2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT) | 979-8-3315-1857-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/CE2CT64011.2025.10939596

Artificial Intelligence in Legal Practice: Enhancing Case Prediction and Legal Research

Manish Kumar
L N Mishra College of Business Management ,
Muzaffarpur , Bihar,
842001, India.
manishsirhere@gmail.com

J. Lalith Kumar
Vellore Institute of Technology, School of Law (VITSOL),
Chennai, India.
lalithkumar.j@vit.ac.in

Rahul J Nikam
Department of Law, Marwadi University
Rajkot-Morbi Highway, Rajkot
360003. India.
rahulsnikam@gmail.com

Jayant Maruti Hudar
DBA , BHS University Switzerland.
DBA in Data Science and AI.
PGDBM. B.E. Electronics.
Diploma in Computer Technology,
Navi Mumbai, India.
jayanthudar@gmail.com

Sameena Hasan
Lords Universal College of Law
Affiliated to University of Mumbai.
javsameena@gmail.com

Sandipta Padhee
SNIL, SOA University,
Bhubaneswar, India.
Sandipta.padhee@gmail.com

***Abstract*-Legal case prediction (LCP) and decision assistance strive to allow systems to forecast the outcome of legal cases after comprehending an explanation of the evidence, using artificial intelligence (AI) in legal practices. This work offers an enhanced supervision-based LCP framework that accounts for the consecutive dependency of every sub-task in the LCP activity. The empirical outcomes confirm the efficiency of the conceptual paradigm and the procedure surveillance process used in this structure. Initially, textual characteristics were extracted using the convolutional neural network (CNN) technique, and then data characteristics were reduced in size using the principal component assessment method. The predictive paradigm built around supervision analysis is then presented initially. When simulating the dependent connection between consecutive sub-data collections, supervision analysis is used to assure the accuracy of the acquired dependent data, while a genetic algorithm is used to maximize the variables and enhance the final case prediction efficiency. When evaluated with the standard approach, the proposed method produced outstanding outcomes on 4 separate publicly accessible legal databases. The implementation of autonomous LCP can not just help lawyers, attorneys, and other specialists make more effective legal decisions, but it additionally offers legal assistance to persons who lack legal understanding.**

Keywords: Legal Case Prediction, Artificial intelligence, Convolution neural network, Genetic algorithm, and Legal practices.

I. INTRODUCTION

Pattern recognition is a strong suit for humans. They are unable to quickly examine a vast database, though, particularly when the data doesn't seem to correspond. Evaluation of data and statistical approaches are two ways to support the identification procedure. However, certain datasets include so much data that it is challenging to conduct a thorough examination in trying to identify trends, even with the aid of scientific techniques [1]. Statistics that include statistics about court cases are an instance of this, as they comprise a lot of data that at first does not correspond [2].

For the average person, the intricacy and competence of legal papers are impossible obstacles, and the outcomes of conventional legal cases are possible only following the examination and comprehension by pertinent individuals [3]. According to initial research, professionals in legal research are created by relevant individuals to address the challenges that individuals face [4]. However, the procedure of creating these systems often requires an extensive number of regulations and descriptions, and maintaining the framework afterward will take a significant amount of time and effort [5]. A strong basis for smart legal research has been established by the numerous outstanding deep neural networking (DNN) architectures that have been suggested as deep learning (DL) techniques have matured [6]. One crucial smart legal subsection is the paper prediction. Lawyers' effectiveness in handling cases can be significantly increased by using the paper's prediction work, which can also forecast the law that will be used in the proceedings and serve as the foundation for the legal verdict on allegations jail sentences, penalties, and other penalties [7]. The use of artificial intelligence (AI) in the legal domain is legal case prediction (LCP), which attempts to allow robots to estimate the outcome of court proceedings after studying the facts [8]. In addition to helping attorneys, judicial officials, and other experts make more informed decisions, the ability to forecast legal analysis automatically can also help those without legal knowledge by offering legal assistance. As DL and NLP (natural language processing) technologies have advanced lately, a growing number of researchers are becoming interested in the study of LCP [9].

The study on LCP started years ago. The legal decision problem is currently thought of as a text categorization duty [10]. Numerous groundbreaking approaches have been offered by scientists who have studied the subject, primarily the DNN-based LCP technique, and the conventional DL technique [11]. Due to their limited flexibility to different circumstances and difficulty in applying to different case kinds, the majority of legal decision systems that rely on classic AI techniques explicitly label characteristics depending on particular case kinds. To accomplish LCP, the DNN-based LCP technique no more depends on a manually created model [12]. Instead, it uses convolutions or cyclic NN to gather contextual data and create characteristic representations depending on the case information. Even though the current techniques have produced reasonable forecasting effects, it is nevertheless challenging to significantly increase accuracy.

To solve this issue this study develops an LCP model based on supervision analysis (SA-LCP) using AI algorithms for legal research.

II. Literature Review

Years ago, studies on LCP started. The constraints on open instances mostly hindered early research, and data-based approaches only produced data for a restricted number of verdicts instead of accurate predictions [13, 14, 15].

The LCP work is now considered a textual classification work due to the advancement of AI and NLP tools. As a result, the majority of LCP activities are task-specific, focusing on finding more useful textual elements to improve crime forecasting using AI techniques [16, 17, 18]. Due to their reliance on manually created shallow text characteristics, high salaries, and inadequate domain adaptability, these conventional approaches are challenging to adjust to different situations.

In 2021, [19] conducted an extensive research assessment of the difficulties encountered by the verdict estimation structure, which uses a DL algorithm to assist attorneys, judicial officials, and non-specialists in forecasting the likelihood of revenue or loss, the duration of penalties, and articles from relevant law for novel instances. The investigators provide a detailed description of the transformer system features, the theoretical works on textual categorization techniques, and the scientific research on LCP techniques.

In [20], studies were conducted to determine if or not a case was deemed to be a breach of an individual's rights by analyzing rulings from the European Court of Human Appeals. [21] classified different file kinds in legal cases of the Trial of Appeals of Minas Gerais using Glove vector words created for the Portuguese languages and Convolutional Neural Network (CNN).

III. Materials and Methods

A. Database

The Indian Legal AI Competition database CAIL2020, the inaugural extensive Indian database for LCP for decision support, is used for enhancing LCP. The statistics are derived from actual legal cases that the Supreme Court has released. The case's facts and the resulting result are used to evaluate every specimen. Since numerous real-world cases contain several defendants, legal prediction will become much more challenging. As a result, the collection of instances with just one defendant is kept in this work. The data show that there is a significant imbalance in the allocation of legal groups in CAIL2020. The majority is made up of basic offenses like robbery and deliberate harm. While the statistics of the 10 regulations with the shortest rate just make up 0.13% of the overall statistics, the data of the 10 regulations with the largest rate make up 78% of the overall statistics. Predicting lower rates and obfuscation regulations is made extremely difficult by the unequal distribution of data groups in CAIL2020. Both the Smaller_CAIL2020 and Larger_CAIL2020 databases make up the CAIL2020 data collection. There are 197,000 tool readings in Smaller_2020 and 1.6 million in Larger_2020. Table I displays the partitioning of the study's 2 databases. To further confirm the accuracy of the suggested framework, this study also added to the CAIL2021 dataset.

Table I CAIL2020 Database

Databases	Larger	Smaller
Training	1450000	150000
Testing	30000	30000
Validation	20000	17000

B. Feature Extraction

CNN is one of the illustrative techniques of DL technological advances. It was initially developed in the 1980s, and as AI and computation energy increased, the CNN rapidly developed. Today, various CNN networks are used in computer vision, image analysis, and other domains. In a basic CNN, the input, pooling layer, and convolution metric compute layer are often present. In order to extract regional characteristics from textual or image data, CNN's central computation is the convolution matrix.

To extract textual semantic characteristics, scientists primarily build emotional dictionaries and develop features, although selecting features and layout requires a significant amount of labor. CNN is a 1-D and is capable of obtaining the text's regional primary semantics in DL. To extract characteristics from the statistics in this research, the CNN system was utilized in this investigation.

C. Dimension Reduction

Variable screening may be used to get significant independent factors during the analysis of higher-dimensional statistics in the framework. This can simplify the framework and guarantee that the filtered independent factors have a robust interpretation of the dependent factors. Researchers from a variety of areas now frequently use this strategy because it gives the theoretical framework the exceptional functionality of a lower-dimensional optimal system. A summary of the factor-selecting technique's foundation is provided. To determine if every independent factor can enter the framework by developing suitable regulations, the independent factor degree is first examined. This type of reasoning, nevertheless, is prone to run into the "size disaster" problem the challenge of complex calculation when dealing with huge sizes. The PCA method was developed and is now widely applied in numerous domains. The PCA approach will be used in this framework to lower the characteristics data size.

This paper presents a sequential decision-based LCP paradigm depending on supervision analysis that is made up of 3 layers: an ensemble fact descriptive coding level, a supervision analysis level depending on the self-attention procedure, and an output estimation layer.

D. Ensemble Fact Description Coding Level

The fact descriptive coding tier serves as the estimations ensemble layer in the architecture used in this work. This work uses BiLSTM for encoding the representation of facts.

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

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

$$o_t = \sigma(W_o\,[x_o + h_{o-1}] + b_o) \tag{3}$$

$$\bar{C_t} = tanh(W_c\,[x_t + h_{t-1}] + b_c) \tag{4}$$

$$C_t = f_t \times C_{t-1} + i_t \times \overline{C_t} \tag{5}$$

$$h_t = o_t \times tanh\,(C_{t-1}) \tag{6}$$

In the BiLSTM programming procedure, the front LSTM system obtains the from left to right characteristic depiction of the fact description, while the reverse LSTM networking encodes and fuses the meanings of the following characteristics.

$$h = BiLSTM(x, \emptyset) \tag{7}$$

In simpler terms, the Bi-LSTM system encodes the input vector x to produce the highest-level semantics response h. The size of the front and back implicit states are adjusted to d/2, whereas the size of the implicit phase following spliced is fixed to d. The variable associated with the computation procedure is ϕ.

E. Supervision Analysis Tier Depending on Self-Attention Procedure

The supervision analysis tier, depending on the self-attention system, primarily provides supervision analysis to gather efficient previous task-specific dependent data, therefore offering a significant characteristic assurance for following consecutive LCP-dependent data layer extract. In the surveillance tier of the LCP procedure, a BiLSTM system depending on the self-attention system is used to extract the fact descriptive characteristics for every sub-data collection by supervising sub-data collection labeled statistics.

The characteristic description for every sub-data collection categorization is then calculated using a balanced total depending on the self-attention volume.

To guarantee that every sub-data collection may acquire its relevant characteristic description from the fact descriptive characteristic in simulation, class labeling for every sub-data collection is included for supervision analysis throughout this stage of training.

Supervision analysis is used to guarantee that it can acquire efficient sub-data set-based characteristic depiction, which guarantees the accuracy of consecutive LCP prediction dependent on the dependent characteristics.

F. Output Estimation Layer

According to the feature-based hdj of integrating reliant activity data, it is linearly modified, and the softmax functioning is used to achieve the final estimation of the task incorporating reliant data while minimizing the cross-entropy.

G. Training

To calculate the ultimate loss value across every estimation activity, this study evaluated the total of the procedure-supervised cross-entropy decline for every sub-data collection and the overall cross-entropy of the activity with reliant data.

In action, this study maintains consistency between the value of the loss functionality of supervision analysis and the value of the subset of data comprising reliant data. This study merely require to determine the measured ratio between the two pieces. In this article, GA is utilized to optimize parameters.

IV. Result and Discussion

For comparing the empirical outcomes, the below 3 algorithms were chosen as benchmark algorithms.
(1) CNN. The literature uses a CNN algorithm with several filter sizes for categorizing and classifying the fact representation.
2) HSLTM. Compared to the hierarchy NN architecture used in the emotional categorization challenge, this study uses LSTM to encode line characteristics and additional LSTM to encode text attributes outlined in facts.
(3) TOPCASE is a legal decision estimation framework considering structural dependence among distinct sub-data collections. LSTM programming depicted fact specification characteristics.

The preliminary empirical outcomes for the 4 sub-data collections are reported in Tables II-V and Fig.1-4. The efficiency of the algorithm given in this work is demonstrated by the empirical findings, which indicate that the SA-LCP paradigm used in this study functions better than the standard technique in 4 classification criteria of 4 databases.

Table II Prediction Outcomes of CAIL2021-Larger

Techniques	Databases	CAIL2021-Larger			
		Accuracy	Precision	Recall	F-value
Standard	CNN	0.82	0.85	0.88	0.82
	HLSTM	0.84	0.87	0.81	0.84
	TOP-CASE	0.84	0.88	0.80	0.84
Proposed	LCP	0.85	0.883	0.81	0.84
	SA-LCP	0.86	0.88	0.84	0.57
	PCA-SA-LCP	0.88	0.88	0.87	0.87

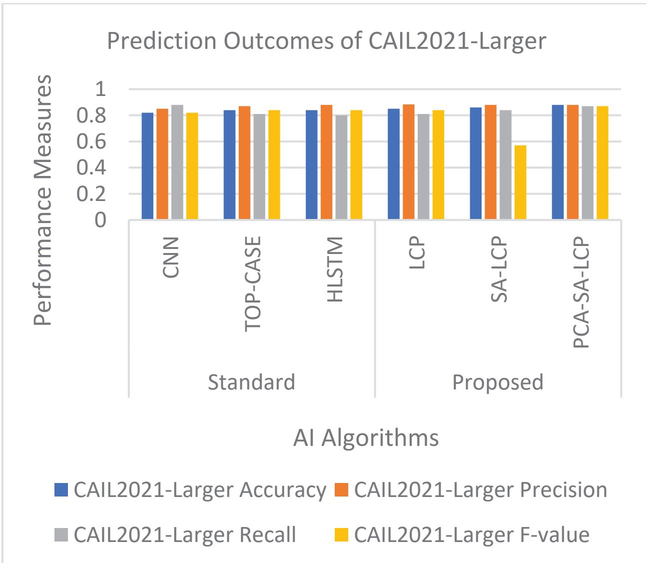

Fig.1. Prediction Outcomes of CAIL2021-Larger

Table III Prediction Outcomes of CAIL2021-Smaller

Techniques	Databases	CAIL2021-Smaller

		Accuracy	Precision	Recall	F-value
Standard	CNN	0.85	0.88	0.89	0.82
	HLSTM	0.88	0.90	0.82	0.85
	TOP-CASE	0.88	0.91	0.81	0.86
Proposed	LCP	0.88	0.91	0.82	0.85
	SA-LCP	0.89	0.90	0.85	0.88
	PCA-SA-LCP	0.91	0.90	0.88	0.88

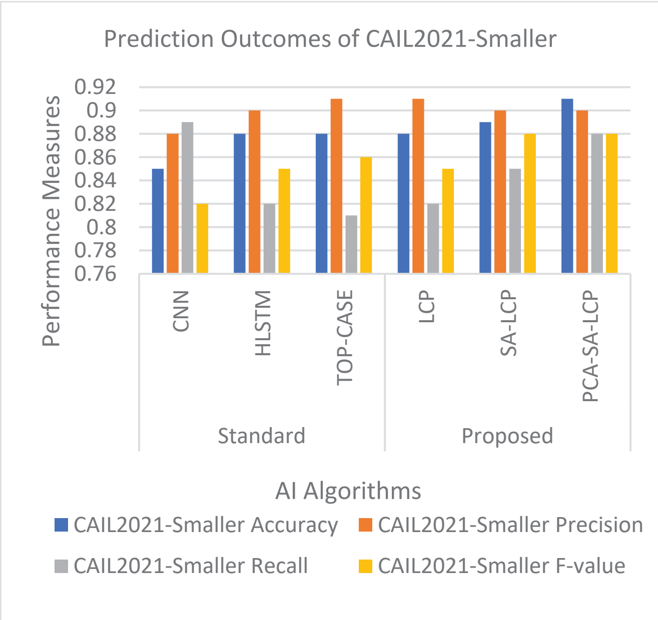

Fig.2. Prediction Outcomes of CAIL2021-Smaller

TABLE IV PREDICTION OUTCOMES OF CAIL2020-LARGER

Techniques	Databases	CAIL2020-Larger			
		Accuracy	Precision	Recall	F-value
Standard	CNN	0.84	0.87	0.89	0.84
	HLSTM	0.87	0.90	0.82	0.86
	TOP-CASE	0.86	0.90	0.81	0.861
Proposed	LCP	0.86	0.91	0.83	0.86
	SA-LCP	0.87	0.90	0.85	0.88
	PCA-SA-LCP	0.89	0.90	0.88	0.89

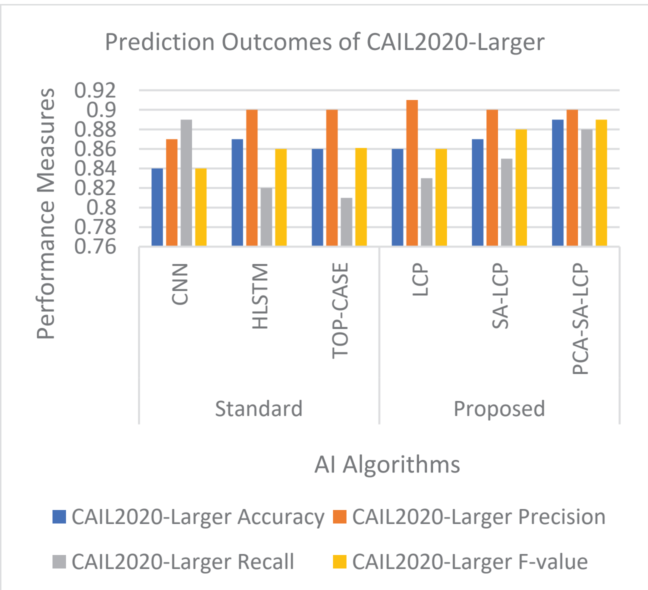

Fig.3. Prediction Outcomes of CAIL2020-Larger

TABLE V PREDICTION OUTCOMES OF CAIL2020-SMALLER

Techniques	Databases	CAIL2020-Smaller			
		Accuracy	Precision	Recall	F-value
Standard	CNN	0.87	0.89	0.93	0.85
	HLSTM	0.90	0.91	0.85	0.87
	TOP-CASE	0.89	0.91	0.84	0.87
Proposed	LCP	0.90	0.92	0.86	0.87
	SA-LCP	0.90	0.91	0.88	0.90
	PCA-SA-LCP	0.92	0.91	0.91	0.89

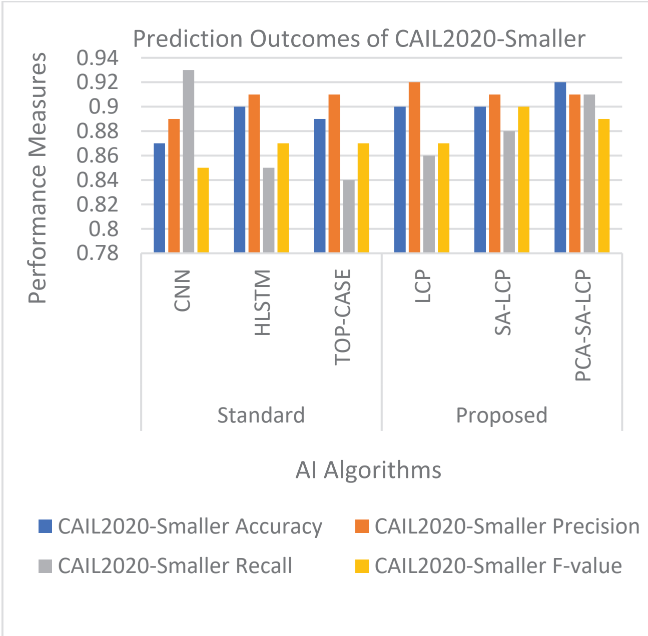

Fig.4. Prediction Outcomes of CAIL2020-Smaller

Tables II-V show that (1) all of the algorithms operate more effectively in the smaller database in comparison to the bigger database across all 4 databases, indicating that the dimension of the database will have a significant effect on the approach's forecasting accuracy; and (2) the suggested framework surpasses the standard approach in all 4 databases compared to the standard and suggested paradigm. This demonstrates that the suggested system outperforms some standard algorithms for LCP; (3) the SA-LCP algorithm suggested performs more effectively rather than the single LCP in 4 distinct variables across 4 databases; and (4) the SA-LCP paradigm depending on PCA that this study suggested outperformed the SA-LCP system without characteristic data preprocessing in the 4 databases.

More significantly, this study discovered that the PCA-SA-LCP paradigm this study suggested produced the best-predicted outcomes out of every algorithm, demonstrating the effectiveness of our suggested approach.

V. CONCLUSION

It has typically been demonstrated that using AI as an instrument to forecast particular traits is a practical and beneficial response in the technical and scientific domains. The investigation of predicting legal cases is a valuable subtask of smart legal practice. Legal professionals can use it to handle cases more effectively and efficiently, but it may additionally assist the general public in comprehending cases and developing psychological projections about how they will turn out. Depending on the study of LCP, this work examines the issues from legal practice and transforms them into textual analysis activities. In this study, various NN architectures are built using the DL algorithm, and training and evaluation are carried out using actual legal databases. In this study, LCP employed the PCA approach to minimize the size of data characteristics and the CNN method for gathering text information. The supervision analysis-based forecasting framework is then presented initially. Supervision analysis is used to assure the accuracy of the generated dependent data when simulating the dependent connection between consecutive sub-data collections. GA is used for optimizing the variables and enhancing the efficiency of the final forecast. The findings of this investigation were extremely positive since the forecast accuracy was extremely elevated, demonstrating that AI techniques could potentially used to estimate LCP. When combined with these methods, the data preprocessing phase proved crucial in raising the techniques' prediction standard and greatly enhancing the outcomes. The suggested methodology produced the best outcomes on four distinct legal publicly accessible databases when contrasted to the standard approach.

REFERENCES

[1] E. Dickhaut, A. Janson, M. Söllner, and J. M. Leimeister, "Lawfulness by design–development and evaluation of lawful design patterns to consider legal requirements," *Eur. J. Inf. Syst.*, vol. 33, no. 4, pp. 441-468, Jul. 3, 2024.

[2] V. Oliveira, G. Nogueira, T. Faleiros, and R. Marcacini, "Combining prompt-based language models and weak supervision for labeling named entity recognition on legal documents," *Artif. Intell. Law*, vol. 1, pp. 1-21, Feb. 15, 2024.

[3] R. A. Shaikh, T. P. Sahu, and V. Anand, "Predicting outcomes of legal cases based on legal factors using classifiers," *Procedia Comput. Sci.*, vol. 167, pp. 2393-2402, Jan. 1, 2020.

[4] J. Cui, X. Shen, and S. Wen, "A survey on legal judgment prediction: Datasets, metrics, models, and challenges," *IEEE Access*, Sep. 19, 2023.

[5] R. Sil and A. Roy, "A novel approach to argument-based legal prediction model using machine learning," in *Proc. 2020 Int. Conf. Smart Electron. Commun. (ICOSEC)*, Sep. 10, 2020, pp. 487-490.

[6] J. Zahir, "Prediction of a court decision from rabic documents using deep learning," *Expert Syst.*, vol. 40, no. 6, p. e13236, Jul. 2023.

[7] S. S. Pandi, A. M. Farook, and W. Kingston, "Scenario-based deep learning prediction for legal judgment using LSTM and CNN," in *Proc. 2023 3rd Int. Conf. Smart Technol., Commun. Robotics (STCR)*, Dec. 9, 2023, vol. 1, pp. 1-6.

[8] R. Singh, B. V. Kumar, K. K. Rawat, S. V. Akram, S. Kathuria, and V. Pachouri, "Intervention of AI in prediction of judgement of cases," in *Proc. 2023 4th Int. Conf. Smart Electron. Commun. (ICOSEC)*, Sep. 20, 2023, pp. 1268-1272.

[9] S. Li, H. Zhang, L. Ye, S. Su, X. Guo, H. Yu, and B. Fang, "Prison term prediction on criminal case description with deep learning," *Comput. Mater. Continua*, vol. 62, no. 3, pp. 1217-1231, 2020.

[10] H. Chen, L. Wu, J. Chen, W. Lu, and J. Ding, "A comparative study of automated legal text classification using random forests and deep learning," *Inf. Process. Manage.*, vol. 59, no. 2, p. 102798, Mar. 1, 2022.

[11] L. Gan, K. Kuang, Y. Yang, and F. Wu, "Judgment prediction via injecting legal knowledge into neural networks," in *Proc. AAAI Conf. Artif. Intell.*, May 18, 2021, vol. 35, no. 14, pp. 12866-12874.

[12] S. Goyal, A. Shankar, K. C. Das, A. Singh, S. Oli, and Madan Mohan Sati, "Manual and Adaptive tuned PID controllers for industrial application," pp. 1953–1958, May 2024, doi: https://doi.org/10.1109/icacite60783.2024.10616874

[13] H. Zhang, Z. Dou, Y. Zhu, and J. R. Wen, "Contrastive learning for legal judgment prediction," *ACM Trans. Inf. Syst.*, vol. 41, no. 4, pp. 1-25, Apr. 21, 2023.

[14] H. Zhong, Y. Wang, C. Tu, T. Zhang, Z. Liu, and M. Sun, "Iteratively questioning and answering for interpretable legal judgment prediction," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 3, 2020, vol. 34, no. 01, pp. 1250-1257.

[15] L. Yue, Q. Liu, B. Jin, H. Wu, K. Zhang, A. Y. An, M. Cheng, B. Yin, and D. Wu, "Neurjudge: A circumstance-aware neural framework for legal judgment prediction," in *Proc. 44th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, Jul. 11, 2021, pp. 973-982.

[16] K. D. Ashley, "A brief history of the changing roles of case prediction in AI and law," *Law Context: Socio-Legal J.*, vol. 36, pp. 93-104, 2019.

[17] Y. Guo, Y. Li, F. Ge, H. Yu, S. Wang, Z. Miao, "Legal judgment prediction via fine-grained element graphs and external knowledge," in *Proc. 2024 Int. Joint Conf. Neural Networks (IJCNN)*, Jun. 30, 2024, pp. 1-8.

[18] L. Yue, Q. Liu, B. Jin, H. Wu, A. Y. An, "A circumstance-aware neural framework for explainable legal judgment prediction," *IEEE Trans. Knowl. Data Eng.*, Apr. 11, 2024.

[19] S. G. Sukanya and J. Priyadarshini, "Hybrid CNN: An empirical analysis of machine learning models for predicting legal judgments," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 7, pp. 1-7, Jul. 1, 2024.

[20] M. Medvedeva, M. Vols, and M. Willing, "Using machine learning to predict decisions of the European Court of Human Rights," *Artif. Intell. Law*, vol. 28, no. 2, pp. 237-266, Jun. 2020.

[21] J. Lopes and C. D'Britto, "Implementing serverless computing architectures for expandable and cost-effective cloud applications," Int. J. Comput. Trends Technol., vol. 72, no. 8, pp. 171-179, 2024. https://doi.org/10.14445/22312803/IJCTT-V72I8P125.