Refining Judgment Prediction in European Human Rights Cases with Deep Learning

Meng-Luen Wu
Dept. of CSIE
Tamkang University
New Taipei City, Taiwan (R.O.C.)
158769@mail.tku.edu.tw

Hyun-Min Choi
Dept. of CSIE
NTUST
Taipei, Taiwan (R.O.C.)
m10715809@mail.ntust.edu.tw

Chao-Wei Liu
Dept. of CSIE
Tamkang University
New Taipei City, Taiwan (R.O.C.)
612410653@mail.tku.edu.tw

***Abstract*—The integration of artificial intelligence (AI) into the legal field represents a transformative opportunity to enhance the efficiency and accuracy of judicial processes, particularly in the realm of human rights. As legal practitioners face overwhelming workloads and the need for informed decision-making, AI technologies can streamline case management and improve access to justice. This study aims to explore advanced AI methodologies that can effectively analyze complex legal texts, thereby promoting fairness and transparency in legal systems.**

To achieve this, we propose an innovative integrated deep learning approach that combines Convolutional Neural Networks (CNN), Bi-directional Gated Recurrent Units (BiGRU), and attention mechanisms for predicting judgments in human rights cases. Each of these components brings unique strengths: CNNs excel at extracting local features and patterns from text, BiGRUs effectively capture long-range dependencies and contextual information, and attention mechanisms enhance the model's ability to focus on the most relevant parts of the input data. By integrating these techniques, our model leverages their complementary strengths, resulting in a significant improvement in prediction accuracy compared to traditional machine learning methods. The experimental results demonstrate the model's effectiveness, showcasing its potential to automate judgment outcomes and support legal practitioners in making more informed decisions in human rights cases.

***Keywords*—Judgment prediction, Integrated model, BiGRU, Convolutional neural network (CNN), European court of human rights (ECHR)**

I. Introduction

The integration of artificial intelligence (AI) in the legal field is growing, enhancing efficiency and accuracy in judicial processes. AI systems help judges manage overwhelming workloads by streamlining case management and decision-making. In human rights case prediction, AI enables legal practitioners to analyze large data sets quickly, leading to more informed judgments and improved access to justice. Additionally, AI identifies patterns in cases, supporting the development of effective legal strategies and promoting fairness and transparency in the justice system.

This study explores advanced AI technologies for predicting judgments in human rights cases, comparing deep learning with traditional machine learning methods. It shifts from conventional techniques like Support Vector Machines (SVM) to more sophisticated approaches such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). The proposed integrated model combines Bi-directional Gated Recurrent Units (BiGRU), CNN, and Attention mechanisms to automate judgment predictions. By utilizing natural language processing (NLP), the model enhances the analysis of legal texts, improving predictive accuracy and supporting legal decision-making.

This paper presents a system based on an integrated deep learning method that combines Convolutional Neural Networks (CNN), Bi-directional Gated Recurrent Units (BiGRU), and Attention mechanisms for extracting, analyzing, and predicting judgments from the European Court of Human Rights (ECHR) dataset, which includes legal articles on various human rights cases. Improving the efficiency and accuracy of resolving these cases is vital for advancing human rights. The process involves pre-processing, vector representation, and deep learning techniques. This study details the architecture and functionality of the proposed integrated model and discusses the natural language processing (NLP) techniques used to enhance predictive accuracy, ultimately aiming to address human rights issues more effectively and equitably.

TABLE I
The Topics and Sections of Utilized Articles for Training

Article	Topic	Sections
Article 2	Right to life	procedure + facts
Article 3	Prohibition of torture	facts
Article 5	Right to liberty and security	facts
Article 6	Right to a fair trial	procedure + facts
Article 8	Respect for private and family life	procedure + facts
Article 10	Freedom of expression	procedure
Article 11	Freedom of assembly and association	procedure
Article 13	Right to an effective remedy	facts
Article 14	Prohibition of discrimination	procedure + facts

II. Related Work

In the realm of judgment prediction for the European Court of Human Rights (ECHR) dataset, traditional machine learning methods have predominantly utilized algorithms such as Support Vector Machines (SVM). SVMs are favored for their effectiveness in managing high-dimensional data and their capability to define complex decision boundaries in judicial contexts. Notable studies, including those by Aletras *et al*. [1], Medvedeva *et al*. [2], and Alexandre *et al*. [3], have demonstrated the successful application of SVMs in legal text classification, employing various N-gram and topic modeling techniques. These studies highlight the SVM's foundational role in legal judgment prediction, showcasing both its strengths and limitations in accurately forecasting case outcomes across different jurisdictions.

The introduction of BERT (Bidirectional Encoder Representations from Transformers) has marked a significant advancement in natural language processing, particularly for legal texts. BERT's pre-trained models can be fine-tuned to address the specific challenges posed by legal language, enhancing comprehension of complex legal precedents. The HIER-BERT model, developed by Chalkidis *et al*. [4], combines BERT's capabilities with hierarchical attention mechanisms, allowing for a deeper understanding of the intricate relationships within legal texts. This innovative approach has led to improved accuracy in predicting legal outcomes, demonstrating the effectiveness of leveraging advanced language processing techniques in the legal domain.

Attention mechanisms have also emerged as a powerful tool in judgment prediction, particularly through models like the Hierarchical Attention Network [5]. Chalkidis *et al*. [4] applied attention mechanisms to both BiGRU and BERT models, utilizing pre-trained word embeddings such as GloVE. Their findings revealed that these models achieved impressive accuracy rates of 87.0% and 90.4%, respectively, underscoring the importance of attention in enhancing prediction performance on the ECHR dataset. This highlights the potential of attention mechanisms to focus on relevant input features, thereby improving the overall effectiveness of judgment prediction models in the legal context.

III. Integrated Method for Judgement Prediction

In this section, we propose a comprehensive method for predicting judgments in human rights cases by integrating advanced text processing techniques and deep learning models to enhance prediction accuracy. Our approach leverages Convolutional Neural Networks (CNN), Bi-directional Gated Recurrent Units (BiGRU), and attention mechanisms to effectively capture intricate patterns and contextual dependencies within legal texts, providing a robust framework for analyzing judgment documents and improving the reliability of case outcome predictions.

CNNs are particularly adept at extracting local features and patterns, making them efficient and robust in noisy environments. However, they may struggle with long-range dependencies and typically require fixed-size input, which can limit their effectiveness in processing variable-length legal texts. On the other hand, BiGRUs excel at capturing contextual information by processing sequences in both forward and backward directions, allowing them to handle variable-length inputs and maintain contextual awareness throughout the text. Yet, they can be more complex to train and are at risk of overfitting, especially with smaller datasets.

To address these limitations, we integrate an attention model, which enhances performance by allowing the model to focus on the most relevant parts of the input data. This mechanism helps identify key terms associated with violations or non-violations, improving the model's ability to capture nuanced contextual information. By combining CNN's local feature extraction capabilities, BiGRU's contextual understanding, and the attention model's ability to prioritize significant information, our integrated method achieves greater predictive accuracy and robustness, ultimately leading to more reliable predictions in human rights case outcomes.

The architecture of the integrated judgement prediction system id depicted in Fig. 1, which integrates CNN, BiGRU, and the attention mechanism. Details of each component are described below.

A. BiGRU layer

In our model, we start with the BiGRU layer, where input words are transformed into word embeddings using Word2Vec to capture semantic relationships in a lower-dimensional space. This transformation allows the model to represent words in a way that reflects their meanings and relationships. The BiGRU layer processes sequential text inputs in both forward and backward directions, effectively updating and resetting information to capture long-range dependencies and provide a comprehensive context for each article. The output from the BiGRU layer is then passed to the subsequent CNN layer, which extracts local features from the ECHR dataset. This combination of techniques enhances the model's ability to understand and analyze the complexities of legal texts, ultimately improving the accuracy of judgment predictions.

B. CNN layer

Next, we enhance the CNN to effectively process the features extracted from the BiGRU layer, focusing on capturing local relationships within the text to derive meaningful features. This section comprises two consecutive convolutional layers, each followed by max-pooling and a dropout layer to mitigate the risk of overfitting. The first convolutional layer utilizes a filter size of 3, while the second layer employs a larger filter size of 5, allowing the model to capture varying n-gram features from the input text.

In contrast to traditional CNN models that operate independently, our adaptation, as illustrated in Fig. 2, integrates the concept of hidden states to reflect the sequential nature of the data. This integration enables the model to retain contextual information while extracting local features. In this framework, c_t represents a vector in the hidden state sequence from the CNN layers at time t, while $\tanh(h_t)$ serves as a learnable activation function. Additionally, a denotes a probability vector, and r is calculated as a weighted average of h_t with weights determined by a_t. This architecture allows the model to effectively identify and prioritize significant local patterns within the text, which are crucial for accurate judgment predictions. The output from these layers is then passed to the Attention layer for further refinement.

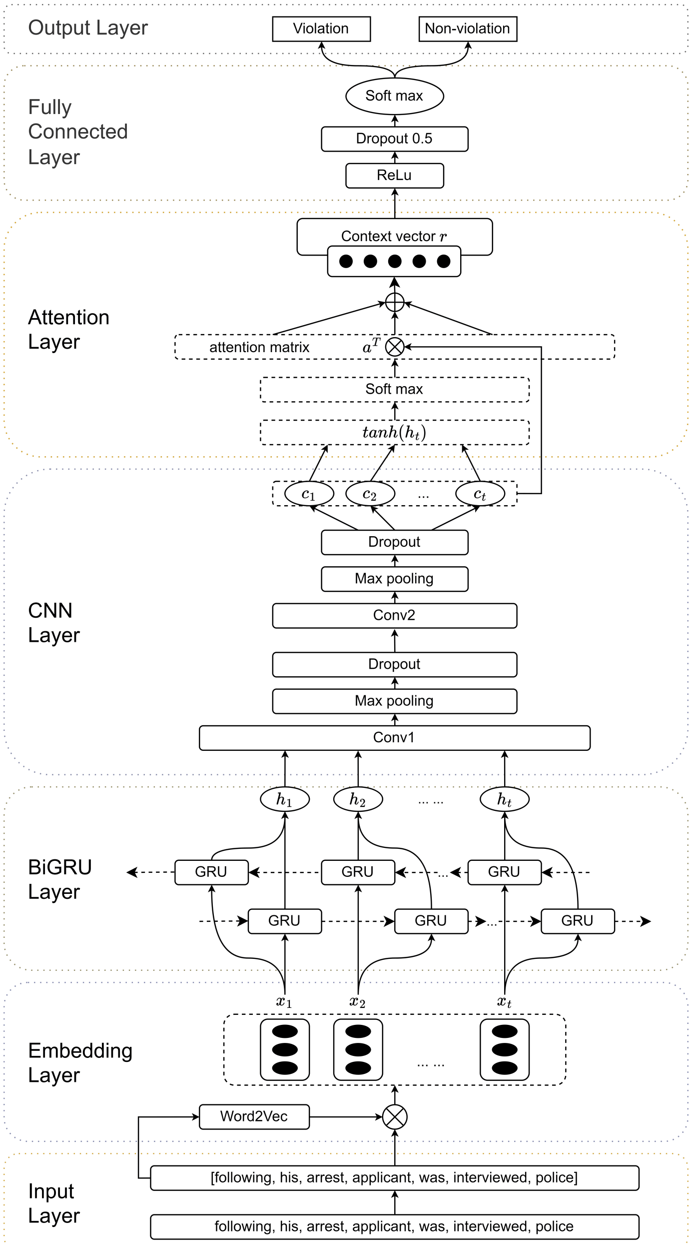

Fig. 1. The architecture of the proposed ensemble model.

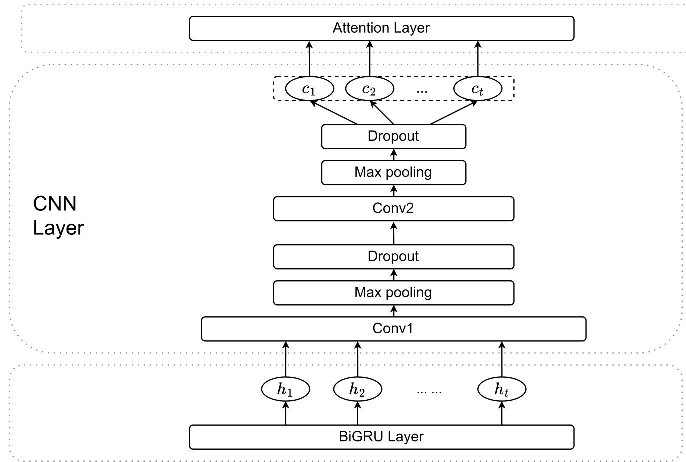

Fig. 2. Illustration of input and output of the CNN layer.

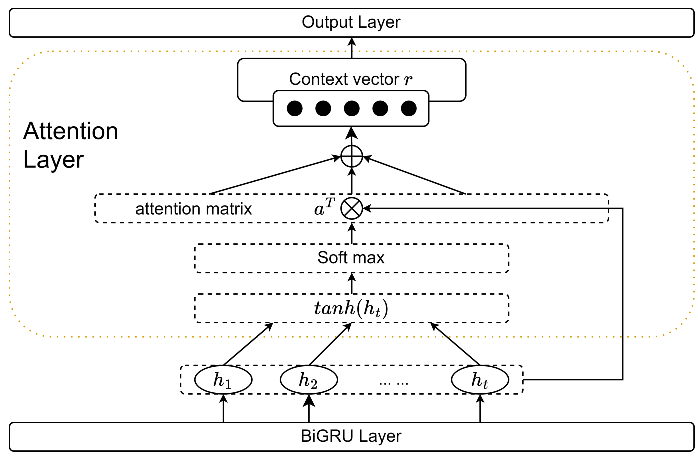

Fig. 3. Illustration of the Attention layer

C. Attention layer

The Attention mechanism highlights the essential elements within a sentence, enhancing the model's understanding of the context derived from the output of the CNN layer. Additionally, it provides insights into the specific words that the model prioritizes during the prediction process. The functionality of the Attention layer is depicted in Fig. 3, which visually represents how the mechanism operates to focus on the most relevant parts of the input data.

D. Fully-connected layer

The judgment prediction classification involves a fully-connected layer with two hidden layers and a dropout layer, followed by an output layer containing two neurons and a softmax function for prediction. The equations for the output layer are as follows:

$$\text{dense}_1 = \text{relu}\left(W_0 \times r + b_0\right) \tag{1}$$

$$\text{dense}_2 = \text{relu}\left(W_1 \times \text{dense}_1 + b_1\right) \tag{2}$$

$$y = \text{softmax}\left(W_2 \times \text{dense}_2 + b_2\right) \tag{3}$$

In this context, r denotes the attended sentence representation obtained from the attention layer, while W_i and b_i signify the weight matrices and bias vectors for $i = 0, \ldots, 2$. Throughout the learning process, back-propagation is utilized to adjust these parameters, which are crucial for optimizing the filter weights. The softmax function produces a probability distribution for the two classes: "violation" and "non-violation." To mitigate the risk of overfitting, we apply dropout, randomly deactivating a portion of the neurons during training, with a dropout rate set at 50

$$\text{Loss}\left(y_j, y_j'\right) = -\Sigma_j y_j * \log\left(y_j'\right) \tag{4}$$

where y_j represents the ground truth judgment for each case, and y'_j indicates the model's predicted judgment. The errors derived from the categorical cross-entropy are subsequently utilized to adjust the weights, thereby enhancing the learning process.

IV. Experimental Result

In this section, we present an overview of the ECHR dataset used for our case prediction, detail the parameters implemented in our model, and offer a thorough analysis of the results regarding judgment predictions on the ECHR dataset, differentiating between violations and non-violations. First, we introduce the ECHR dataset along with the parameters applied in our deep learning model. Next, we divide the dataset by individual Articles, performing training and testing separately for each Article.

A. ECHR Dataset

The ECHR dataset, developed by Medvedeva et al. [2], consists of textual cases organized under different Articles and classified as either "violation" or "non-violation." This dataset is carefully balanced between the two categories, which helps maintain the objectivity of our model and mitigates bias towards particular judgment outcomes. Below, we provide a table that outlines the data counts for violations and non-violations for each Article.

TABLE II
The ECHR Dataset [2]

	ECHR dataset		Total	Test set
	violation	Non-violation		(only violation)
Article 2	57	57	114	398
Article 3	284	284	568	851
Article 5	150	150	300	1118
Article 6	458	458	916	4092
Article 8	220	229	458	496
Article 10	106	106	212	252
Article 11	32	32	64	89
Article 13	106	106	212	1060
Article 14	144	144	288	44
Total	1566	1566	3132	8400

B. Parameters for deep learning model

The parameters chosen for this study are designed to optimize model performance. They include an embedding size of 300 dimensions, a categorical cross-entropy loss function, the Adam optimization method with a learning rate of 0.001, and a maximum sequence length of 600. These parameters are essential for effective training and accurate predictions in legal judgment outcomes.

The study also utilizes Word2Vec embedding with 300 dimensions to effectively represent the textual data. Additionally, a 10-fold cross-validation method is implemented to evaluate the model's performance and ensure robustness in assessing the predictive accuracy of the deep learning models for legal judgment prediction.

C. Result and Discussion

We evaluate our models using precision, recall, and F1-score metrics derived from a confusion matrix of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). These metrics are crucial for assessing model performance in legal judgment prediction. The formulas are as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

$$F1 - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

We evaluate the performance of our models and compare it with the method proposed by Medvedeva et al. [2]. This evaluation encompasses a range of Articles to analyze the predictive accuracy of the models across various legal contexts and case scenarios. The results are presented in Table III.

TABLE III
Experimental on Separate Articles

	Medvedeva et al. [2]	SVM	Integrated Model
Article 2	73 %	66.66 %	75.55 %
Article 3	80 %	78.66 %	84.44 %
Article 5	71 %	73.1 %	82.91 %
Article 6	80 %	79.38 %	90.00 %
Article 8	72 %	71.11 %	83.61 %
Article 10	61 %	61.9 %	78.82 %
Article 11	83 %	76.00 %	94.00 %
Article 13	83 %	82.14 %	79.41 %
Article 14	75 %	76.75 %	78.26 %
Average	75 %	73.96 %	83.00 %

The superior performance of our proposed integrated model compared to Medvedeva's model can be attributed to its advanced architecture, which integrates Bi-directional Gated Recurrent Units (BiGRU), Convolutional Neural Networks (CNN), and Attention mechanisms. This integration approach facilitates a more thorough analysis of contextual dependencies and the extraction of relevant features from the data, resulting in improved prediction accuracy.

In contrast, while the SVM-based model shows competitive results in certain areas, it may lack the robustness of the integrated model due to potential overfitting issues observed in its test accuracy for specific Articles. This suggests that the SVM model may struggle to generalize to unseen data, whereas the integrated model exhibits more stable performance across various scenarios.

V. Conclusion

In this paper, we focus on enhancing judgment prediction in human rights cases, specifically within the European Court of Human Rights (ECHR), using advanced deep learning techniques. By combining Bi-directional Gated Recurrent Units (BiGRU), Convolutional Neural Networks (CNN), and

Attention mechanisms, we have developed an integrated model that excels in capturing contextual dependencies and extracting relevant features for accurate judgment prediction. This model showcases significant improvements over traditional machine learning approaches, demonstrating superior robustness and generalization capabilities.

Notably, the integrated model achieved the highest performance with an average F1-score of 83%, indicating its effectiveness in accurately predicting case outcomes and handling both violation and non-violation classifications. The experimental results underscore the superiority of our integrated model in comparison to the SVM-based model, showcasing its ability to provide stable and reliable predictions across various legal contexts within the ECHR dataset. Our research not only contributes to the field of legal informatics by offering a scalable method for predicting case outcomes but also paves the way for future exploration in applying Natural Language Processing (NLP) techniques within the legal domain, thereby advancing decision-making processes in human rights cases.

ACKNOWLEDGEMENT

The authors thank the National Science and Technology Counsil of Taiwan (R. O. C.) for supporting this work in part under Grant NSTC 113-2221-E-032-023-.

REFERENCES

[1] N. Aletras, D. Tsarapatsanis, D. Preoţiuc-Pietro, and V. Lampos. Predicting judicial decisions of the european court of human rights: a natural language processing perspective. *PeerJ Computer Science*, 2:e93, Oct 2016.

[2] Masha Medvedeva, Michel Vols, and Martijn Wieling. Using machine learning to predict decisions of the european court of human rights. *Artificial Intelligence and Law*, 28(2):237–266, 2020.

[3] Alexandre Quemy. European court of human right open data project. *arXiv preprint arXiv:1810.03115*, 2018.

[4] Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. Neural legal judgment prediction in english. *arXiv preprint arXiv:1906.02059*, 2019.

[5] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489, 2016.