A Study of Legal Judgment Prediction Based on Deep Learning Multi-Fusion Models—Data from China

Yu Wen[1] and Ping Ti[2] (iD)

Abstract

Legal judgment prediction implies predicting judgment results based on the case description suitable for judgment predictions in legal terms, crimes, fines, and punishment. This study examines Chinese legal cases, focusing on legal predictions based on deep learning multiple fusion models. It aims to analyze case descriptions and predict legal terms and fines. The use of artificial intelligence models to predict the law can solve the problem of increasing workload of legal institutions and personnel because of the increasing number of cases; and simultaneously, reduce the differences in judgment. This will improve the efficiency and fairness of legal judgments. This study used the BDCI2017 dataset for experiments and applied deep learning algorithms to improve the prediction accuracy. Various models, such as TextCNN, TextRNN, Wide & TextCNN, and TextDenseNet classify cases and fines. Results revealed that TextDenseNet is better than the other model structures in terms of predictive accuracy.

The use of artificial intelligence (AI) models in legal judgment prediction has acquired significant attention in recent years with the aim of improving the efficiency and fairness of the legal system. This is particularly important because of the increasing burden on legal staff caused by continuous changes in global legislation and a growing number of precedents. Using AI models, legal outcomes can be predicted, and legal recommendations can be provided in a timely and fair manner (Medvedeva et al., 2020).

The application of AI models in legal judgment prediction can be traced back to the 1980s, when early research focused on statistical models and neural networks (Susskind, 1991). Subsequently, the use of decision trees and case-based reasoning indicated promising results in predicting court judgments (Bruninghaus & Ashley, 2003; Chaphalkar et al., 2015). Over time, various machine learning models such as support vector machines, logistic regression, and natural language processing techniques have been employed to enhance legal prediction and decision-making (Kelly et al., 2019). The effectiveness of AI models in legal prediction and decision-making has been demonstrated by the development of intelligent question-answering robots, legal big

data analysis platforms, and prediction tools (Gavaghan, 2017; Maharg, 2017; Murimi, 2021; Semmler & Rose, 2017). For example, in the US, the AI platform "Lex Machina" can help lawyers or companies formulate more accurate litigation strategies by collecting and integrating past trial cases and structurally extracting key information, and thereafter analyzing the judges' approval tendencies (Ashley, 2018; Joshua, 2020; Kufakwababa, 2021; Lettieri et al., 2018). Similarly, the French Courts of Appeal in Douai and Rennes have begun the trial of the Prédictice software to assist in judging cases (Alfaia Sampaio et al., 2022; N. Zhang & Xu, 2019).

Furthermore, in China, the adoption of AI models in legal judgment prediction is growing, with collaborations

[1]Guilin University of Tourism, Guilin, China
[2]Guangxi Normal University, Guilin, China

Corresponding Author:
Ping Ti, Guangxi Normal University, No. 1 Yanzhong Road, Yanshan, Guilin, Guangxi 541004 China.
Email: 2312126055@qq.com

Data Availability Statement included at the end of the article

between AI companies and law firms to provide legal services (J. Li, 2019). Notably, the Beijing Internet Court introduced the world's first "AI Virtual Judge" in 2019, enabling convenient online litigation services through an intelligent litigation service center (China News Network, 2019). AI models are used to analyze case data and previous judgments, and generate prediction reports, thereby facilitating the prediction of success rates. The Hangzhou Internet Court is another example of using AI models to predict legal judgments, particularly in civil disputes such as e-commerce cases (Legaldaily, 2022). By using natural language processing and machine learning algorithms, the court's AI system analyzes case data and identifies patterns that assist in decision making. In 2022, the Shanghai Financial Court launched an AI-powered platform for judgment prediction, employing natural language processing and machine learning algorithms to analyze and compare past legal cases with specific details of the current case (Zeng & Pan, 2023).

Therefore, the application of deep learning models to legal judgments and punishment prediction indicates considerable potential for improving the accuracy of legal systems (Ahmad et al., 2022). Deep learning models use artificial neural networks to analyze large amounts of data and extract meaningful patterns, surpassing traditional methods in terms of prediction accuracy (S. Li et al., 2019; Miotto et al., 2018). These neural networks can be trained to recognize patterns in legal documents, similar to human legal analyses, and learn from previous judgments to predict more accurately in future cases (Law, 2000). Additionally, deep learning models can learn from mistakes and adapt their methods accordingly, meaning that they can continuously improve their predictions over time (Zhu, Guo, et al., 2020).

However, most studies on deep learning models only consider the content described by the case and ignore the issues of existing keyword information in legal provisions, the crime of confusion, and the amount of related penalties. This study applies deep learning-based algorithms to the problem of prediction and fine prediction and obtains a model with higher accuracy indicators through the adjustment of the existing corpus training model. Moreover, it attempts to solve two primary problems in forecasting legal judgments.

Problem 1: The amount of fine is not uniform.
Problem 2: Easy to confuse crimes.

The amount of fine related to the above problems is not uniform. This study assumes that the amounts involved are first uniformly converted into Arabic numerals, thereafter, a mapping table is constructed for the amounts, and finally, the numbers are denoised to solve the problem. Regarding the confusion of crimes

aforementioned in Problem 2, this study assumes that an additional input is added to solve the problem by using the input as a keyword vector for the case description.

Related Studies

This section examines the research development of artificial intelligence (AI) in legal judgment prediction and reviews the literature and work related to machine learning (ML) and deep learning (DL) models in legal judgment prediction. These studies serve as the theoretical basis for the present study.

Studies on Attitudes Toward the Use of Artificial Intelligence in Legal Judgment Prediction

The origin of AI in the legal field can be traced back to the early expert systems developed in the 1980s. These systems used knowledge representation techniques and inference rules to model legal reasoning processes (Leith, 2016; Wagner, 2017). Legal professionals and scholars have recognized the potential benefits of AI in improving the speed and accuracy of legal analyses. The increasing availability of large legal databases and advancements in computational power have facilitated the development of AI models that can process and analyze vast amounts of legal information (Susskind & Susskind, 2015). In 2018, the European Commission presented the "Ethics Guidelines for Trustworthy AI," emphasizing the importance of transparency, accountability, and fairness in AI applications (Nikolinakos, 2023). This document influenced subsequent discussions and initiatives in the field, promoting the responsible development and use of AI in legal decision-making processes.

However, critics have raised concerns about potential biases in training data, lack of interpretability in AI decision-making processes, and ethical implications. As O'Sullivan et al. (2019) proposed, AI should be used judiciously to ensure transparency, accountability, and adherence to legal ethics. Mainstream scholars have expressed both enthusiasm and caution regarding their application, stressing the need for a complementary role of AI alongside human judgment (Sourdin, 2018). Overall, the trend in AI in legal judgment prediction research is toward the use of more sophisticated algorithms, with machine learning and deep learning models demonstrating promising results in predicting legal judgments.

Studies on Machine Learning Models for Legal Judgment Prediction

The study on early machine learning models for legal judgment predictions, made predictions using text

classification (S. Li et al., 2020). Gonçalves and Quaresma (2005) assessed the importance of the pre-processing phase in text classification. They applied a Support Vector Machine (SVM) paradigm to two datasets. They evaluated various techniques for feature reduction/construction, feature subset selection, and term weighting to identify the document representation that yields the highest SVM performance for each dataset. Similarly, Hachey and Grover (2006) discussed the potential of summarization techniques in legal information management systems using rhetorical annotation schemes to help structure summaries to meet the needs of different types of users. Later, a machine learning model was used to improve the accuracy of legal judgment predictions by manually labeling cases and extracting word-level characteristics (Chen et al., 2022). For example, Lin et al. (2012) proposed a classification and sentencing forecast of a case using the labels of legal elements. Medvedeva et al. (2020) proposed using machine learning and natural language processing techniques to predict the decisions of the European Court of Human Rights (ECtHR). They demonstrated the potential of language analysis and automatic information extraction to predict judicial decisions with an accuracy of 75%.

However, machine learning models often struggle to capture intricate and nonlinear relationships in complex datasets (Ong et al., 2019). Deep learning models, with their multiple layers of interconnected neurons, are better equipped to capture and model complex patterns and dependencies in data. Therefore, the development of deep learning models has received attention, as they have demonstrated improved performance compared with traditional machine learning models in predicting legal judgments (H. Zhang et al., 2023).

Studies on Deep Learning Models for Legal Judgment Prediction

To better analyze the semantic and syntactic features of legal texts and incorporate contextual information surrounding legal cases, the exploration of deep learning multi-fusion models offers a promising direction for future research on legal judgment prediction (Zahir, 2023). Scholars have used deep learning models to study legal judgment predictions, which are generally divided into two categories (Wang et al., 2020). One is to use more complicated and novel deep learning models and algorithms for providing the accuracy of legal judgment prediction. For example, Hu et al. (2018) introduced the legal attributes of charges into the charge prediction task. They proposed a multi-task learning framework that jointly infers the attributes and charges of a case, using an attribute attention mechanism to learn attribute-aware fact representations. Zhu, Ma, et al. (2020)

focused on multi-charge prediction in legal judgment documents, proposed a novel hierarchical nested attention structure model, and used the LSTM attention mechanism to improve the accuracy of multi-charge classification. The other type is how to introduce legal provisions, legal professional terms, and even specific scenarios into specific in-depth learning models, resulting in the relevant characteristics of extraction to be more legal to discriminate. For example, Ye et al. (2018) proposed a label-conditioned Seq2Seq model with attention. This model decodes court views conditioned on encoded charge labels, thereby generating charge-discriminative court views. The experimental results demonstrated the effectiveness of the proposed method. Wang et al. (2020) considered judgment prediction as a multi-label text classification problem and trained word embeddings using BERT on a large dataset of judicial case text files. They combined BERT with other deep learning models such as CNN, LSTM, DPCNN, and RCNN to predict judgments in judicial cases. Experimental results demonstrate that deep learning models based on BERT achieve significant improvements compared with the six baseline methods, with an increase in prediction accuracy of 8% to 10%. The proposed approach has the potential to assist judges and lawyers in making accurate judgments.

Chief Task and Model Classification Method

Chief Task

In recent years, the Chinese Computer Society has organized several large-scale big data analyses and applied artificial intelligence models to solve problems in various fields. Many high-quality legal referee documents have been published since then. Forecasting research offers several possibilities for this purpose. This study applies deep learning multi-labeling and multi-class classification task models to the problem of prediction and fine prediction, and adjusts the parameters of the existing corpus training model to obtain a model with a higher accuracy index.

Model Classification Method

Cyclical neural network is a category of recursive neural networks (RNN) based on sequence data as the input, recursions in the evolution direction of the sequence, and all nodes connected by a chain connection (Goodfellow et al., 2016). RNN are highly effective in analyzing data with sequence characteristics. They can extract the timing and semantic information in the data, therefore, they have begun to be widely used in natural language processing. The sequence characteristics are in accordance with the time order, logical order, and so on. Convolutional neural networks are a type of feed neural
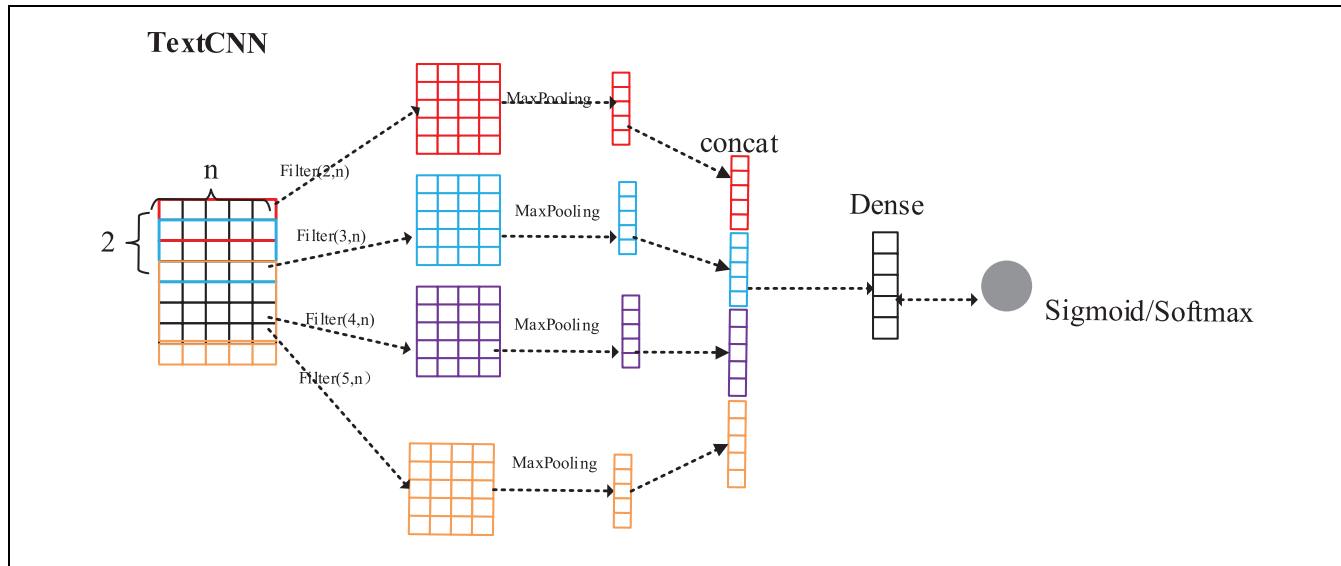
Figure 1. Structure of TextCNN model.

network that contains convolutional calculations, has a deep structure, and is a representative algorithm of deep learning. Convolutional neural networks are widely used in language and image recognition and natural language processing. This experiment adopts this method and improves the algorithm as a basis for subsequent research. The classification methods used in this study are the TextCNN, Wide & TextCNN, and TextDenseNet models.

TextCNN Model. Kim (2014) proposed a classic text classification algorithm, TextCNN, which is a commonly used model for text classification. The common procedure is to set the convolution kernel as $2 \times n$ to convolve two adjacent words. Multiple convolution kernels can generate multiple feature maps, and finally, Max Pooling is used to reduce the model complexity and extract the key information of the feature map. Moreover, related papers are found in which different kernels can be defined to simultaneously construct multiple feature maps, and thereafter, concatenate the features learned under each kernel to obtain better results. The structure of the TextCNN model is illustrated in Figure 1.

Wide & TextCNN Model. The Wide & Deep model proposed by Zheng et al. (2012) is applied to the field of recommendation systems, establishing a logical regression model for some features, and some features to establish a neural network model. It is a fusion model. The use of Wide and Deep parts to handle the characteristics of different sources can be regarded as another method for processing the polygonal characteristics. Later, using the idea of the Wide & Deep model, the Wide & TextCNN

was proposed, which is a modification of TextCNN. In the data analysis part, a series of features were counted, called the Wide part. The red dotted box is the artificial statistical features neural network structure, and three layers of full links will be used to learn the content of the wide part and add BatchNormalize after each full link to prevent the gradient disappearance problem. Wide & TextCNN can realize the combination of artificial and automatic features of neural networks, which improves the effect of CNN. The Wide & TextCNN model is presented in Figure 2.

TextDenseNet Model. Traditional convolutional networks have only one connection per layer in a forward process. ResNet adds residual connections to increase the flow of information from one layer to the next, and FractalNets repeatedly combine several sequences of parallel layers with different numbers of convolutional blocks to increase the nominal depth while maintaining the short path of the network forward propagation. DenseNet proposes a new connection model, dense connections based on this model. DenseNet uses dense connections to form three dense blocks, which are then convolved and pooled into several layers. The dense connections in the dense block are implemented using the concat. The structure of the TextDenseNet model is illustrated in Figure 3.

Text Pre-Processing

Data Processing. The original corpus often contains redundant information that is not suitable for direct training, and usually needs to be cleaned first. In this
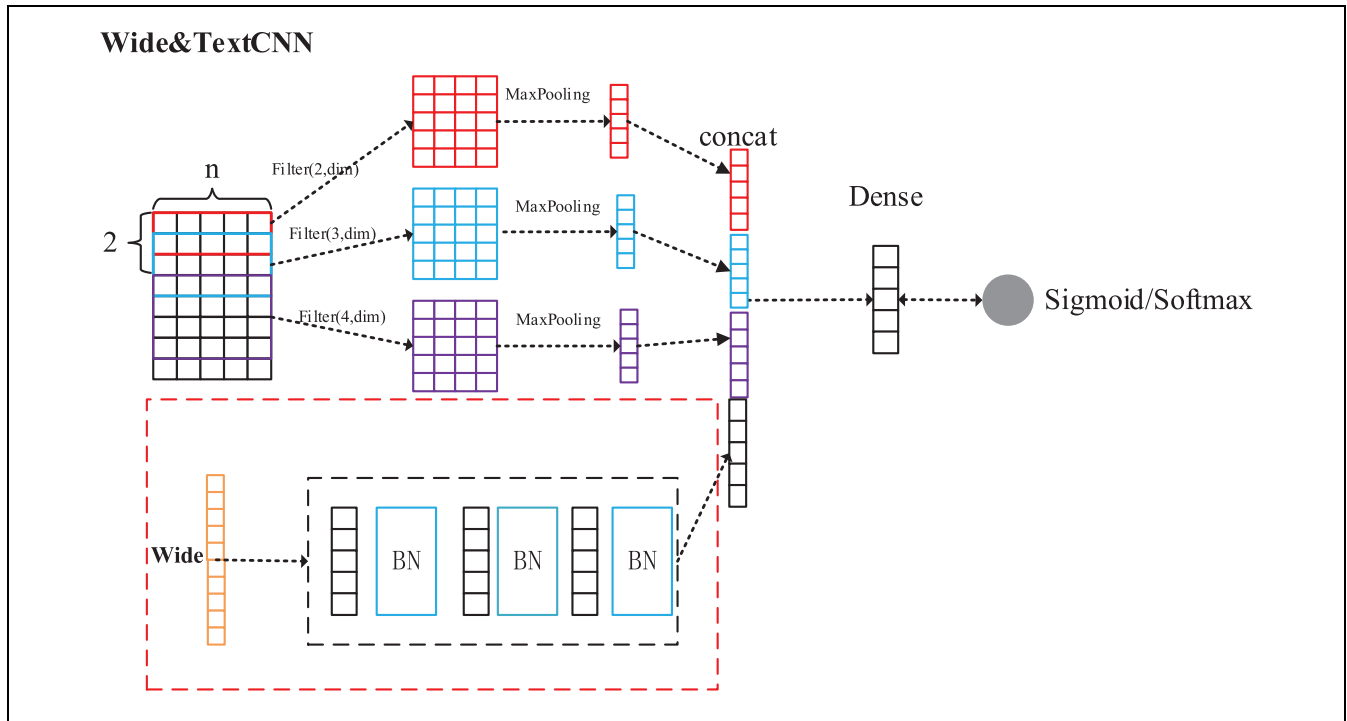
Figure 2. Structure of Wide&TextCNN model.

study, the text in the case corpus contained many special symbols, garbled codes, stop words, and other irregularities, particularly the amount and the number of regulation entries. For example, the following case is that of the public prosecutor, Beitang District People's Procuratorate of Wuxi City: "The defendant ... Criminal detention for suspected credit card fraud. In May 2008, the defendant... Applied for a credit card with a credit limit of 28,000 yuan... Repayment to CITIC Bank of 4519.81 yuan, 3345.85 yuan, 1000 yuan, respectively... The truthful confession of his crime... Family members withdrew \$19,100 from the account on behalf of ... Offend the provisions of Article 263, Article 120." The above cases of 4519.81 and 3345.85 yuan are in the same range of fines, however, are treated as two terms after the division of words, as well as the Article two hundred and sixty-three may be divided into two hundred and six and thirteen.

To address these issues, the data need to be pre-processed. First, the amount involved in the text is uniformly converted into Arabic numerals, and then an amount mapping table is constructed to map the amount of an interval to a uniform identifier. In addition, the regulation entries are filtered out by regular expressions, and the numbers are noise-reduced. The amount mapping Table 1 is as follows.

By amount implicitly, the above case can be converted after sub-word to "... The defendant applied for a credit line QL credit card... Repayment QH QG QE... Withdrawal of stolen funds QK Offense Two hundred and sixty-three, one hundred and twenty..."

Statistical Analysis. Statistical analysis was performed on the case data to determine the length distribution under each penalty category, and it was found that the heavier the penalty, the longer the length. This is illustrated in Figure 4.

Overall Framework

In this study, the input text data of each case are cleaned and divided into words, then converted into feature vectors and input into the neural network, and the classification results of fines and penalties are output using the classifier layer, as presented in Figure 5.

Experiment on AI Judge Prediction Case Analysis Based on Fusion Model

Experimental Data Set

The CCF Big Data and Computing Intelligence Competition (BDCI) is sponsored by the Chinese Computer Society and is among the most authoritative big data events in China. The contest aims to meet the needs of big data key industries and application fields. It
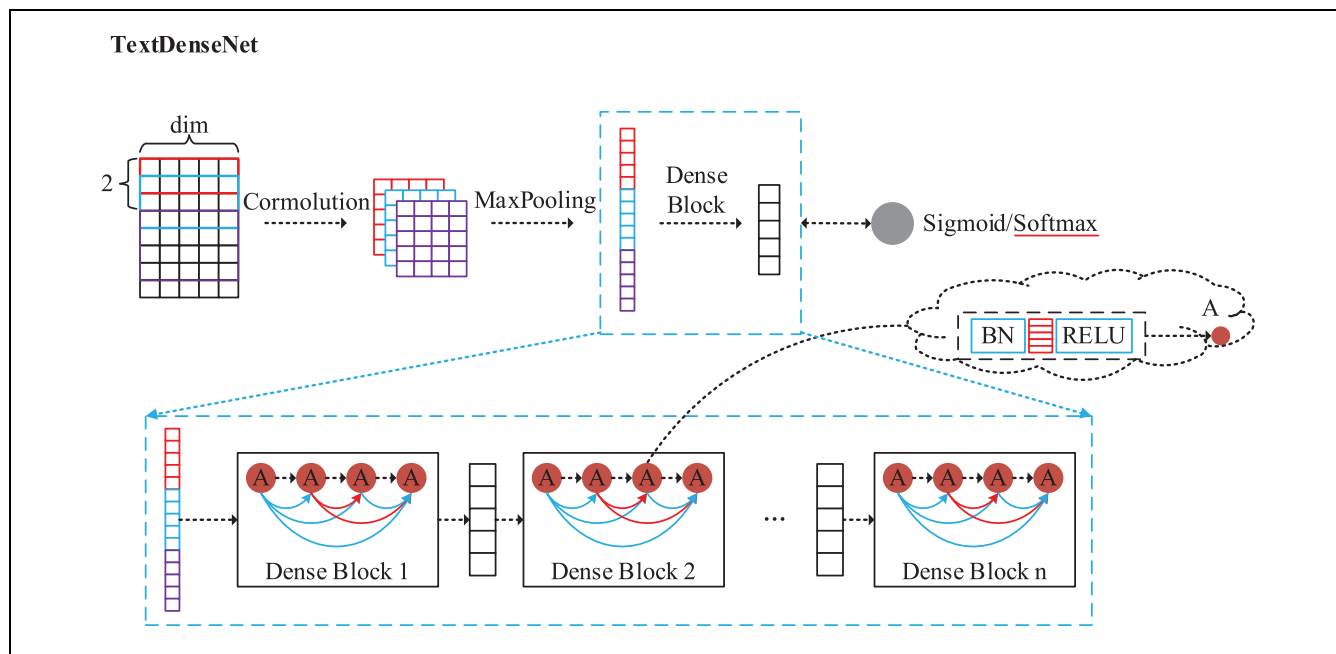
Figure 3. Structure of TextDenseNet model.

Table 1. Amount Mapping Table.

Amount	Identifiers
<30	QA
[30,100)	QB
[100,1000)	QC
[1000,2000)	QD
…	…
[50000, +)	QP

Figure 4. Distribution of the length of penalty categories.

is guided by cutting-edge technologies and industrial applications in big data and the goal of promoting industrial development and upgrading with big data. The development of big data technology and industrial ecology should be promoted. One of the tasks of the event provided the participants with several legal referee documents with labels, which promoted the application and development of natural language processing technology in legal referee documents, and has the reference significance and value of studying legal judgments and predictions. The experimental data source used for this study was the dataset of the BDCI2017 preliminary round, including method, training, and testing data. All formats were .txt formats. Among them, there were 452 legal datasets, 40,000 training datasets, and 10,000 test sets. The training set was divided into a training and validation set in the ratio of 9:1, and the validation set was used to select the optimal parameters as well as the optimal number of rounds; 10 models were trained using different models according to 10 different SEEDs. The last
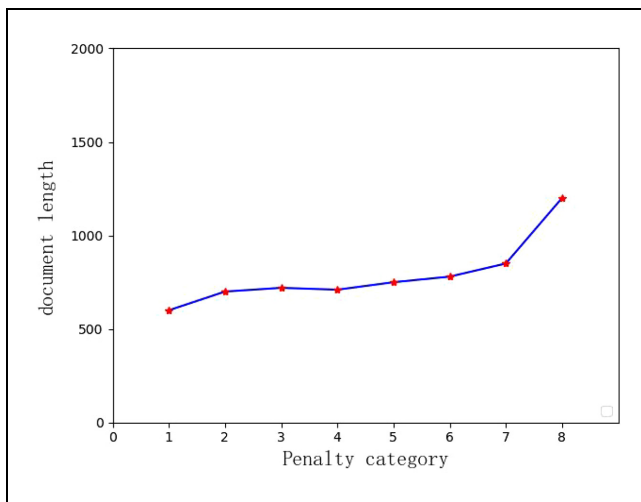
three models were fused according to the ratio of 0.75, 0.15, and 0.10.

Experimental Design and Parameter Selection

This study pre-processed the case dataset according to the method aforementioned, and then input the processed data and converted the vocabulary into a vector to represent it by embedding the overall embedding matrix, which represent the information of the whole case file, after which the neural networks TextCNN, TextRNN, Wide&TextCNN, and TextDenseNet models
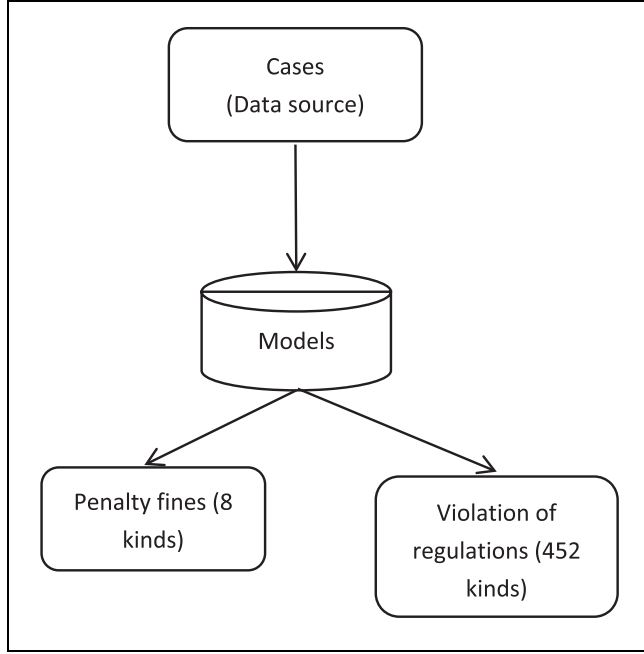
Figure 5. Modeling diagram.

Figure 6. Classification model design diagram.

were passed. To complete the task of text classification, this study added Sigmoid and SoftMax layers to the output layer to perform regulation and penalty classification, respectively, for the input cases. The design is illustrated in Figure 6.

In this experiment, a model with high classification accuracy was obtained by adjusting the values of the above five parameters without considering running time and memory problems. The default preset values of the five parameters were as follows: epochs were set to 4, batch_size was set to 16, pad_size was set to 1000, learning_rate was set to 1×10^{-5}, and hidden_size was set to 768 (Table 2).

Evaluation Indicators

The key to evaluating the model performance is selecting the right model performance evaluation indicator. This study uses Micro-Averad F1 and Jaccard similar coefficient values as its evaluation indicators, for the classification task of prediction and fine prediction. In this study, the fine classifier uses the Micro-Averaged F1 metric to measure model performance and the French classifier uses the Jaccard similarity coefficient measure,

(1) Micro-Averaged F1 calculation method:

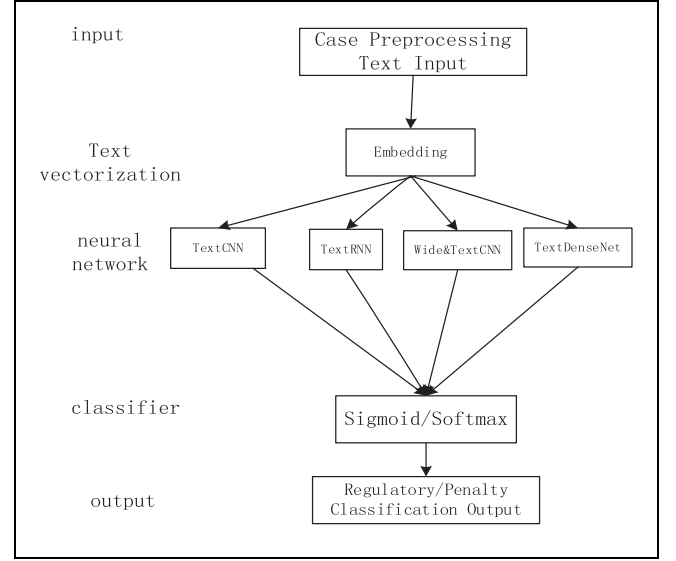$$micro - avg - F1 = \frac{1}{N}\sum_{i=1}^{m} w_i f_i \qquad (1)$$

where m denotes the number of categories, N the total number of samples, w_i the number of test samples belonging to category i, and f_i the F1 value for category i.

(2) Calculation of Jaccard similarity coefficient:

$$P_i = \frac{\left| L^i \cap L_g^i \right|}{\left| L^i \cup L_g^i \right|} \qquad (2)$$

$$p = \frac{1}{N}\sum_{i=1}^{n} p_i \qquad (3)$$

where L^i is the set of real legal provisions for case i and L_g^i is the set of legal provisions for case i of the test result.

Analysis of Results

The experimental results revealed that deepening the convolution in the form of DenseNet did not have a positive effect on the results, and the training time became increasingly slower. However, adjusting the full-link structure after concat significantly impacted the results. A traditional concat, followed by a full linkage, is insufficient to exploit this information. Therefore, the features were deepened after pooling instead of deepening the convolution. An A-structure, which comprises BN, full link, and RELU, was established so that the feature activation can avoid gradient disappearance as much as possible, and each of our A-structures uses the information

Table 2. Model Parameter Configuration.

Parameters	Parameter name	Parameter value
epochs	Number of training sessions	4
batch_size	Batch size	16
pad_size	Sentence length	1000
learning_rate	Learning rate	1×10^{-5}
hidden_size	Number of hidden layers	768

Table 3. Comparison of the Results of Different Neural Network Models.

Task name	Penalty forecast	Law forecast
Models	Micro-Averaged F1	Jaccard
TextCNN	0.5892	0.7264
TextRNN	0.6031	0.7135
Wide&TextCNN	0.6378	0.7536
TextDenseNet	0.7086	0.8175

of all the previous A-structures so that the information of each layer can be close to the output layer during deep learning. Each dense block is composed of four A-structures. Four dense blocks are the best in terms of penalty and one to two in terms of regulations. In addition, the effects of the various models were compared. The traditional model TextCNN/RNN was the worst, Wide & TextCNN was better than the traditional TextCNN/RNN, and TextDenseNet was much better than the first three model structures (Table 3).

The amount of fines related to the aforementioned problems are not uniform. In this study, by converting the amount involved into Arabic numerals, the mapping table of the construction amount is finally solved by the number of noise reductions; Problem 2, which is prone to confusion, adopts an input. Compared with the unprecedented original single-input model, the keyword vector described, improved the performance of the model described by the case description.

Conclusion and Influence

This study aimed to predict legal judgments. The primary research content included predictions and fine predictions. This experimental data used the dataset of BDCI2017 events, which primarily analyzed the text data described by the case, filtered out the numbers that are not related to the fine, and classified the different quota ranges of the fines, and the amount of penalty. The keyword vector described by the case was used as an input, which reduced easily confusing error rates. Various fusion models were adopted, and several deep learning models (TextCNN, TextRNN, Wide & TextCNN, and TextDenseNet) were analyzed and compared. The experimental results revealed that the TextDenseNet model has the highest accuracy and proves the prediction of applicable legal judgments.

Theoretically, this study used deep learning models to analyze datasets of legal judgment documents. The training classification model effectively predicted the case and its category. This has improved the accuracy and reliability of legal judgment prediction systems. Many tasks in the field of legal judgment prediction are primarily based

on text analysis and research and require semantic analysis. It primarily involves natural language processing (NLP) technology, such as automatic conviction, which involves text classification and semantic analysis, and intelligent research.

In practice, a legal judgment prediction model can improve the efficiency of judicial personnel in handling cases. For example, technologies such as dossier retrieval and case recommendation can help judicial personnel promptly identify relevant cases and laws and regulations to improve the work efficiency of judicial personnel and facilitate judicial justice. For example, technologies such as deep learning conviction are typically based on the characteristics of information learning, and conviction is based on several historical cases. Thus, the possibility of human errors could be reduced to a certain extent.

Limitation and Future Research

Limitations

1. To train the model fully, this study focused on the training of several common case data samples without focusing on the sample category of some special cases. In the data pre-processing stage, models can predict the most common crimes and laws, however, they are unsatisfactory in uncommon crimes and predictions. In real life, human judges encounter unusual cases that may involve uncommon crimes and laws. To be an effective judicial auxiliary tool, the legal judgment prediction should assist the judge in real trial cases far more than the present status. Current prediction performance is of sample category.
2. The dataset used in this experiment comprised 50,000 legal referee documents, and there are currently more than two million legal referee documents. If sufficient computing resources are available, the results of all the datasets may be convincing.
3. Many applicable clauses inevitably have vague places, neither directly nor qualitatively, and it is

difficult to decide which clause machines are applicable. This affects the judgment accuracy.

4. The scope of sentencing of some provisions is large, and artificial intelligence cannot provide accurate suggestions, and factors of emotion and humanities are not considered. Each judge differs in principle. Smart software/AI temporarily lacks this feature.

Future Research

Future studies should simulate the judgment prediction steps of the judge, combined with the information of the judgment as much as possible and not simply the description of the case by the referee, which makes the results more accurate and in accordance with reality. A legal judgment prediction system established in this manner can only be more convincing and realize justice intelligence in the true sense.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iD

Ping Ti (iD) https://orcid.org/0009-0004-8505-4354

Data Availability Statement

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

References

Ahmad, S., Asghar, M. Z., Alotaibi, F. M., & Al-Otaibi, Y. D. (2022). A hybrid CNN + BILSTM deep learning-based DSS for efficient prediction of judicial case decisions. *Expert Systems with Applications*, 209, 118318.

Alfaia Sampaio, E., Seixas, J. J., & Gomes, P. J. (2022). *Artificial Intelligence and the judicial ruling*. Themis Competition 2019, Semi-Final D, Team Portugal I.

Ashley, K. D. (2018). Automatically extracting meaning from legal texts: Opportunities and challenges. *Georgia State University Law Review*, 35, 1117.

Bruninghaus, S., & Ashley, K. D. (2003, June). *Predicting outcomes of case based legal arguments* [Conference session].

Proceedings of the 9th International Conference on Artificial Intelligence and Law* (pp. 233–242).

Chaphalkar, N. B., Iyer, K. C., & Patil, S. K. (2015). Prediction of outcome of construction dispute claims using multilayer perceptron neural network model. *International Journal of Project Management*, 33(8), 1827–1835.

Chen, J., Du, L., Liu, M., & Zhou, X. (2022). Mulan: A multiple residual article-wise attention network for legal judgment prediction. *Transactions on Asian and Low-Resource Language Information Processing*, 21(4), 1–15.

China News Network. (2019). *Sogou and Beijing Internet Court launch world's first AI virtual judge*. https://baijiahao.baidu.com/s?id=1637476278209748950&wfr=spider&for=pc

Gavaghan, C. (2017). Lex Machina: Techno-regulatory mechanisms and rules by design. *Otago Law Review*, 15, 123

Gonçalves, T., & Quaresma, P. (2005, July). *Evaluating preprocessing techniques in a text classification problem* [Conference session]. *Proceedings of the Conference of the Brazilian Computer Society*.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* (Vol. 1, pp. 326–366). MIT Press.

Hachey, B., & Grover, C. (2006). Extractive summarisation of legal texts. *Artificial Intelligence and Law*, 14, 305–345.

Hu, Z., Li, X., Tu, C., Liu, Z., & Sun, M. (2018, August). *Few-shot charge prediction with discriminative legal attributes* [Conference session]. *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 487–498).

Joshua. (2020). *Lex Machina: America's "Legal Big Data" product*. https://zhuanlan.zhihu.com/p/326547648

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17, 1–9.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Kufakwababa, C. Z. (2021). *Artificial intelligence tools in legal work automation: The use and perception of tools for document discovery and privilege classification processes in Southern African legal firms* [Doctoral dissertation, Stellenbosch University].

Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management*, 21(4), 331–340.

Legaldaily. (2022). *Hangzhou Internet Court builds an Internet judicial digital intelligence brain to comprehensively improve the level of judicial intelligence*. http://www.legaldaily.com.cn/IT/content/2022-07/21/content_8755704.html

Leith, P. (2016). The rise and fall of the legal expert system. *International Review of Law, Computers & Technology*, 30(3), 94–106.

Lettieri, N., Altamura, A., Giugno, R., Guarino, A., Malandrino, D., Pulvirenti, A., Vicidomini, F., & Zaccagnino, R. (2018). Ex machina: Analytical platforms, law and the challenges of computational legal science. *Future Internet*, 10(5), 37.

Li, J. (2019). The legal profession of China in a globalized world: Innovations and new challenges. *International Journal of the Legal Profession*, 26(2–3), 217–264.

Li, S., Zhang, H., Ye, L., Guo, X., & Fang, B. (2019). Mann: A multichannel attentive neural network for legal judgment prediction. *IEEE Access*, 7, 151144–151155.

Li, S., Zhang, H., Ye, L., Su, S., Guo, X., Yu, H., & Fang, B. (2020). Prison term prediction on criminal case description with deep learning. *Computers, Materials and Continua*, *62*(3), 1217–1231.

Lin, W. C., Kuo, T. T., Chang, T. J., Yen, C. A., Chen, C. J., & Lin, S. D. (2012). Exploiting machine learning models for Chinese legal documents labeling, case classification, and sentencing prediction. *Processsdings of ROCLING*, *17*(4), 140.

Maharg, P. (2017). Let's get digital. Paul Maharg explores the potential for AI and legal education. *New Law Journal*, *167*(7731), 20.

Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, *28*, 237–266.

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*, *19*(6), 1236–1246.

Murimi, R. (2021). *When humans judge other humans using machines*. MIT Computational Law Report. https://law.mit.edu/pub/whenhumansjudgeotherhumansusingmachines

Nikolinakos, N. T. (2023). Ethical principles for trustworthy AI. In N. T. Nikolinakos (Ed.), *EU policy and legal framework for artificial intelligence, robotics and related technologies—The AI act* (pp. 101–166). Springer.

Ong, K., Haw, S. C., & Ng, K. W. (2019, November). *Deep learning based-recommendation system: An overview on models, datasets, evaluation metrics, and future trends* [Conference session]. *Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems* (pp. 6–11).

O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., Holzinger, K., Holzinger, A., Sajid, M. I., & Ashrafian, H. (2019). Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *15*(1), e1968.

Semmler, S., & Rose, Z. (2017). Artificial Intelligence: Application today and implications tomorrow. *Duke Law & Technology Review*, *16*, 85.

Sourdin, T. (2018). Judge v Robot?: Artificial intelligence and judicial decision-making. *The University of New South Wales Law Journal*, *41*(4), 1114–1133.

Susskind, R. E. (1991). *Expert systems in law: A jurisprudential inquiry*. Clawendon Press.

Susskind, R. E., & Susskind, D. (2015). *The future of the professions: How technology will transform the work of human experts*. Oxford University Press.

Wagner, W. P. (2017). Trends in expert system development: A longitudinal content analysis of over thirty years of expert system case studies. *Expert Systems with Applications*, *76*, 85–96.

Wang, Y., Gao, J., & Chen, J. (2020, October). *Deep learning algorithm for judicial judgment prediction based on BERT* [Conference session]. 2020 5th International Conference on Computing, Communication and Security (ICCCS) (pp. 1–6). IEEE.

Ye, H., Jiang, X., Luo, Z., & Chao, W. (2018). Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions. *arXiv preprint arXiv:1802*.08504.

Zahir, J. (2023). Prediction of court decision from Arabic documents using deep learning. *Expert Systems*, *40*(6), e13236.

Zeng, Y., & Pan, W. (2023). *Shanghai Financial Court adopts "test court" for the first time to resolve financial disputes*. Kankannews Knews. https://baijiahao.baidu.com/s?id=175 8978806189984177&wfr=spider&for=pc

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.

Zhang, N., & Xu, J. (2019). Computational law: Interdisciplinary research between law and artificial intelligence. *Modern Law Science*, *41*(6), 78–91.

Zheng, F. R., Miao, D. Q., Zhang, Z. F., & Gao, C. (2012). News topic detection approach on Chinese microblog. *Computer science*, *39*(1), 138–141.

Zhu, K., Guo, R., Hu, W., Li, Z., & Li, Y. (2020). Legal judgment prediction based on multiclass information fusion. *Complexity*, *2020*, 1–12.

Zhu, K., Ma, B., Huang, T., Li, Z., Ma, H., & Li, Y. (2020). Sequence generation network based on hierarchical attention for multi-charge prediction. *IEEE Access*, *8*, 109315–109324.