accuracy of 98.0% and an F1-score of 0.96 on the sentiment classification task, outperforming the 78.99% accuracy of Royyan et al. using Word2Vec [5] and the Rouge-1 F1 score of 0.436 obtained by Anand et al. with a 1DCNN-LSTM model [6]. Compared to GloVe-HeBiLSTM (93.70%, [7]) and pure BERT-CNN (93%, [8]), the proposed model achieves at least a 4.3% improvement, thanks to its multi-level feature extraction and deep contextual embedding design. In judicial judgment prediction, the BERT-BiLSTM-MHA variant yields a conviction prediction accuracy of 96% and an F1-score of 0.94 on the CAIL-big dataset, surpassing the 92.05% accuracy of CNN-LSTM reported by Alghazzawi et al. [21] and the 0.75 F1-score reported by Sengupta et al. [20], with greater robustness across diverse case types.

The substantial improvements in model performance can be attributed to the synergistic effect of deep contextual representation and joint multi-task optimization. BERT's bidirectional encoding effectively captures long-range dependencies and legal-specific semantic nuances, compensating for the limitations of static embeddings like Word2Vec and GloVe, which are unable to resolve semantic ambiguity. BiLSTM further enhances temporal feature modeling, tightly coupling

emotional dynamics with contextual semantics. CNN, through its multi-channel convolutional structure, extracts fine-grained sentiment signals, which are then weighted and fused with contextual representations via the multi-head attention (MHA) mechanism. This design maximizes the influence of emotional cues on legal outcome prediction and compensates for the limitations of traditional single-task models in capturing therelationship between emotion and judgment.

From an architectural perspective, three design choices are pivotal: shared BERT-BiLSTM encoders, multi-channel CNN modules, and multi-head attention (MHA). The shared encoders reduce parameter redundancy while promoting knowledge transfer across tasks. The CNN component captures sentiment features at varying granularities, offering MHA a rich and diverse input representation. MHA then selectively integrates and balances feature weights from different tasks via parallel attention heads, enabling coordinated optimization across sub-tasks such as legal article recommendation, conviction prediction, and sentencing length estimation. The effectiveness of this architecture is further validated by ablation studies: the removal of any single module results in significant performance degradation, underscoring the critical role of each component.

Future research may focus on enhancing the robustness and generalization of the proposed framework, as well as exploring its integration into real-world intelligent legal judgment systems. Additionally, studies in other domains further validate the broad applicability of multi-task learning. For instance, Informatica has demonstrated the use of MTL beyond natural language processing by combining it with Q-learning for joint optimization in logistics tasks such as path selection, scheduling, and resource allocation [40]. Similarly, multi-source deep MTL frameworks have been developed for simultaneous smile detection, emotion recognition, and gender classification [41], further underscoring the versatility and practical potential of the MTL paradigm.

9 Conclusions

This paper adopts an MTL framework and has achieved remarkable results in legal text SA and legal judgment prediction. The experiment constructs a BERT-BiLSTM shared model, utilizes context-related representations to capture the characteristics of legal text, combines a CNN for sentiment feature extraction, and employs MHA to handle tasks related to legal judgment. Experimental results show that the MTL framework outperforms the traditional Word2Vec model in terms of accuracy and F1 value and can effectively improve the performance of legal text analysis. The framework also performs well in tasks such as law recommendation, conviction prediction, and sentence prediction, demonstrating its potential for application in the legal field. This study still has room for improvement in dealing with more complex legal scenarios and multimodal data. Future research can further enhance the robustness of the model, explore additional types of legal cases, and improve the model's ability to comprehend unstructured court statement texts, thereby promoting the in-depth development of intelligent legal judgment systems.

References

[1] Krasadakis, P., Sakkopoulos, E., & Verykios, V. S. (2024). A survey on challenges and advances in natural language processing with a focus on legal informatics and low-resource languages. Electronics, 13(3), 648. https: //doi.org/10.3390/electronics13030648

[2] Maree, M., Al-Qasem, R., & Tantour, B. (2024). Transforming legal text interactions: leveraging natural language processing and large language models for legal support in Palestinian cooperatives. International Journal of Information Technology, 16(1), 551-558. https://doi.org/10.1007/s41870-023-01584-1

[3] DeMattee, A. J. (2023). A grammar of institutions for complex legal topics: Resolving statutory multiplicity and scaling up to jurisdiction-level legal institutions. Policy Studies Journal, 51(3), 529-550. https://doi.org/10.1111/psj.12488

[4] Allison, N. G. (2023). From semantic weight to legal ontology via classification of concepts in legal texts. The Law Teacher, 57(2), 201-217. https://doi.org/10.1080/03069400.2023.2173918

[5] Royyan, A. R., & Setiawan, E. B. (2022). Feature expansion Word2Vec for sentiment analysis of public policy in Twitter. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 6(1), 78-84.

[6] Anand, D., & Wagh, R. (2022). Effective deep learning approaches for summarization of legal texts. Journal of King Saud University-Computer and Information Sciences, 34(5), 2141-2150. https://doi.org/10.1016/j.jksuci.2019.11.015

[7] Mahto, D., & Yadav, S. C. (2024). Emotion prediction for textual data using GloVe based HeBi-CuDNNLSTM model. Multimedia Tools and Applications, 83(7), 18943-18968. https://doi.org/10.1007/s11042-023-16062-w

[8] Bello, A., Ng, S. C., & Leung, M. F. (2023). A BERT framework to sentiment analysis of tweets. Sensors, 23(1), 506. https://doi.org/10.3390/s23010506

[9] Khan, L., Amjad, A., Ashraf, N., & Chang, H. T. (2022). Multi-class sentiment analysis of urdu text using multilingual BERT. Scientific Reports, 12(1), 5436. https://doi.org/10.1038/s41598-022-09381-9

[10] Abimbola, B., de La Cal Marin, E., & Tan, Q. (2024). Enhancing Legal Sentiment Analysis: A Convolutional Neural Network–Long Short-Term

Memory Document-Level Model. Machine Learning and Knowledge Extraction, 6(2), 877-897. https://doi.org/10.3390/make6020041

[11] Rajapaksha, I., Mudalige, C. R., Karunarathna, D., de Silva, N., Ratnayaka, G., & Perera, A. S. (2022). SigmaLaw PBSA-A Deep Learning Approach For Aspect Based Sentiment Analysis in Legal Opinion Texts. J. Data Intell., 3(1), 101-115.

[12] Abimbola, B., Tan, Q., & De La Cal Marín, E. A. (2024). Sentiment analysis of Canadian maritime case law: a sentiment case law and deep learning approach. International Journal of Information Technology, 16(6), 3401-3409. https://doi.org/10.1007/s41870-024-01820-2

[13] Zhang, N., Xiong, J., Zhao, Z., Feng, M., Wang, X., Qiao, Y., & Jiang, C. (2024). Dose my opinion count? A CNN-LSTM approach for sentiment analysis of Indian general elections. Journal of Theory and Practice of Engineering Science, 4(05), 40-50. https://doi.org/10.53469/jtpes.2024.04(05).06

[14] Pavan Kumar, M. R., & Jayagopal, P. (2023). Context-sensitive lexicon for imbalanced text sentiment classification using bidirectional LSTM. Journal of Intelligent Manufacturing, 34(5), 2123-2132. https://doi.org/10.1007/s10845-021-01866-0

[15] Sirisha, U., & Bolem, S. C. (2022). Aspect based sentiment & emotion analysis with ROBERTa, LSTM. International Journal of Advanced Computer Science and Applications, 13(11).

[16] Zulqarnain, M., Ghazali, R., Aamir, M., & Hassim, Y. M. M. (2024). An efficient two-state GRU based on feature attention mechanism for sentiment analysis. Multimedia Tools and Applications, 83(1), 3085-3110. https://doi.org/10.1007/s11042-022-13339-4

[17] Chauhan, A., Sharma, A., & Mohana, R. (2025). An Enhanced Aspect-Based Sentiment Analysis Model Based on RoBERTa For Text Sentiment Analysis. Informatica, 49(14):193-202. https://doi.org/10.31449/inf.v49i14.5423

[18] Yelisetti, S., & Geethanjali, N. (2023). Emotion-based sentiment analysis using conv-BiLSTM with frog leap algorithms. Acta Informatica Pragensia, 12(2), 225-242.

[19] Zhang, H., Dou, Z., Zhu, Y., & Wen, J. R. (2023). Contrastive learning for legal judgment prediction. ACM Transactions on Information Systems, 41(4), 1-25. https://doi.org/10.1145/3580489

[20] Sengupta, S., & Dave, V. (2022). Predicting applicable law sections from judicial case reports using legislative text analysis with machine learning. Journal of Computational Social Science, 5(1), 503-516. https://doi.org/10.1007/s42001-021-00135-7

[21] Alghazzawi, D., Bamasag, O., Albeshri, A., Sana, I., Ullah, H., & Asghar, M. Z. (2022). Efficient prediction of court judgments using an LSTM+ CNN neural network model with an optimal feature set. Mathematics, 10(5), 683. https://doi.org/10.3390/math10050683

[22] Esan, A. O. (2024). PERFORMANCE EVALUATION OF MACHINE LEARNING ALGORITHMS FOR JUDICIAL PREDICTION SYSTEM. LAUTECH Journal of Engineering and Technology, 18(1), 192-203.

[23] Hao, S., Zhang, P., Liu, S., & Wang, Y. (2023). Sentiment recognition and analysis method of official document text based on BERT–SVM model. Neural Computing and Applications, 35(35), 24621-24632. https://doi.org/10.1007/s00521-023-08226-4

[24] Wicaksono, G. W., Azhar, Y., Hidayah, N. P., & Andreawana, A. (2023). Automatic summarization of court decision documents over narcotic cases using bert. JOIV: International Journal on Informatics Visualization, 7(2), 416-422. http://dx.doi.org/10.30630/joiv.7.2.1811

[25] Üveges, I., & Ring, O. (2023). HunEmBERT: a fine-tuned BERT-model for classifying sentiment and emotion in political communication. IEEE Access, 11, 60267-60278. https://doi.org/10.1109/ACCESS.2023.3285536

[26] Jlifi, B., Abidi, C., & Duvallet, C. (2024). Beyond the use of a novel Ensemble based Random Forest-BERT Model (Ens-RF-BERT) for the Sentiment Analysis of the hashtag COVID19 tweets. Social Network Analysis and Mining, 14(1), 88. https://doi.org/10.1007/s13278-024-01240-x

[27] Gou, Z., & Li, Y. (2023). Integrating BERT embeddings and BiLSTM for emotion analysis of dialogue. Computational Intelligence and Neuroscience, 2023(1), 6618452. https://doi.org/10.1155/2023/6618452

[28] Yue, W., & Li, L. (2023). Sentiment analysis using a CNN-BiLSTM deep model based on attention classification. International Information Institute (Tokyo). Information, 26(3), 117-162.

[29] Shelar, A., & Moharir, M. (2024). Judgment prediction from legal documents using Texas wolf optimization based deep BiLSTM model. Intelligent Decision Technologies, 18(2), 1557-1576. https://doi.org/10.3233/IDT-230566

[30] Wankhade, M., Annavarapu, C. S. R., & Abraham, A. (2024). CBMAFM: CNN-BiLSTM multi-attention fusion mechanism for sentiment classification. Multimedia Tools and Applications, 83(17), 51755-51786. https://doi.org/10.1007/s11042-023-17437-9

[31] Başarslan, M. S., & Kayaalp, F. (2023). MBi-GRUMCONV: A novel Multi Bi-GRU and

Multi CNN-Based deep learning model for social media sentiment analysis. Journal of Cloud Computing, 12(1), 5. https://doi.org/10.1186/s13677-022-00386-3

[32] Erkantarci, B., & Bakal, G. (2024). An empirical study of sentiment analysis utilizing machine learning and deep learning algorithms. Journal of Computational Social Science, 7(1), 241-257. https://doi.org/10.1007/s42001-023-00236-5

[33] Jun, W., Tianliang, Z., Jiahui, Z., Tianyi, L., & Chunzhi, W. (2023). Hierarchical multiples self-attention mechanism for multi-modal analysis. Multimedia Systems, 29(6), 3599-3608. https://doi.org/10.1007/s00530-023-01133-7

[34] Zhao, S., Liu, R., Cheng, B., & Zhao, D. (2022). Classification-labeled continuousization and multi-domain spatio-temporal fusion for fine-grained urban crime prediction. IEEE Transactions on Knowledge and Data Engineering, 35(7), 6725-6738. https://doi.org/10.1109/TKDE.2022.3180726

[35] Reyad, M., Sarhan, A. M., & Arafa, M. (2023). A modified Adam algorithm for deep neural network optimization. Neural Computing and Applications, 35(23), 17095-17112. https://doi.org/10.1007/s00521-023-08568-z

[36] Zhang, Y., Chen, C., Shi, N., Sun, R., & Luo, Z. Q. (2022). Adam can converge without any modification on update rules. Advances in neural information processing systems, 35, 28386-28399.

[37] Amit Chauhan, Aman Sharma, Rajni Mohana. "An Enhanced Aspect-Based Sentiment Analysis Model Based on RoBERTa For Text Sentiment Analysis." Informatica, vol. 49, no. 14, 2025. https://doi.org/10.31449/inf.v49i14.5423

[38] Shaha T. Al-Otaibi, Amal A. Al-Rasheed. "A Review and Comparative Analysis of Sentiment Analysis Techniques." Informatica, vol. 46, no. 6, 2022. https://doi.org/10.31449/inf.v46i6.3991

[39] Aliea Sabir, Huda Adil Ali, Maalim A. Aljabery. "ChatGPT Tweets Sentiment Analysis Using Machine Learning and Data Classification." Informatica, vol. 48, no. 7, 2024. https://doi.org/10.31449/inf.v48i7.5535

[40] Xianfeng Zhu. "Multi-Task Deep Reinforcement Learning for Intelligent Logistics Path Planning and Scheduling Optimization." Informatica, vol. 49, no. 20, 2025. https://doi.org/10.31449/inf.v49i20.7996

[41] Sang Dinh Viet, Cuong Le Tran Bao. "Effective Deep Multi-Source Multi-Task Learning Frameworks for Smile Detection, Emotion Recognition and Gender Classification." Informatica, vol. 42, no. 3, 2018. https://doi.org/10.31449/inf.v42i3.2301.