

## Human Rights Case Analysis Using Ai and Transformer Models

Dipali S. Jadhav<sup>1</sup>, Dr. N. R. Wankhade<sup>2</sup>, Santhosh R. Agrawal<sup>3</sup>, Ashwini Gaikwad<sup>4</sup>

Priyanka U. Mandlik<sup>5</sup>

<sup>1</sup>PG Computer Engineering, Late G. N. Sapkal College of Engineering, Nashik, Maharashtra, India

<sup>2,3</sup>Professor, Computer Engineering, Late G. N. Sapkal College of Engineering, Nashik, Maharashtra, India

<sup>4,5</sup>PG Computer Engineering, Late G. N. Sapkal College of Engineering, Nashik, Maharashtra, India

**Emails:** [deepalitarle10@gmail.com](mailto:deepalitarle10@gmail.com)<sup>1</sup>, [nileshw\\_2000@yahoo.com](mailto:nileshw_2000@yahoo.com)<sup>2</sup>, [s.ragrawal1976@gmail.com](mailto:s.ragrawal1976@gmail.com)<sup>3</sup>, [gaikwadashwini481@gmail.com](mailto:gaikwadashwini481@gmail.com)<sup>4</sup>, [priyankamandlik759@gmail.com](mailto:priyankamandlik759@gmail.com)<sup>5</sup>

### Abstract

The increasing volume of human rights cases calls for intelligent systems capable of analyzing legal documents efficiently and accurately. Manual review is time-consuming, labor-intensive, and susceptible to human error, making it challenging to detect patterns or similarities across extensive datasets. This project leverages AI techniques—specifically Transformer-based Legal-BERT and Doc2Vec combined with Support Vector Machine (SVM)—to automate the analysis of human rights case documents. These methods enable semantic understanding, feature extraction, and classification of legal texts, assisting stakeholders in identifying violations and relevant precedents more effectively. Legal-BERT, a BERT variant pretrained on legal corpora, captures domain-specific language, legal terminology, and contextual nuances, generating high-quality embeddings for case documents. This model supports tasks such as case classification, violation detection, and semantic similarity analysis, allowing the system to identify related cases and recurring patterns in human rights violations. In parallel, Doc2Vec transforms entire documents into vector representations, which are subsequently classified using SVM to predict categories or potential case outcomes. By combining deep contextual embeddings with robust traditional machine learning, this hybrid approach enhances overall performance. The proposed system not only accelerates human rights case analysis but also ensures scalability and continuous learning as new cases are incorporated. By integrating Legal-BERT with Doc2Vec + SVM, the project establishes a comprehensive AI-powered framework for legal document processing, pattern recognition, and decision support, thereby improving the efficiency of legal professionals, NGOs, and policymakers in protecting human rights.

**Keywords:** Human Rights, Legal-BERT, Doc2Vec, Support Vector Machine (SVM), NLP, Case Analysis, Legal Document Processing, Semantic Similarity, Classification

### 1. Introduction

The protection of human rights remains one of the most critical responsibilities of modern legal systems. With the increasing number of human rights cases worldwide, legal institutions, NGOs, and policymakers face the daunting challenge of efficiently reviewing and analyzing extensive volumes of case documents. Traditional manual review processes are labor-intensive, time-consuming, and prone to errors, which can delay justice and reduce the effectiveness of interventions. As such, there is a growing need for intelligent systems that can assist in the accurate and timely analysis of legal texts. Advancements in artificial intelligence (AI) and natural language processing

(NLP) offer promising solutions to these challenges. AI-powered systems can analyze complex textual data, extract meaningful patterns, and classify documents with high accuracy. Specifically, transformer-based models like BERT have demonstrated remarkable capabilities in understanding the context and semantics of language. By adapting these models to the legal domain, researchers and practitioners can process large volumes of legal documents while capturing domain-specific terminology and nuances. Legal-BERT, a variant of BERT pretrained on extensive legal corpora, has emerged as a powerful tool for legal text analysis. It enables semantic understanding, case

classification, and similarity analysis by generating high-quality embeddings for legal documents. These embeddings capture subtle linguistic and contextual cues present in case texts, allowing the system to identify relevant precedents, recurring violations, and patterns that may not be immediately apparent to human reviewers. In addition to transformer-based models, traditional machine learning approaches remain valuable for structured classification tasks. Doc2Vec, for instance, converts entire legal documents into fixed-length vector representations, which can then be classified using algorithms like Support Vector Machine (SVM). This combination allows the system to leverage both deep contextual understanding from Legal-BERT and the predictive robustness of SVM for document classification and outcome prediction. By integrating these complementary AI techniques, the proposed framework offers a scalable and automated solution for human rights case analysis. It reduces the manual workload on legal professionals, accelerates the identification of violations, and provides actionable insights for decision-making. Moreover, the system can continuously learn from new case data, ensuring that it remains up-to-date and relevant in a rapidly evolving legal landscape. AI-driven approaches such as Legal-BERT combined with Doc2Vec and SVM represent a significant advancement in the field of legal analytics. This project demonstrates the potential of these technologies to transform human rights case processing, improve the efficiency of legal research, and support NGOs and policymakers in safeguarding human rights more effectively.

### 1.1. Abbreviations And Acronyms

AI (Artificial Intelligence) refers to systems capable of performing tasks that typically require human intelligence; NLP (Natural Language Processing) enables machines to understand and process human language; BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based language model, while Legal-BERT is its legal-domain adaptation trained on legal corpora; Doc2Vec converts documents into vector representations; SVM (Support Vector Machine) is a supervised machine learning algorithm for classification; ML (Machine Learning) and DL (Deep Learning) represent data-driven learning approaches;

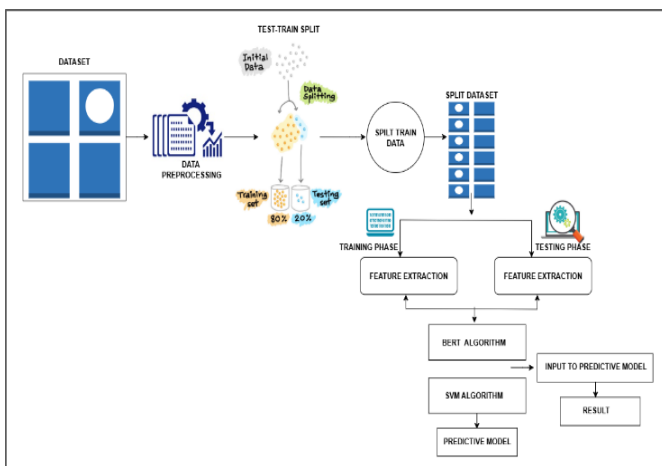
TF-IDF (Term Frequency–Inverse Document Frequency) is used for text feature extraction; LSTM (Long Short-Term Memory) supports sequential text modeling; KNN (K-Nearest Neighbors), RF (Random Forest), CPU (Central Processing Unit), GPU (Graphics Processing Unit), API (Application Programming Interface), IT (Information Technology), HR (Human Rights), NGO (Non-Governmental Organization), and ECHR (European Court of Human Rights) are commonly used terms in the proposed AI-based human rights legal analysis framework.

### 2. Synthesis Of Samples

To develop an effective AI-based framework for human rights case analysis, several sample documents from diverse sources were collected. These included court judgments, legal petitions, NGO reports, and case summaries related to human rights violations. The samples were carefully curated to represent a wide range of cases, including civil rights infringements, labor rights issues, refugee cases, and freedom of expression violations. Each sample contained unstructured textual data, such as narratives, facts of the case, judgments, and legal references, reflecting[1] the diversity and complexity of real-world legal documents. The collected samples underwent preprocessing to standardize the text for AI analysis. This involved tokenization, removal of stop words, normalization of terms, and handling of legal-specific language, such as citations, statutes, and legal jargon. Such preprocessing ensures that AI models like Legal-BERT and Doc2Vec can capture semantic meaning without being distracted by irrelevant or noisy data. Semantic embeddings were generated using Legal-BERT, allowing the system to understand the context, relationships, and patterns within the documents. Doc2Vec was also applied to produce vector representations of entire case files, which were then classified using Support Vector Machine (SVM) to predict case categories, potential violations, or outcomes[2]. The combination of these methods provided a hybrid approach that leverages both deep contextual understanding and robust classification. Analysis of the sample cases revealed several insights: recurring types of human rights violations, common legal arguments, and patterns in judgments. These patterns were not easily discernible

through manual review but were highlighted effectively by the AI models. The system also demonstrated the ability to identify semantic similarity between cases, helping legal professionals and NGOs quickly locate relevant precedents. The synthesized samples validated the effectiveness of the proposed hybrid approach. The integration of transformer-based embeddings (Legal-BERT) and vector-based machine learning (Doc2Vec + SVM) enabled accurate classification, efficient semantic search, and pattern recognition across a diverse set of human rights case documents. This synthesis provides a strong foundation for building a scalable, automated system for legal document analysis. As shown in Figure 1 System Architecture[3 – 10]

### 3. System Architecture

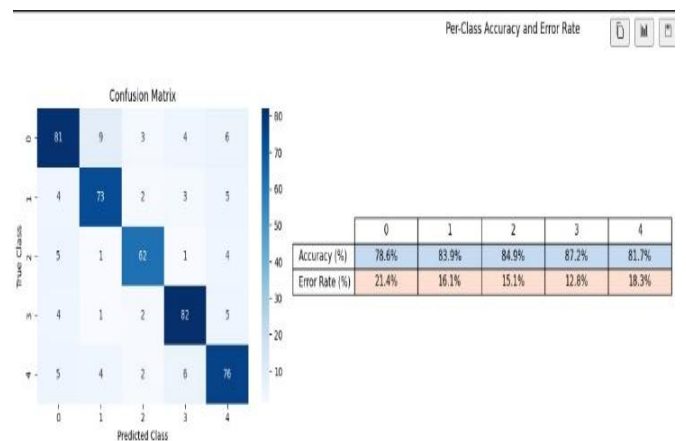


**Figure 1** System Architecture

The diagram illustrates the overall workflow of a machine learning-based predictive system for legal or textual data analysis. It begins with the dataset, where raw data is collected and passed through a preprocessing stage to clean, normalize, and structure the information. The processed data is then divided using a test-train split strategy, typically allocating 80% of the data for training and 20% for testing. This splitting ensures that the model learns patterns from one portion of the data while its performance is evaluated on unseen data. The separated training and

testing datasets are further prepared for the next stages of analysis. In the training and testing phases, feature extraction is applied to identify meaningful attributes from the data. These features are then processed using the BERT algorithm to capture contextual and semantic relationships within the text. After this, the SVM[11] algorithm is used to build a predictive model based on the extracted features. The trained model receives input from the processed testing data and generates predictions, which are presented as the final results. This pipeline demonstrates an integrated approach combining deep learning (BERT) and traditional machine learning (SVM) for accurate and efficient prediction[12].

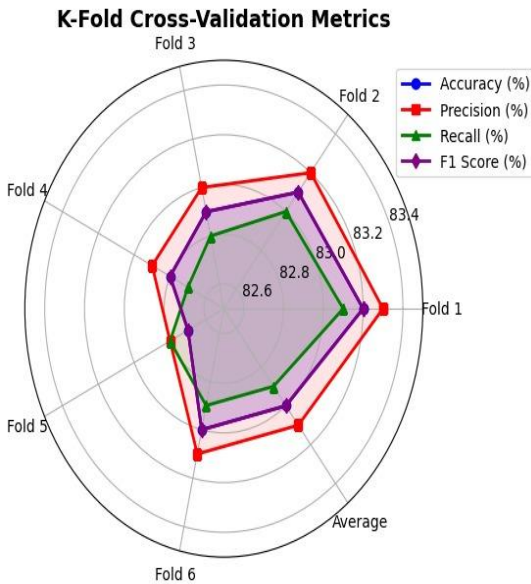
### 4. Experimental Analysis and Results



**Figure 2** Confusion Matrix & Per Class Accuracy With Error Rate Analysis

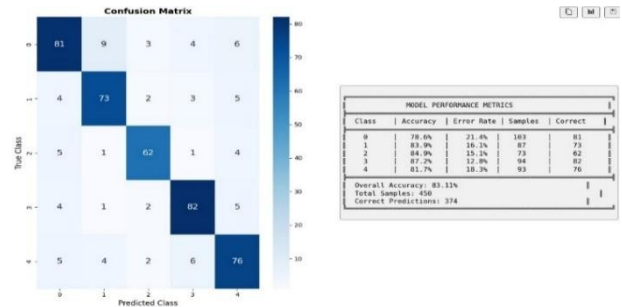
The figure presents the performance evaluation of the classification model using a confusion matrix and per-class accuracy/error rate metrics. The confusion matrix on the left shows how correctly and incorrectly the model predicts each class, with higher values along the diagonal indicating better classification accuracy. The table on the right summarizes the accuracy and corresponding error rate for each class, highlighting how well the model performs individually across categories. Overall, most classes demonstrate high accuracy levels with relatively low error rates, indicating that the model is effective in distinguishing between classes, although some misclassifications still occur and suggest scope for further optimization and training improvements. As shown in Figure 2 Confusion

Matrix & Per Class Accuracy With Error Rate Analysis, Figure 3 K Fold Cross Validation Metrics Showing Accuracy Precision Recall & F1 Score Across Different Folds[13]



**Figure 3 K Fold Cross Validation Metrics Showing Accuracy Precision Recall & F1 Score Across Different Folds**

The radar chart illustrates the performance of the model using K-fold cross-validation by comparing four key metrics—accuracy, precision, recall, and F1-score—across multiple folds and their overall average. Each fold represents a different subset of the dataset used for training and testing, helping evaluate the model’s consistency and generalization capability. The metrics remain relatively stable across folds, with precision and accuracy slightly higher than recall and F1-score in most cases, indicating reliable prediction performance with minor variations. The average values demonstrate balanced model behavior, suggesting that the model performs consistently across different data splits and is not heavily dependent on a specific dataset portion. As shown in Figure 4 Confusion Matrix & Model Performance Metrics For Multiclass Classification



**Figure 4 Confusion Matrix & Model Performance Metrics for Multiclass Classification**

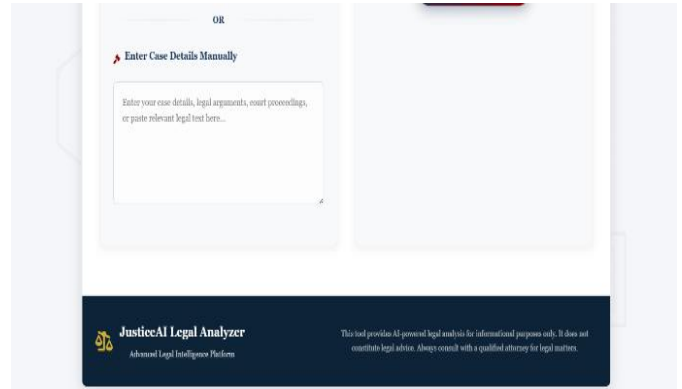
The figure displays the evaluation of a multi-class classification model using a confusion matrix and a performance summary table. The confusion matrix shows the distribution of correct and incorrect predictions for each class, where higher values along the diagonal indicate accurate classifications and off-diagonal values represent misclassifications. The accompanying table provides per class accuracy, error rate, total samples, and correctly predicted instances, offering a detailed view of model effectiveness across different categories. The overall accuracy of 83.11 % demonstrates strong predictive performance, with most classes achieving high accuracy and relatively low error rates, indicating that the model generalizes well while maintaining balanced classification across classes. As shown in Figure 5 Justice Ai Legal Analyzer Smart Fast & Reliable[14 – 20].



**Figure 5 Justice Ai Legal Analyzer Smart Fast & Reliable**

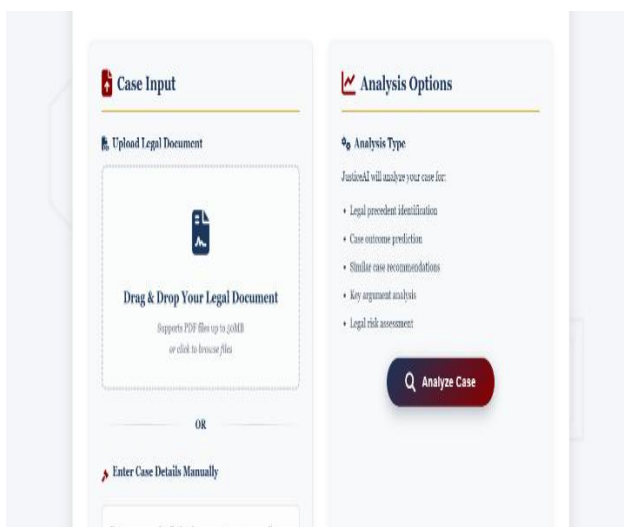
The Justice AI Legal Analyzer is an advanced AI driven platform designed to assist legal professionals by providing accurate, efficient, and data-backed legal analysis. By allowing users to upload legal documents, the system intelligently analyzes cases to

identify relevant legal precedents, predict possible case outcomes, and recommend similar cases across multiple jurisdictions. With high accuracy, round-the-clock availability, and support for diverse legal domains, Justice AI streamlines complex legal research and decision-making, saving time while enhancing precision and confidence in legal strategies. As shown in Figure 6 Upload Analyses & Decide, As shown in Figure 7 Manual Case Entry with Responsible Ai Assistance



**Figure 7 Manual Case Entry with Responsible Ai Assistance**

This section of the JusticeAI Legal Analyzer allows users to manually enter case details, legal arguments, court proceedings, or relevant legal text, providing an alternative to document uploads for greater flexibility. It ensures that even incomplete or draft information can be analyzed effectively using AI-powered legal intelligence. The footer emphasizes responsible usage by clearly stating that the platform offers informational legal analysis and does not replace professional legal advice, reinforcing trust, transparency, and ethical use while supporting informed legal research and decision-making. As shown in Figure 8 Analyze Legal Case



**Figure 6 Upload Analyses & Decide**

This interface of the Justice AI Legal Analyzer enables users to seamlessly input legal cases either by uploading PDF documents or by manually entering case details, ensuring flexibility and ease of use. Once the case data is provided, users can choose from multiple AI-powered analysis options such as legal precedent identification, case outcome prediction, key argument analysis, similar case recommendations, and legal risk assessment. With a clean, user-friendly layout and a single “Analyze Case” action, the platform simplifies complex legal analysis into a fast, accurate, and structured process, empowering legal professionals to make informed decisions efficiently.



**Figure 8 Analyze Legal Case**

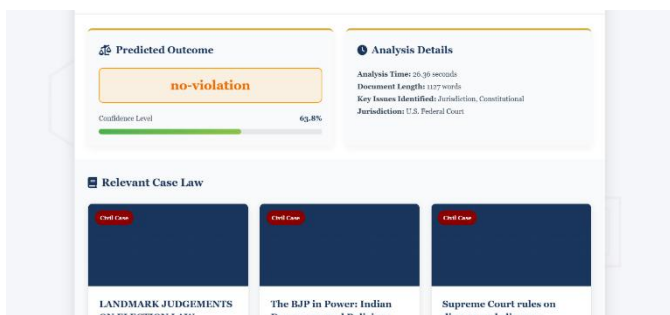
This screen represents the real-time processing stage of the JusticeAI Legal Analyzer, where the system actively analyzes the submitted legal case. During this phase, the AI extracts key legal arguments, interprets case details, and compares them with an extensive case law database to identify relevant precedents and patterns. The progress indicator reassures users that complex legal data is being systematically processed, highlighting the platform’s

ability to deliver accurate, structured, and intelligent legal insights efficiently. As shown in Figure 9 Analysis Results: Ai Generated Case Summary



**Figure 9 Analysis Results: Ai Generated Case Summary**

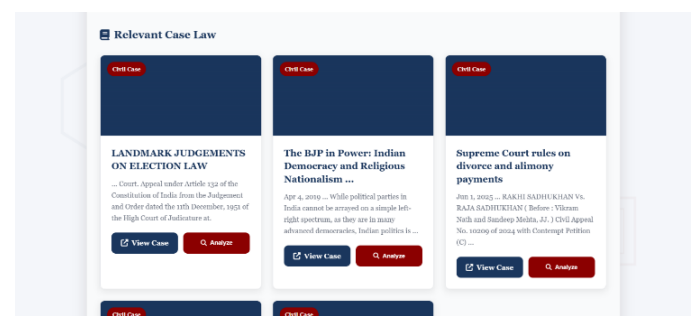
This section displays the Analysis Results generated by the JusticeAI Legal Analyzer, presenting a concise and structured summary of the legal case. The AI automatically extracts essential facts, judicial observations, and legal provisions involved, such as relevant sections of criminal law and constitutional considerations. By condensing complex judgments into an easy-to-understand overview, this feature helps users quickly grasp the core issues, reasoning, and outcome of the case, enabling faster review and more informed legal decision-making. As shown in Figure 10 Predicted Outcome



**Figure 10 Predicted Outcome**

This screen presents the core decision-support insights generated by the JusticeAI Legal Analyzer, including the predicted case outcome with an associated confidence level, detailed analysis metrics, and relevant case law references. The predicted outcome section summarizes the AI's assessment based on legal patterns and precedents,

while the confidence bar indicates the reliability of the prediction. The analysis details provide transparency by highlighting processing time, document length, identified legal issues, and jurisdiction. Additionally, the relevant case law panel suggests comparable civil cases and landmark judgments, enabling users to contextualize the prediction within established judicial precedents and strengthen legal understanding and strategy. As shown in Figure 10 Relevant Case Law & Precedent Recommendations.



**Figure 10 Relevant Case Law & Precedent Recommendations**

This section displays a curated list of relevant case law identified by the JusticeAI Legal Analyzer based on the current case context. Each card highlights a civil case with a brief summary, date, and legal relevance, helping users quickly understand comparable judgments and legal principles. The “View Case” and “Analyze” actions allow deeper exploration of individual rulings, enabling legal professionals to study precedents, draw parallels, and strengthen arguments through informed, data-driven legal research.

**Conclusion**

The rapid growth in human rights case documentation has highlighted the need for intelligent, automated systems capable of handling large-scale legal text analysis. This project demonstrates how advanced AI techniques, particularly Legal-BERT and Doc2Vec combined with Support Vector Machine (SVM), can significantly enhance the efficiency and accuracy of human rights case analysis. By enabling semantic understanding, reliable feature extraction, and effective classification, the proposed framework

reduces dependence on manual review while improving the identification of violations, patterns, and relevant legal precedents. The integrated AI-based approach offers a scalable and adaptable solution for legal document processing and decision support. As new cases are continuously added, the system can learn and evolve, ensuring long-term relevance and performance. By supporting legal professionals, NGOs, and policymakers with actionable insights and data-driven analysis, this framework contributes meaningfully to the protection and promotion of human rights in an increasingly complex legal landscape.

### References

- [1]. Al-Shareef, Y., "CHRExpert: An AI-Driven Court of Human Rights Expert Assistant for Legal Practitioners Utilizing Transformer Models," *IEEE Access*, vol. 13, 2025.
- [2]. N. Faruqui, N. V. D. S. S. V. P. Raju, S. Sivakumar, N. Patel, S. V. Bhaskaran, S. Khanam, and T. Bhuiyan, "Gen-optimizer: A generative AI framework for strategic business cost optimization," *Computers*, vol. 14, no. 2, p. 59, Feb. 2025.
- [3]. K. Sathupadi, S. Achar, S. V. Bhaskaran, N. Faruqui, and J. Uddin, "BankNet: Real-time big data analytics for secure internet banking," *Big Data Cognit. Comput.*, vol. 9, no. 2, p. 24, Jan. 2025.
- [4]. K. M. Richmond, S. M. Muddamsetty, T. Gammeltoft-Hansen, H. P. Olsen and T. B. Moeslund, "Explainable AI and law: An evidential survey," *Digit. Soc.*, vol. 3, no. 1, p. 1, May 2024.
- [5]. M. Saul, "Local authorities at the European court of human rights," *Int. J. Hum. Rights*, vol. 28, no. 7, pp. 1060–1085, Aug. 2024.
- [6]. N. Giri and J. Cheng, "De novo atomic protein structure modeling for cryoEM density maps using 3D transformer and HMM," *Nature Commun.*, vol. 15, no. 1, p. 5511, Jun. 2024.
- [7]. Z. Qazi, W. Shiao, and E. E. Papalexakis, "GPT-generated text detection: Benchmark dataset and tensor-based detection method," in *Proc. Companion ACM Web Conf.*, May 2024, pp. 842–846.
- [8]. Witt, A. Huggins, G. Governatori, and J. Buckley, "Encoding legislation: A methodology for enhancing technical validation, legal alignment and interdisciplinarity," *Artif. Intell. Law*, vol. 32, no. 2, pp. 293–324, Jun. 2024.
- [9]. D. Chandler and B. Hashimoto, "Here-, there-, and every where-: Exploring the role of pronominal adverbs in legal language," *Appl. Corpus Linguistics*, vol. 4, no. 1, Apr. 2024, Art. no. 100087.
- [10]. Á. Pérez Ragoné, "An approach and general overview to framing the structure of the court system and case management," in *Proc. Int. Assoc. Procedural Law Conf.*, Tianjin, China. Springer, Nov. 2023, pp. 143–189.
- [11]. H. Mentzingen, N. António, and F. Bacao, "Automation of legal precedents retrieval: Findings from a literature review," *Int. J. Intell. Syst.*, vol. 2023, no. 1, Jan. 2023, Art. no. 6660983.
- [12]. V. V. Rossolov, "Compensation for non-pecuniary damage in labor law is an important guarantee of the protection of workers' rights," *Uzhhorod Nat. Univ. Herald. Series: Law*, vol. 2, no. 83, pp. 29–33, Aug. 2024.
- [13]. M. Klymenko, S. M. Doesburg, G. Medvedev, P. Xi, U. Ribary, and V. A. Vakorin, "Byte-pair encoding for classifying routine clinical electroencephalograms in adults over the lifespan," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 4, pp. 1881–1890, Apr. 2023.
- [14]. N. Faruqui, P. Thatoi, R. Choudhary, I. Roncevic, H. Alqahtani, I. H. Sarker, and S. Khanam, "AI-analyst: An AI-assisted SDLC analysis framework for business cost optimization," *IEEE Access*, vol. 12, pp. 195188–195203, 2024.
- [15]. T. Zhao and N. Faruqui, "EconoFormer: A novel macroeconomic policy analysis and implementation planner using generative transformer model," *IEEE Access*, vol. 12, pp. 184714–184725, 2024.
- [16]. S. Liu, Y. Liu, J. Zheng, Y. Liao, G. Zheng, and Y. Zhang, "DFFNet: A rainfall nowcasting model based on dual-branch

- feature fusion,” *Electronics*, vol. 13, no. 14, p. 2826, Jul. 2024.
- [16]. Suhaeni and H.-S. Yong, “Enhancing imbalanced sentiment analysis:A GPT-3-based sentence-by-sentence generation approach,” *Appl. Sci.*, vol. 14, no. 2, p. 622, Jan. 2024.
- [17]. M. E. Hossain, N. Faruqui, I. Mahmud, T. Jan, M. Whaiduzzaman, and A. Barros, “DPMS: Data-driven promotional management system of universities using deep learning on social media,” *Appl. Sci.*, vol. 13, no. 22,
- [18]. Á. Pérez Ragone, “An approach and general overview to framing the structure of the court system and case management,” in *Proc. Int. Assoc. Procedural Law Conf., Tianjin, China*. Springer, Nov. 2023, pp. 143–189.
- [19]. H. Mentzingen, N. António, and F. Bacao, “Automation of legal precedents retrieval: Findings from a literature review,” *Int. J. Intell. Syst.*, vol. 2023, no. 1, Jan. 2023, Art. no. 6660983.
- [20]. Jain, M. D. Borah, and A. Biswas, “A sentence is known by the company it keeps: Improving legal document summarization using deep clustering,” *Artif. Intell. Law*, vol. 32, no. 1, pp. 165–200, Mar. 2024.