Incorporating Legal Logic into Deep Learning: An Intelligent Approach to Probation Prediction

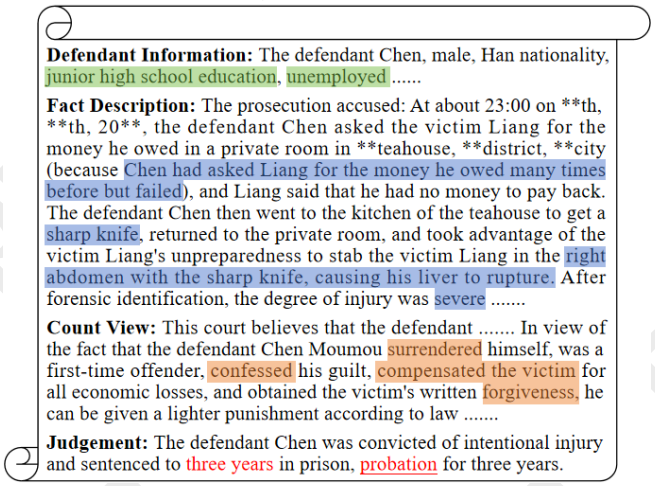
Qinghua Wang[1] , **Xu Zhang**[1] , **Lingyan Yang**[1] ,
Rui Shao[1] , **Bonan Wang**[2] , **Fang Wang**[1] , **Cunquan Qu**[1]
[1]Shandong University
[2]University of Macau
{qinghuawang, xuzhang, 202321227, 202420955}@mail.sdu.edu.cn, bonan.wang@connect.um.edu.mo,
{wangfang226, cqqu}@sdu.edu.cn

Abstract

Probation is a crucial institution in modern criminal law, embodying the principles of fairness and justice while contributing to the harmonious development of society. Despite its importance, the current Intelligent Judicial Assistant System (IJAS) lacks dedicated methods for probation prediction, and research on the underlying factors influencing probation eligibility remains limited. In addition, probation eligibility requires a comprehensive analysis of both criminal circumstances and remorse. Much of the existing research in IJAS relies primarily on data-driven methodologies, which often overlooks the legal logic underpinning judicial decision-making. To address this gap, we propose a novel approach that integrates legal logic into deep learning models for probation prediction, implemented in three distinct stages. First, we construct a specialized probation dataset that includes fact descriptions and probation legal elements (PLEs). Second, we design a distinct probation prediction model named the Multi-Task Dual-Theory Probation Prediction Model (MT-DT), which is grounded in the legal logic of probation and the *Dual-Track Theory of Punishment*. Finally, our experiments on the probation dataset demonstrate that the MT-DT model outperforms baseline models, and an analysis of the underlying legal logic further validates the effectiveness of the proposed approach.

1 Introduction

In recent years, the application of artificial intelligence (AI) in the Intelligent Judicial Assistant System (IJAS) has increased steadily [Zhong *et al.*, 2020; Medvedeva and Mcbride, 2023; Dong and Niu, 2021; Niklaus *et al.*, 2021], primarily with the aim of improving judicial efficiency. Research has mainly focused on improving the performance of AI from two perspectives. The first focuses on continuously refining the extraction and learning of key information from judgment documents [Miao *et al.*, 2024; Bhattacharya *et al.*, 2022; Xiao *et al.*, 2021]. These methods, however, mainly concentrate on superficial information and fail to deeply analyze the legal interpretation of texts. The second adopts data-driven techniques to build effective neural network models for legal judgment prediction (LJP) [Liu *et al.*, 2023; Zhang *et al.*, 2023; Zhang *et al.*, 2024a]. However, these approaches often overlook the inherent legal logic.


Defendant Information: The defendant Chen, male, Han nationality, junior high school education, unemployed .....

Fact Description: The prosecution accused: At about 23:00 on **th, **th, 20**, the defendant Chen asked the victim Liang for the money he owed in a private room in **teahouse, **district, **city (because Chen had asked Liang for the money he owed many times before but failed), and Liang said that he had no money to pay back. The defendant Chen then went to the kitchen of the teahouse to get a sharp knife, returned to the private room, and took advantage of the victim Liang's unpreparedness to stab the victim Liang in the right abdomen with the sharp knife, causing his liver to rupture. After forensic identification, the degree of injury was severe .....

Count View: This court believes that the defendant ..... In view of the fact that the defendant Chen Moumou surrendered himself, was a first-time offender, confessed his guilt, compensated the victim for all economic losses, and obtained the victim's written forgiveness, he can be given a lighter punishment according to law .....

Judgement: The defendant Chen was convicted of intentional injury and sentenced to three years in prison, probation for three years.

Figure 1: Illustrative example of information related to probation conditions. Highlight: information related to probation conditions. Red text: the prerequisite of probation. Red underlined text: probation eligibility.

Probation is a crucial component of sentencing [Matczak, 2021; Berryessa, 2022; Ruhland and Scheibler, 2022]. On the one hand, it offers criminals an opportunity for rehabilitation through non-custodial measures, enabling them to reform and reducing the likelihood of re-offending [Applegate *et al.*, 2009; Phelps *et al.*, 2022; Norman and Ricciardelli, 2022]. On the other hand, probation helps alleviate the burden on the prison system and contributes to social harmony by facilitating the reintegration of criminal [McNeill and Dawson, 2014; Harding *et al.*, 2022; Okonofua *et al.*, 2021]. Despite its importance, current IJAS lack dedicated methods for probation prediction, and research on the underlying factors influencing probation decisions remains limited.

Article 72 of the *Criminal Law of the People's Republic of*

China[1] (*Art. 72*) stipulates that criminals sentenced to detention or fixed-term imprisonment of three years or less may be granted probation if they meet the following conditions at the same time:

- **Condition (a).** Mild circumstances of the crime.
- **Condition (b).** Evidence of remorse.
- **Condition (c).** No danger of re-offending.
- **Condition (d).** No significant negative impact on the community.

In addition, criminals sentenced to detention or fixed-term imprisonment of 3 years or less must be granted probation if they are under 18 years old, pregnant, or over 75 years old. In the Chinese judicial system, both the prerequisite and substantive conditions for probation form a crucial theoretical foundation for determining probation eligibility, which guide judges to assess a criminal's probation eligibility. Figure 1 illustrates a case example of intentional injury, highlighting key information relevant to probation eligibility of criminal. In this case, the defendant Chen was sentenced to a three-year prison term, meeting the prerequisites for probation. In addition, his evidence of remorse and a low risk of recidivism further support his probation eligibility.

Due to limited AI adoption in probation research, we aim to enhance deep learning models with legal logic for more accurate predictions. However, this poses several challenges:

Diverse Legal Information in LJP Tasks. Different prediction tasks focus on distinct types of legal information. For example, in charge prediction tasks [Zhao *et al.*, 2022; Liu *et al.*, 2021; Zhang *et al.*, 2024b], it is essential to incorporate not only the fact description but also the specific elements that constitute each charge, enabling a deeper understanding of its legal implications. In case matching tasks [Bi *et al.*, 2022; Ge *et al.*, 2021], the relevant legal articles and the semantic relationships among entities must be integrated for similarity matching. Similarly, in probation prediction tasks, the PLEs play a particularly crucial role. Therefore, combining fact description with PLEs is critical for improving the performance of probation prediction models.

Legal Logic in Multi-Task Learning. Multi-task learning approaches [Yao *et al.*, 2020] have been employed to predict multiple labels simultaneously. However, these methods typically share the same input across all tasks, neglecting the logical dependencies that exist between tasks. However, in probation prediction task, different information must be provided for each subtask, based on the specific legal logic underlying probation decisions. This task-specific approach aligns with legal logic and has the potential to significantly enhance both the accuracy and interpretability of the predictions.

To address the first challenge, we construct a dataset tailored for probation tasks, focusing on the crime of intentional injury as a case study. This crime exemplifies the conflict between its inherently high subjective malice and the "minor circumstances" required for probation eligibility. To navigate this conflict in probation cases involving intentional injury, we begin by analyzing the interpretation of "minor circumstances" within the substantive conditions of probation. Currently, two mainstream viewpoints exist. First, the crime must involve relatively minor circumstances, indicating a low degree of personal danger posed by the criminal, thereby fulfilling the retribution purpose of probation, referred to as the "retributive factor". Second, the minor circumstances of the crime also suggest a low risk of re-offending, which satisfy the preventive purpose of probation, referred to as the "preventive factor". In this study, we extract the fact description to represent the retributive factors. Additionally, guided by legal expertise, we obtain PLEs reflecting preventive factors and other substantive conditions.

To address the second challenge, we first propose the **T**wo-**S**tage Probation Prediction Model Based on P**LE**s (TS-LE), which focuses on preventive factors using a cascading prediction framework. Building upon TS-LE, we introduce the **T**wo-**S**tage **D**ual-**T**heory Probation Prediction Model (TS-DT), which integrates both retributive and preventive factors, aligning with the *Dual-Track Theory of Punishment* to achieve a more balanced decision-making approach. Finally, extending TS-DT, the **M**ulti-**T**ask **D**ual-**T**heory Probation Prediction Model (MT-DT) adopts a multi-task framework, treating the prediction of prerequisites as an auxiliary task. This enhancement allows for the simultaneous optimization of prerequisites and substantive conditions, further improving prediction accuracy and interpretability. These models incorporate legal interpretation encoding of PLEs to better simulate the judicial decision-making process. By combining fact description with PLEs, these models provide a comprehensive and systematic assessment of probation eligibility. Validation on our constructed dataset demonstrates their effectiveness, with each successive model offering improved performance and deeper insights into the legal logic underlying probation decisions.

To sum up, this paper has three main contributions, summarized as follows:

- **Dataset Construction.** We construct a probation prediction dataset, which includes fact description and PLEs with legal interpretations.
- **Legal Logic Analysis and Model Development.** We conduct an in-depth analysis of the legal logic behind probation eligibility and proposed three progressively advanced models, TS-LE, TS-DT and MT-DT, to simulate the decision-making process for probation.
- **Model Validation and Insights.** Experimental results on the constructed dataset show that the MT-DT model effectively integrates both the retributive and preventive factors, while also addressing the "error amplification" problem in cascading tasks.

2 Related Work

2.1 Exploration of Factors in Probation

Several studies have explored the factors influencing probation eligibility. Cordier *et al.* [Cordier *et al.*, 2021] analyzed the link between legal measures and recidivism risk, offering insights into probation outcomes. Similarly, Sims *et al.*

[1] http://xingfa.org/

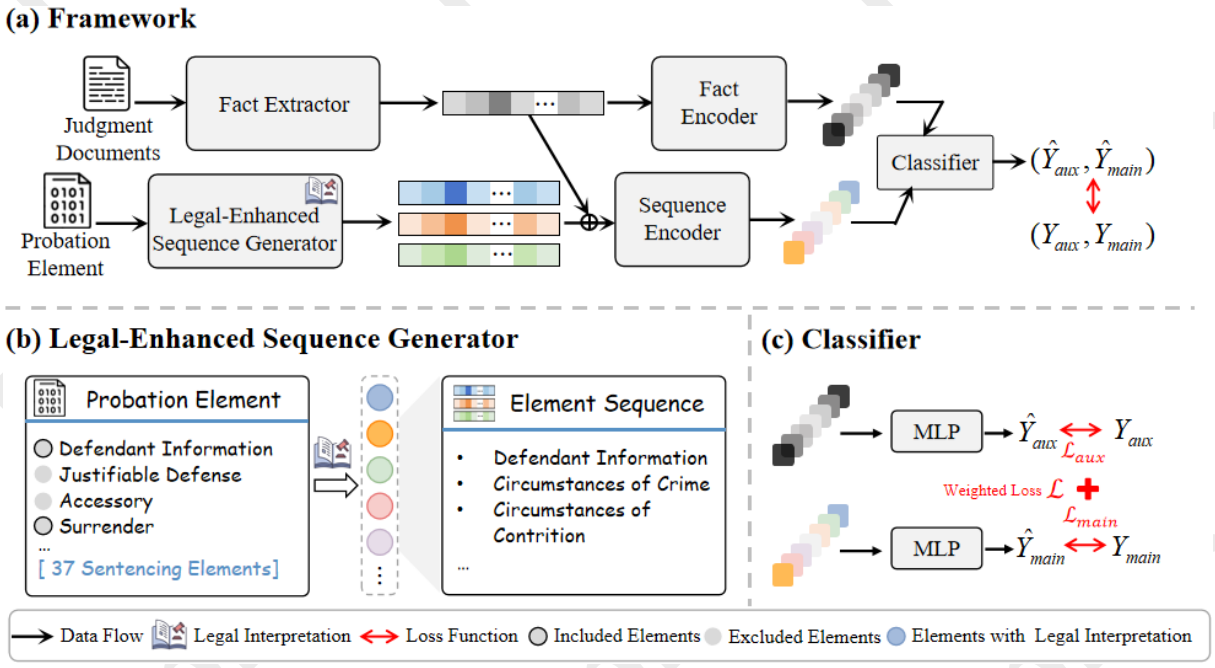


Figure 2: The overall architecture of the proposed MT-DT. It consists of three main steps: (1) Data Processing. This step includes the Fact Extractor and the Legal-Enhanced Sequence Generator. (2) Text Encoder. This step mainly encodes the text to obtain their vector representation. (3) Multi-Task Classifier. This module simulates the probation decision-making process and predicts probation eligibility.

[Sims and Jones, 1997] studied 2,850 felony probationers in North Carolina and identified key factors such as demographics, criminal history, and probation conditions. Furthermore, Louden *et al.* [Louden and Skeem, 2012] used simulation-based experiments to highlight the complex role of mental health issues in probation management. Their findings emphasized the need for customized support and diverse intervention strategies. Williams *et al.* [Williams *et al.*, 2014] developed a tool to assess recidivism risk among juvenile offenders. All of these approaches primarily rely on mathematical and statistical methods.

2.2 Application of Multi-Task Models in LJP

Multi-task learning models have been applied to various LJP tasks. For instance, Lyu *et al.* [Lyu *et al.*, 2022] proposed a method to enhance legal judgment prediction by reinforcing the extraction of criminal elements, integrating these elements into content inputs for multi-task predictors, with the goal of improving model accuracy. Yao *et al.* [Yao *et al.*, 2020] introduced a gated hierarchical multi-task learning network, which models different legal tasks at multiple granularities, thereby improving the predictive performance of judicial decisions. Additionally, Yue *et al.* [Yue *et al.*, 2021] proposed the NeurJudge, an environment-aware approach that divides fact description into different scenarios using intermediate subtask results, which are then leveraged to predict other subtasks. These studies provide important insights for our research on intelligent probation prediction models.

To the best of our knowledge, this is the first work to integrate legal logic with AI techniques for probation prediction. Our approach focuses on developing deep learning models that incorporate legal logic, with the aim of achieving efficient and reliable probation prediction.

3 Method

3.1 Overview

Figure 2 showcases the overall pipeline of our model—data Processing, text Encoder, and multi-task classifier—into a cohesive system.

3.2 Preliminaries

Probation Prediction. According to *Art. 72*, probation applies only to criminals sentenced to three years or less in prison or detention. Therefore, the probation prediction consists of two subtasks: **Task 1:** Predict whether the criminal's sentence meets the prerequisite of three years or less, or detention. **Task 2:** Predict the probation eligibility of criminal.

Retributive (Preventive) Factors. By analyzing the substantive conditions for probation, **Condition (a)** incorporates retributive or preventive factors. Retributive factors refer to elements that embody the principle of retribution within the probation system, primarily focusing on the criminal's personal danger or culpability. In this study, we represent retributive factors using fact description. Preventive factors, on the other hand, reflect the preventive purpose of probation and are included in the PLEs.

The Dual (Single)-Track Theory of Punishment. When **Condition (a)** accounts for the retribution factor, it is known as the *Dual-Track Theory of Punishment*. In contrast, when only the prevention factor is considered, it is known as the *Single-Track Theory of Punishment*. This distinction arises

because the retribution factor has been incorporated into the sentencing process as a prerequisite for probation. Re-incorporating this factor would constitute double retribution.

Probation Legal Elements. Guided by legal expertise, we analyze the substantive conditions for probation eligibility as outlined in *Art. 72*, as follows:

- **Mild Circumstances of the Crime.** For retribution, the fact description includes numerous retributive factors, such as the criminal's modus operandi and the consequences caused. For prevention, the underlying causes are considered, along with other subjective and objective circumstances, such as whether the criminal was forced to commit the crime.
- **Evidence of Remorse.** According to Article 225 of the *Interpretation on the Application of the Criminal Procedure Law of the People's Republic of China*, remorse requires a comprehensive assessment of the criminal's post-crime behavior, including actions such as compensation, expressions of remorse, and whether forgiveness has been obtained.
- **No Danger of Re-offending.** This condition should be analyzed from two perspectives. One is the dangerousness of the criminal's behavior, which focuses on the motivations behind the crime. Elements such as excessive self-defense (i.e., the criminal was coerced into committing the offense) are indicative of a lower risk of recidivism. The other perspective is the dangerousness of the individual, which considers the criminal history, as well as personal factors such as age, occupation, education, family environment, and social circumstances.
- **No Significant Negative Impact on the Community.** This condition consists of a variety of factors, including the criminal's motives, methods, and the outcomes of the crime, to determine the extent of social harm caused by the offense.

Based on the analysis of the substantive conditions, we identify 33 distinct probation legal elements. The distribution of their frequencies is presented in Figure 3.

3.3 Problem Formulation

We assume that the fact description is denoted as F and the text sequence of PLEs as Q. Based on F and Q, the probation prediction task aims to learn a function ϕ, where $\hat{y} = \phi(F, Q)$, to predict the probation eligibility of the criminal. This task consists of two subtasks: $T = \{T_{\text{aux}}, T_{\text{main}}\}$. The corresponding prediction results are $\hat{y}_{\text{aux}} \in \{0, 1\}$ and $\hat{y}_{\text{main}} \in \{0, 1\}$.

3.4 Legal-Enhanced Sequence Generator

As illustrated in Figure 2 (b), to provide legal implication to PLMs, we obtain the interpretation of each element from the encyclopedia entries and generate a corresponding text sequence. More specifically, we first apply regularization to match the elements of each case, as follows:

$$\mathcal{V}_i = \mathcal{H}(R, E), \quad i = 1, 2, \ldots, N, \tag{1}$$

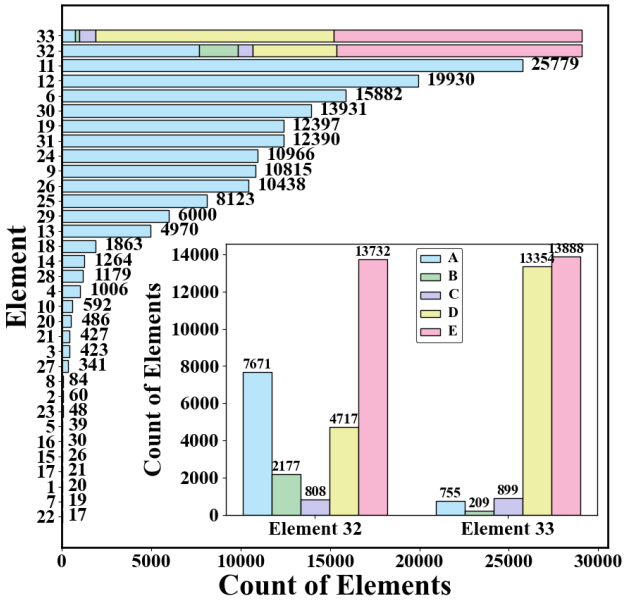


Figure 3: Distribution of frequencies for each element in the dataset, where elements 32 and 33 each contain five distinct values.

where $\mathcal{V}_i = \{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\}$ is a vector that represents the presence of elements in the i-th judgement document, and $v_{i_k} \in \{0, 1\}$ indicates whether the k-th element is present ($v_{i_k} = 1$) or absent ($v_{i_k} = 0$). Here, $\mathcal{H}$ is the extraction function, R is the extraction rule, E represents the set of PLEs, and N is the number of judgement document.

Then, we extract the legal interpretation for each element from the encyclopedic entries, forming $\mathcal{A}$ as:

$$\mathcal{A} = \{A_1, A_2, \ldots, A_k\}, \tag{2}$$

where A_k mains the legal meaning of the k-th element.

Finally, we search the knowledge base for the legal interpretation corresponding to the elements in the case and generate a text sequence of elements unique to the case Q_i as:

$$Q_i = \mathcal{G}(\mathcal{A}, \mathcal{V}_i), \tag{3}$$

where $\mathcal{G}$ represents the generating function.

3.5 Training and Prediction

In the multi-task training of MT-DT, the fact description is encoded as input for the auxiliary task, represented as:

$$w_{\text{aux}} = \text{ENC}_{\text{aux}}(F), \tag{4}$$

where F denotes a sequence representing the fact description, $F = \{f_1, f_2, \ldots, f_m\}$, and m is the length of sequence.

According to the *Dual-Track Theory of Punishment*, we concatenate the fact description and the text sequences of PLEs, encoding them jointly as input for the main task. Specifically, this process can be represented as follows:

$$w_{\text{main}} = \text{ENC}_{\text{main}}(F, Q), \tag{5}$$

where Q denotes a word sequence representing the elements, $Q = \{q_1, q_2, \ldots, q_n\}$, and n is the length of sequence.

	Framework	Input Information	Legal Logic
TS-LE	Cascaded Two-Stage Model	Task1: F Task2: Q	Single-Track Theory of Punishment
TS-DT	Cascaded Two-Stage Model	Task1: F Task2: F & Q	Dual-Track Theory of Punishment
MT-DT	Multi-task Model	F & Q	Dual-Track Theory of Punishment

(a) Models with varying legal logics and framework.

Model		ACC(%)	MP(%)	MR(%)	F1(%)
TS-LE	Task 1	95.17	83.55	86.77	85.06
	Task 2	69.65	68.16	67.12	66.95
TS-DT	Task 1	95.17	83.55	86.77	85.06
	Task 2	72.88	70.28	73.13	72.45
MT-DT	Task 2	88.15	84.01	87.57	85.47

(b) Results on TS-LE, TS-DT, and MT-DT

Table 1: Performance comparison of three legal logic enhanced probation prediction models.

To obtain prediction results for the auxiliary and main task, we use two MLP classifiers:

$$\begin{aligned} \hat{y}_{\text{aux}} &= \text{softmax}(\text{MLP}(w_{\text{aux}})), \\ \hat{y}_{\text{main}} &= \text{softmax}(\text{MLP}(w_{\text{main}})). \end{aligned} \tag{6}$$

In order to train the multi-task probation predictor, we compute the cross-entropy loss function for both the auxiliary and main task. The loss for the auxiliary task is formally computed as:

$$\mathcal{L}_{aux} = -\sum_{c=1}^{N} y_{aux_c} \log(\hat{y}_{aux_c}), \tag{7}$$

where y_{aux_c} represents the ground truth for the auxiliary task prediction associated with the c-th judgment document. Similarly, the loss for the main task is formally computed as:

$$\mathcal{L}_{main} = -\sum_{c=1}^{N} y_{main_c} \log(\hat{y}_{main_c}), \tag{8}$$

where y_{main_c} represents the ground truth for the main task prediction associated with the c-th judgment document.

During multi-task training, we sum the losses of the different subtasks as the total loss to train the multi-task model:

$$\mathcal{L} = \mathcal{L}_{\text{main}} + \lambda \mathcal{L}_{aux}, \tag{9}$$

where λ represents the weight of the auxiliary task loss.

4 Analysis

In this section, we aim to answer the following research questions:

- **(RQ1)** How can the underlying legal logic of the methods be interpreted? (Section 4.1)
- **(RQ2)** Is the text sequence of PLEs valid? (Section 4.2)
- **(RQ3)** What is the impact of the hyperparameter λ in the multi-task model? (Section 4.3)

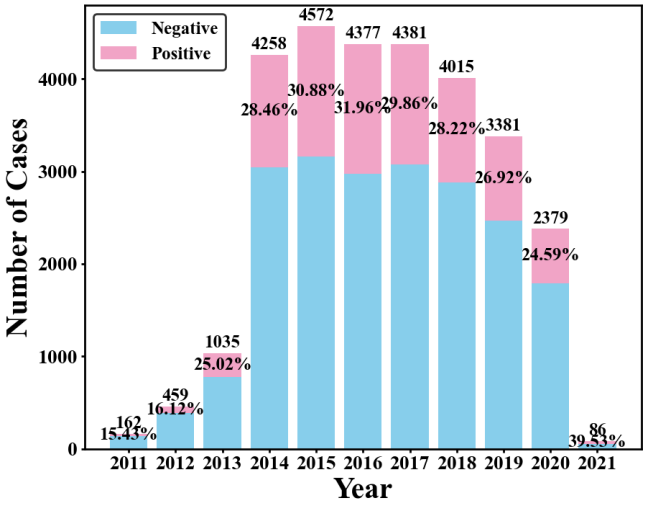


Figure 4: Case distribution by year. From 2011 to 2021, probation accounted for only 28.69% of all cases.

4.1 Legal Logic Analysis

As shown in Table 1 (a), to model the legal logic in the probation decision process, we establish three models with different frameworks and legal logic. In Table 1 (b), the results show that the multi-task model framework based on the *Dual-Track Theory of Punishment* performs the best.

TS-LE. According to the two subtasks of the probation prediction task, in the TS-LE model, we use the fact description as input for Task 1. If the prediction result meets the prerequisites for probation eligibility, we use the text sequence of PLEs as input for Task 2 to predict the probation eligibility of criminals.

As shown in Table 1 (b), the TS-LE method achieves an accuracy of 95.17% in Task 1. The legal logic behind this lies in Article 234 of the *Criminal Law of the People's Republic of China*, which specifies that in cases of intentional injury, sentencing is determined by the means employed by the criminal and the resulting consequences. Therefore, determining whether a criminal meets the prerequisites for probation relies on the fact description, which includes the criminal's actions and the outcomes.

Next, we select cases from the dataset where the prediction result of Task 1 meets the prerequisites, and we input the text sequence of PLEs into the model to predict the probation eligibility of criminals. The results show that the accuracy is only 69.65%, which is less than the accuracy of the MT-DT method (88.15%). This reflects a legal logic where the independent training approach employed by the TS-LE method causes the two models trained in Task 1 and Task 2 to focus respectively on retributive and preventive factors. This approach fails to capture the dual retribution principle in *Dual-Track Theory of Punishment*, which requires revisiting the retributive factors in Task 2.

TS-DT. Based on the analysis in TS-LE, we propose a revised two-stage probation prediction method that aligns more closely with the *Dual-Track Theory of Punishment* in legal

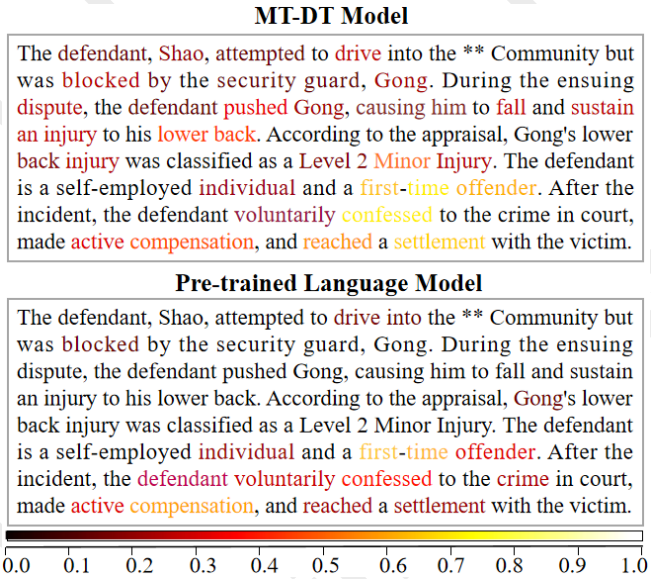


Figure 5: An example of attention score visualization of the MT-DT model and the pre-trained language model. The color changes from black to yellow, indicating that the model pays more and more attention to the text.

logic. Specifically, in Task 2, we incorporate both the fact description and the text sequence of PLEs as joint inputs to the model for prediction. The experimental results in Table 1 (b) demonstrate that this model outperforms the TS-LE method in probation prediction.

Furthermore, as shown in Figure 4, due to class imbalance in dataset, the cascade method used for training in Task 2 may lead to suboptimal model performance. The imbalance significantly hampers the model's ability to effectively predict probation eligibility. Moreover, since the prediction accuracy of Task 1 does not reach 100%, any errors made in this task are likely to be amplified in Task 2. In practice, for cases where the true sentence is three years or less, but the Task 1 prediction incorrectly exceeds three years, these cases will not proceed to Task 2. Consequently, PLEs will not be considered and the probation eligibility cannot be evaluated. This results in disproportionately harsh sentencing outcomes for such criminals.

MT-DT. To address the error amplification issue discussed in TS-DT, we adopt a joint training approach by modifying Task 1 to an implicit intermediate task. Thus, we propose a multi-task probation prediction model based on the *Dual-Track Theory of Punishment*. This model treats Task 1 as an auxiliary task, jointly training Tasks 1 and 2.

In practice, we redefine the loss function of the model to ensure that, during training, the model simultaneously focuses on both the retributive and preventive factors of the criminal. According to the experimental results in Table 1 (b), the MT-DT method improves accuracy by 15.27% compared to the TS-DT.

Additionally, we visualize the attention mechanism of the model, as shown in Figure 5. The visualization demonstrates that the model effectively captures both retributive and preventive factors in MT-DT. This indicates that our joint training framework effectively enhances the focus of the model on retributive factors, leading to better alignment with the legal logic of the *Dual-Track Theory of Punishment*.

Input Information	Ablated Variants		
	A	B	C
Fact description	✔	✔	✔
The vector of PLEs	✗	✔	✗
The sequence of PLEs	✗	✗	✔

(a) Different combinations of input information.

Ablated Variants	ACC(%)	MP(%)	MR(%)	F1(%)
A	82.30	77.35	77.82	77.58
B	83.42	78.90	80.85	79.75
C	88.15	84.01	87.57	85.47

(b) Results with different input information

Table 2: Ablation study on input information

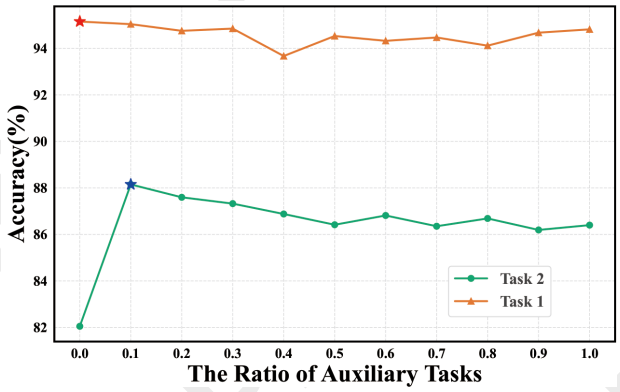


Figure 6: Performance of the MT-DT model across different λ. The parameter is varied from 0 to 1 to assess its influence on the model's overall performance. The stars indicate maximum accuracy

4.2 Ablation Studies

As shown in Table 2 (a), we perform an ablation study on the MT-DT model. To examine the contribution of PLEs to performance, we consider the following three model variants:

- **A.** Fact description only.
- **B.** Fact description and the vector of PLEs.
- **C.** Fact description and the text sequence of PLEs.

As shown in Table 2 (b), replacing the vector with the text sequence of PLEs as input improves the accuracy by 4.73%. Furthermore, the accuracy of both configurations involving PLEs (vector and text) exceeds that of using fact description alone. This suggests that the legal elements provide additional semantic information, thereby enhancing the model's ability to make accurate predictions.

4.3 Influences of the Parameter λ

We further investigate the effect of the auxiliary task loss ratio in the overall loss on the performance of the MT-DT model. The results are shown in Figure 6. From the results, we observe the following:

	Model	ACC(%)	MP(%)	MR(%)	F1(%)
Neural Network Models	TextCNN [Kim, 2014]	81.70	76.72	74.34	75.36
	TextRNN [Liu *et al.*, 2016]	81.56	77.30	72.01	73.87
	Att-BLSTM [Zhou *et al.*, 2016]	81.19	76.00	78.93	77.13
	TextRCNN [Lai *et al.*, 2015]	82.29	77.28	76.31	76.77
	FastText [Joulin *et al.*, 2017]	82.04	76.91	76.28	76.58
Pre-trained Language Models	Lawformer [Xiao *et al.*, 2021]	<u>84.42</u>	80.55	79.88	80.21
	LagelBERT [Chalkidis *et al.*, 2020]	74.95	71.12	56.28	55.02
	ELECTRA [Clark *et al.*, 2020]	83.24	78.60	80.08	79.25
	ALBERT [Lan *et al.*, 2020]	82.40	77.85	76.11	76.89
	BERT [Devlin *et al.*, 2019]	82.05	<u>82.34</u>	<u>82.04</u>	<u>82.01</u>
	MT-DT (ours)	**88.15**	**84.01**	**87.57**	**85.47**

Table 3: Evaluation of our model and baselines. The optimal performances are shown in bold font. The underlined values denote the optimal results of baselines.

- When the auxiliary task loss is excluded from the loss function, the approach is comparable to directly using the pre-trained model for probation prediction. As a result, the model demonstrates suboptimal performance.
- When the auxiliary task loss is incorporated, the performance of the MT-DT model stabilizes in the range of 86% to 88%, significantly surpassing the baseline. This demonstrates the robustness and effectiveness of the proposed method.

5 Experiments

5.1 Dataset

We obtain 29,105 judgment documents from publicly available legal documents on China judgements[2]. Of these, 8,351 cases involve probation, representing 28.69% of the total. The dataset consists of first-instance intentional injury cases with a single defendant and a single charge, spanning from 2011 to 2021. The dataset was randomly split in an 8:1:1 ratio into training sets (23284 cases), validation sets (2911 cases), and test sets (2910 cases).

5.2 Experimental Settings

We set the encoder feature dimension at 768, with the maximum document length limited to 512 words. To prevent overfitting, dropout regularization was applied to the feature vectors. The batch size was set to 16. The models were fine-tuned using a batch contrastive loss function. For optimization, we employ the Adam optimizer [Kingma and Ba, 2014] with a learning rate of 10^{-5} and a dropout rate of 0.3. The model was trained for 10 epochs. The hyperparameters that performed the best, determined in the validation set, were evaluated in the test set, and the process was repeated six times to compute the average prediction results. The hyperparameters were set as follows: $\lambda = 0.1$.

For evaluation, we use the Accuracy (Acc), Macro Precision (MP), Macro Recall (MR), and Macro F1 (F1) metrics.

[2]https://wenshu.court.gov.cn/

5.3 Baseline

In this section, we conduct a comprehensive comparison between our proposed MT-DT model and a range of conventional text classification models, as well as pre-trained language models widely used in natural language processing tasks. The detailed performance metrics for all models are summarized in Table 3. The empirical results clearly demonstrate that MT-DT consistently outperforms all baseline models across various evaluation criteria, establishing its superior predictive capability in the context of the given tasks. Among the baseline models, LawFormer achieves the highest prediction accuracy, which is designed to effectively capture intricate patterns and perform sophisticated information extraction from large-scale, domain-specific legal corpora. However, despite its impressive accuracy, the complexity of the model leads to longer prediction times, which presents a challenge for real-time applications.

6 Conclusion

In this study, we propose a dataset for probation prediction, grounded in legal knowledge, and design an intelligent prediction model based on legal logic. By analyzing the substantive conditions for probation, we extract legal elements from legal documents and generate a text sequence, enriching the model with semantic information beyond the fact descriptions. We also introduce a novel loss function that enables the multi-task model to simultaneously focus on both facts and legal elements, thereby improving accuracy. Extensive experiments demonstrate that our method outperforms others and aligns with the legal logic of probation. Additionally, we assess the sensitivity of the model to parameters.

In future research, we will conduct a deeper investigation into the key factors influencing probation sentence lengths, as well as systematically analyze the underlying legal principles governing such judicial determinations. Furthermore, we plan to extend our dataset to encompass cases involving a broader range of charges and incorporate more sophisticated legal reasoning frameworks to enhance the model's robustness and generalizability.

Ethical Statement

The dataset used in this study has been fully anonymized to protect the privacy of individuals. All data processing and analysis were conducted in compliance with relevant ethical guidelines and legal requirements.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant T2293773 and Grant 12471488. Qinghua Wang and Xu Zhang contributed equally to this work. Please ask Dr. Cunquan Qu (cqqu@sdu.edu.cn) for correspondence.

References

[Applegate *et al.*, 2009] Brandon K. Applegate, Hayden P. Smith, Alicia H. Sitren, and Nicolette Fariello Springer. From the inside: The meaning of probation to probationers. *Criminal Justice Review*, 34:80–95, 2009.

[Berryessa, 2022] Colleen M Berryessa. Modeling "remorse bias" in probation narratives: Examining social cognition and judgments of implicit violence during sentencing. *Journal of Social Issues*, 78(2):452–482, 2022.

[Bhattacharya *et al.*, 2022] Paheli Bhattacharya, Kripabandhu Ghosh, Arindam Pal, and Saptarshi Ghosh. Legal case document similarity: You need both network and text. *Information Processing & Management*, 59:103069, 2022.

[Bi *et al.*, 2022] Sheng Bi, Zafar Ali, Meng Wang, Tianxing Wu, and Guilin Qi. Learning heterogeneous graph embedding for chinese legal document similarity. *Knowledge-Based Systems*, 250:109046, 2022.

[Chalkidis *et al.*, 2020] Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. LEGAL-BERT: The muppets straight out of law school. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2898–2904, Online, 2020. Association for Computational Linguistics.

[Clark *et al.*, 2020] Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. Electra: Pre-training text encoders as discriminators rather than generators. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

[Cordier *et al.*, 2021] Reinie Cordier, Donna Chung, Sarah Wilkes-Gillan, and Renée Speyer. The effectiveness of protection orders in reducing recidivism in domestic violence: A systematic review and meta-analysis. *Trauma, Violence, & Abuse*, 22:804–828, 2021.

[Devlin *et al.*, 2019] Jacob Devlin, MingWei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, 2019. Association for Computational Linguistics.

[Dong and Niu, 2021] Qian Dong and Shuzi Niu. Legal judgment prediction via relational learning. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 983–992, 2021.

[Ge *et al.*, 2021] Jidong Ge, Yunyun Huang, Xiaoyu Shen, Chuanyi Li, and Wei Hu. Learning fine-grained fact-article correspondence in legal cases. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3694–3706, 2021.

[Harding *et al.*, 2022] David J Harding, Bruce Western, and Jasmin A Sandelson. From supervision to opportunity: Reimagining probation and parole, 2022.

[Joulin *et al.*, 2017] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431, Valencia, Spain, 2017. Association for Computational Linguistics.

[Kim, 2014] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, 2014. Association for Computational Linguistics.

[Kingma and Ba, 2014] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[Lai *et al.*, 2015] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. Recurrent convolutional neural networks for text classification. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 2267—-2273, Austin, Texas, 2015. AAAI Press.

[Lan *et al.*, 2020] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, Radu Soricut, Google Research, and Mailton de Carvalho. Albert: A lite bert for self-supervised learning of language representations. 2020.

[Liu *et al.*, 2016] Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. Recurrent neural network for text classification with multi-task learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 2873—-2879, New York, New York, USA, 2016. AAAI Press.

[Liu *et al.*, 2021] Xiao Liu, Da Yin, Yansong Feng, Yuting Wu, and Dongyan Zhao. Everything has a cause: Leveraging causal inference in legal text analysis. *arXiv preprint arXiv:2104.09420*, 2021.

[Liu *et al.*, 2023] Yifei Liu, Yiquan Wu, Yating Zhang, Changlong Sun, Weiming Lu, Fei Wu, and Kun Kuang. Ml-ljp: Multi-law aware legal judgment prediction. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1023—-1034, New York, NY, USA, 2023. Association for Computing Machinery.

[Louden and Skeem, 2012] Jennifer Eno Louden and Jennifer L. Skeem. How do probation officers assess and manage recidivism and violence risk for probationers with mental disorder? an experimental investigation. *Law and human behavior*, 37:22–34, 2012.

[Lyu *et al.*, 2022] Yougang Lyu, Zihan Wang, Zhaochun Ren, Pengjie Ren, Zhumin Chen, Xiaozhong Liu, Yujun Li, Hongsong Li, and Hongye Song. Improving legal judgment prediction through reinforced criminal element extraction. *Information Processing & Management*, 59:102780, 2022.

[Matczak, 2021] Anna Matczak. The penal narratives of community sentence and the role of probation: The case of the wrocław model of community service. *European Journal of Probation*, 13(1):72–88, 2021.

[McNeill and Dawson, 2014] Fergus McNeill and Matt Dawson. Social solidarity, penal evolution and probation1. *The British Journal of Criminology*, 54:892–907, 2014.

[Medvedeva and Mcbride, 2023] Masha Medvedeva and Pauline Mcbride. Legal judgment prediction: If you are going to do it, do it right. In *Proceedings of the Natural Legal Language Processing Workshop 2023*, pages 73–84, Singapore, 2023. Association for Computational Linguistics.

[Miao *et al.*, 2024] Yingzhi Miao, Fang Zhou, Martin Pavlovski, and Weining Qian. Learning legal text representations via disentangling elements. *Expert Systems with Applications*, 249:123749, 2024.

[Niklaus *et al.*, 2021] Joel Niklaus, Ilias Chalkidis, and Matthias Stürmer. Swiss-judgment-prediction: A multilingual legal judgment prediction benchmark. *arXiv preprint arXiv:2110.00806*, 2021.

[Norman and Ricciardelli, 2022] Mark Norman and Rosemary Ricciardelli. Operational and organisational stressors in community correctional work: Insights from probation and parole officers in ontario, canada. *Probation Journal*, 69(1):86–106, 2022.

[Okonofua *et al.*, 2021] Jason A Okonofua, Kimia Saadatian, Joseph Ocampo, Michael Ruiz, and Perfecta Delgado Oxholm. A scalable empathic supervision intervention to mitigate recidivism from probation and parole. *Proceedings of the National Academy of Sciences*, 118(14):e2018036118, 2021.

[Phelps *et al.*, 2022] Michelle S Phelps, Ingie H Osman, Christopher E Robertson, and Rebecca J Shlafer. Beyond "pains" and "gains": untangling the health consequences of probation. *Health & Justice*, 10(1):29, 2022.

[Ruhland and Scheibler, 2022] Ebony Ruhland and Esther Scheibler. Probation officer discretion in monitoring and violating supervision conditions. *Probation Journal*, 69(2):177–196, 2022.

[Sims and Jones, 1997] Barbara Sims and Mark Jones. Predicting success or failure on probation: Factors associated with felony probation outcomes. *Crime & Delinquency*, 43:314–327, 1997.

[Williams *et al.*, 2014] Lela Williams, Craig LeCroy, and John Vivian. Assessing risk of recidivism among juvenile offenders: The development and validation of the recidivism risk instrument. *Journal of evidence-based social work*, 11:318–327, 2014.

[Xiao *et al.*, 2021] Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. Lawformer: A pre-trained language model for chinese legal long documents. *AI Open*, 2:79–84, 2021.

[Yao *et al.*, 2020] Fanglong Yao, Xian Sun, Hongfeng Yu, Yang Yang, Wenkai Zhang, and Kun Fu. Gated hierarchical multi-task learning network for judicial decision prediction. *Neurocomputing*, 411:313–326, 2020.

[Yue *et al.*, 2021] Linan Yue, Qi Liu, Binbin Jin, Han Wu, Kai Zhang, Yanqing An, Mingyue Cheng, Biao Yin, and Dayong Wu. Neurjudge: A circumstance-aware neural framework for legal judgment prediction. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 973—-982, New York, NY, USA, 2021. Association for Computing Machinery.

[Zhang *et al.*, 2023] Han Zhang, Zhicheng Dou, Yutao Zhu, and Ji-Rong Wen. Contrastive learning for legal judgment prediction. *ACM Trans. Inf. Syst.*, 41:1–25, 2023.

[Zhang *et al.*, 2024a] Yunong Zhang, Xiao Wei, and Hang Yu. Hd-ljp: A hierarchical dependency-based legal judgment prediction framework for multi-task learning. *Know.-Based Syst.*, 299:112033, 2024.

[Zhang *et al.*, 2024b] Yunong Zhang, Xiao Wei, and Hang Yu. Hd-ljp: A hierarchical dependency-based legal judgment prediction framework for multi-task learning. *Knowledge-Based Systems*, 299:112033, 2024.

[Zhao *et al.*, 2022] Jie Zhao, Ziyu Guan, Cai Xu, Wei Zhao, and Enze Chen. Charge prediction by constitutive elements matching of crimes. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 4517–4523. International Joint Conferences on Artificial Intelligence Organization, 2022.

[Zhong *et al.*, 2020] Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. How does NLP benefit legal system: A summary of legal artificial intelligence. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5218–5230, Online, July 2020. Association for Computational Linguistics.

[Zhou *et al.*, 2016] Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 207–212, Berlin, Germany, 2016. Association for Computational Linguistics.