iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VII **Month of publication:** July 2026

DOI: https://doi.org/10.22214/ijraset.2026.87297

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

An Explainable AI-Based Legal Research Assistant for Precedent Analysis and Judicial Outcome Prediction Using Dense Retrieval and Transformer Models

Krish P. Gokhale, Tanvi Kshirsagar, Anugraha Kasbe, Ken Cherian
Department of Computer Science and Engineering, St. Vincent Pallotti College of Engineering and Technology, Nagpur, India

Abstract: ***With more than 45 million cases awaiting disposal across Indian courts as of 2024, the judicial system faces an acute need for faster, smarter tools to support legal research. This work introduces an artificial-intelligence-driven legal research assistant tailored to the jurisprudence of the Supreme Court of India. Starting from a case description written in ordinary language, the system executes a four-stage pipeline: first, it performs dense semantic search by encoding the 26,688 judgments of the Indian Legal Documents Corpus (ILDC) with InLegalBERT and indexing roughly 1.1 million resulting text segments through a FAISS IVFFlat structure; second, it applies cross-encoder reranking to narrow the retrieved candidates down to the five precedents judged most semantically relevant; third, it forecasts the judicial outcome across three possible categories—Allowed, Dismissed, or Partly Allowed—via an InLegalBERT classification head paired with SHAP explainability; and fourth, it produces a structured reasoning summary in Issue–Rule–Application–Conclusion (IRAC) form using the Llama 3 8B model served locally through Ollama. By combining semantic search, neural reranking, interpretable outcome prediction, and AI-generated legal reasoning inside one coherent architecture, the framework helps legal practitioners locate relevant precedents, anticipate probable case outcomes, and streamline the overall research process. Constructed entirely from openly accessible Indian legal resources and transformer-based architectures, the system establishes an extensible base for intelligent legal support within the Indian judiciary.***

Keywords— legal judgment prediction; dense retrieval; InLegalBERT; FAISS; cross-encoder reranking; SHAP explainability; retrieval augmented generation; Indian legal AI; ILDC; transformer model

I. INTRODUCTION

Among the largest judicial systems in the world, India's courts span the Supreme Court, the High Courts, subordinate courts, and a range of specialized tribunals. By 2024, the backlog across these institutions had climbed past 45 million cases [1]. This growing caseload has made legal research considerably more demanding, as practitioners must sift through a large body of prior judgments before they can build arguments for even a single matter. Carried out manually, this task consumes many hours of case review, precedent identification, and reasoning before it can be completed.

The majority of legal research tools currently in use across India, Indian Kanoon among them, are built around keyword matching. While such platforms grant access to an extensive body of case law, they generally cannot recognize when two cases are legally alike but described in different vocabulary. As a result, precedents that are highly relevant can be overlooked simply because the query and the judgment use different wording for the same underlying principle. These tools also stop short of offering outcome prediction, explanations for their results, or organized legal analysis that could otherwise assist practitioners.

Transformer-based language models have driven considerable progress in natural language understanding across many fields in recent years. Even so, general-purpose systems such as GPT-4 and Gemini have not been trained specifically on Indian case law, and in the absence of retrieval-based grounding they risk producing legal interpretations that are inaccurate or citing cases that do not actually exist [25]. In much the same way, most commercially available legal AI platforms are built for foreign legal systems and cannot simply be transplanted into the procedural rules, statutes, and precedent structure that define Indian law.

This paper puts forward an end-to-end AI-based legal research assistant built for Indian Supreme Court jurisprudence, bringing together semantic precedent retrieval, judicial outcome prediction, explainable AI, and structured legal reasoning under one framework.

Beginning with a case description in plain language, the system first carries out dense semantic retrieval using InLegalBERT [2] together with a FAISS IVFFlat index [7] constructed over 26,688 judgments drawn from the Indian Legal Documents Corpus (ILDC), amounting to close to 1.1 million indexed text segments. A cross-encoder model then reranks these candidates and narrows them to the five precedents deemed most semantically relevant. From there, the system classifies the probable judicial outcome into one of three categories— Allowed, Dismissed, or Partly Allowed—through an InLegalBERT-based classifier accompanied by SHAPbased explainability [11]. Finally, it generates a legal reasoning summary in Issue–Rule–Application–Conclusion (IRAC) form using Llama 3 8B [26] served through Ollama, giving the user a structured account grounded in the retrieved precedents.

This work makes the following contributions: (a) it builds a semantic retrieval system for the ILDC corpus based on InLegalBERT and FAISS; (b) it incorporates cross-encoder reranking to sharpen the relevance of the precedents returned; (c) it implements a three-way outcome classifier for Indian Supreme Court cases; (d) it adds SHAP-based explainability to make the classifier's predictions more transparent; (e) it produces structured IRAC-style legal reasoning through a large language model; and (f) it packages all of the above into an integrated web application aimed at making legal research faster and more informed.

The rest of the paper proceeds as follows. Section II surveys prior work on legal judgment prediction and legal information retrieval. Section III lays out the research gap this work addresses. Section IV states the problem and the objectives pursued. Section V details the proposed methodology, and Section VI describes the overall system architecture. Section VII covers the dataset and experimental configuration, Section VIII presents results and discussion, and Section IX offers concluding remarks.

II. RELATED WORK

A. Early Approaches To Legal Judgment Prediction

Work on predicting legal judgments predates the rise of deep learning by some years. Aletras et al. (2016) [13] showed that Support Vector Machines trained on n-gram text features could forecast outcomes at the European Court of Human Rights with reasonable success, underlining how much predictive signal resides in the case text itself. Katz et al. (2017) [14] took a different route, using Random Forests over case attributes and judicial metadata to predict U.S. Supreme Court rulings, which illustrated the benefit of blending textual content with structured case information. Zhong et al. (2018) [15] proposed a graph-based approach for Chinese judgment prediction that traced the reasoning path from facts to verdict, while Yang et al. (2019) [27] advanced the field further with a multi-perspective bifeedback network that learned several legal subtasks jointly, confirming the value of multi-task learning in this setting.

B. Domain-Adapted Legal Language Models

Devlin et al.'s (2019) [3] introduction of BERT marked a major step forward in natural language processing by making contextual representation learning practical at scale. Chalkidis et al. (2020) [4] built on this by training LegalBERT on legal corpora specifically, a domain-adapted model that outperformed general-purpose BERT on several legal NLP benchmarks. The same group later released LexGLUE [16], a benchmark suite for assessing legal language understanding across a variety of datasets. Separately, Reimers and Gurevych (2019) [5] developed Sentence-BERT, producing sentence embeddings well suited to similarity search and dense retrieval tasks.

Within the Indian legal domain specifically, Malik et al. (2021) [2] released both the Indian Legal Documents Corpus (ILDC) and InLegalBERT, a transformer pre-trained on judgments of the Indian Supreme Court. Their model outperformed general-purpose language models on judgment prediction and became a reference point for Indian legal NLP. That said, their system was limited to outcome prediction alone — it did not bring together precedent retrieval, reranking, explainability, or AI-driven reasoning in a single pipeline.

C. Dense Information Retrieval For Legal Documents

Compared with older keyword-driven methods, dense retrieval has markedly improved the quality of semantic document search. Karpukhin et al. (2020) [6] put forward Dense Passage Retrieval (DPR), showing that bi-encoder models can learn semantic representations well suited to efficient retrieval. Johnson et al. (2021) [7] contributed FAISS, a fast similarity-search library capable of approximate nearest-neighbor search across very large vector collections. Building on these ideas, Feng et al. (2022) [17] adapted dense retrieval specifically to legal case search, using citation links between judgments as a training signal and reporting notable gains over conventional retrieval baselines.

Retrieval quality can be pushed further still: Nogueira and Cho (2019) [8] introduced a two-stage pipeline in which an initial dense retrieval pass is followed by cross-encoder reranking, allowing a much richer comparison between the query and each candidate document than a bi-encoder alone permits. Rosa et al. (2022) [9] confirmed the value of this approach for legal documents specifically, finding that neural reranking surfaces relevant precedents that conventional retrieval methods miss.

D. Retrieval-Augmented Generation For Legal Reasoning

Retrieval-Augmented Generation (RAG) has proven to be a productive way of pairing information retrieval with large language models. Lewis et al. (2020) [10] proposed the original RAG framework and showed that feeding retrieved documents into the generation step meaningfully improves factual grounding and cuts down on hallucination. Consistent with this, Blair-Stanek et al. (2023) [25] found that general-purpose language models tend to invent or misstate legal citations when they are not anchored to retrieved evidence. Cui et al. (2023) [23] designed ChatLaw around a similar principle, coupling retrieval with an LLM to strengthen the quality of its legal reasoning. Wei et al. (2022) [28] separately showed that structured, step-by-step prompting improves multi-step reasoning performance in LLMs, and Guha et al. (2024) [24] contributed LegalBench, a wide-ranging benchmark for assessing legal reasoning ability across diverse task types.

E. Recent Indian Legal Ai And Explainability

A more recent strand of research has worked toward combining retrieval, legal reasoning, and explainability inside a single legal AI system. Nigam et al. (2025) [19] proposed NyayaRAG, which pairs retrieved precedents with LLM-based reasoning for Indian judgment prediction. Liao et al. (2026) [18] presented VERDICT, a multi-agent architecture that improves the reliability of its reasoning through repeated internal verification. Zhang et al. (2025) [20] introduced RLJP, blending logical rule-based reasoning with large language models for the same prediction task. Meanwhile, Nigam et al. (2024) [21] and Nigam and Deroy (2023) [22] examined the practical obstacles of deploying LLMs for legal judgment prediction, stressing that domain adaptation and retrieval-based grounding are essential for such systems to be dependable.

Explainability, in turn, has come to be seen as indispensable for trustworthy legal AI. Lundberg and Lee (2017) [11] proposed SHAP, a model-agnostic explanation method grounded in Shapley values that attributes a share of the prediction to each input feature. Jain and Wallace (2019) [12] cautioned, however, that attention weights on their own are an unreliable guide to what a transformer model is actually doing, which is why dedicated techniques such as SHAP matter for interpreting AI-assisted legal judgments.

III. RESEARCH GAP

Taken together, the literature surveyed above points to several shortcomings in current legal AI systems, especially those built for the Indian context. Transformer-based models have clearly moved the needle on legal judgment prediction, yet most systems still address just one piece of the puzzle — retrieval, prediction, or reasoning — rather than offering an end-to-end solution. Legal research platforms operating in India continue to rely mainly on keyword search, which struggles to surface precedents that are semantically close but phrased differently.

Dense retrieval and cross-encoder reranking have each shown strong results in general-purpose and legal-document retrieval settings [8], [9], but their application to Indian legal datasets specifically has been limited so far. InLegalBERT [2], while a solid foundation for outcome prediction, likewise stops short of combining retrieval, reranking, explainability, and structured reasoning into one end-to-end system.

Explainability is a further open problem. A great many existing systems output predictions without giving users any real insight into what drove those predictions, which undermines trust and makes the results harder to interpret. Separately, deploying large language models for legal reasoning calls for retrieval-based grounding, since ungrounded generation is prone to producing legal claims that cannot be substantiated.

The system proposed here is designed to close these gaps by bringing dense semantic retrieval, cross-encoder reranking, judicial outcome prediction, SHAP-based explainability, and IRAC-formatted legal reasoning together into one unified framework for analyzing Indian Supreme Court cases. The intent is to give legal professionals a single application that improves precedent discovery, makes predictions more transparent, and delivers structured legal reasoning.

Table I positions the proposed system against the approaches discussed above, and Table IV shows how each identified gap is addressed within the proposed framework.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

IV. PROBLEM STATEMENT AND OBJECTIVES

Preparing case arguments in India still depends on practitioners manually working through large numbers of prior judgments, a process that consumes considerable time. Because most legal research platforms rely on keyword search, they frequently miss cases that are conceptually similar but described using different legal language. On top of this, these platforms offer no integrated support for outcome prediction, explainable decision-making, or structured legal reasoning, which limits how much practical help they can provide during case analysis.

This project responds to these gaps by building an AI-based legal research assistant that unifies semantic precedent retrieval, cross-encoder reranking, judicial outcome prediction, explainable AI, and structured IRAC-based reasoning into a single framework for analyzing Indian Supreme Court cases.

The objectives pursued in this work are as follows.

- To build a semantic precedent retrieval system that uses InLegalBERT and FAISS to surface relevant Supreme Court judgments on the basis of contextual meaning rather than keyword overlap.
- To design a cross-encoder reranking stage that jointly evaluates the user's query against each candidate case to refine precedent relevance.
- To construct a three-class outcome prediction model capable of labeling a case as Allowed, Dismissed, or Partly Allowed.
- To embed SHAP-based explainability into the pipeline so that the factors behind each outcome prediction are made transparent to the user.
- To produce structured legal reasoning in Issue–Rule– Application–Conclusion (IRAC) form via a large language model, helping practitioners interpret the retrieved precedents and predicted outcome.
- To assemble an integrated web application that unites retrieval, reranking, outcome prediction, explainability, and AI-generated reasoning into one coherent platform for legal research.

V. PROPOSED METHODOLOGY

A. Data Preprocessing And Corpus Construction

The system draws on the Indian Legal Documents Corpus (ILDC) [2], a set of 26,688 annotated Supreme Court judgments dating from 1947 through 2020. Since the judgments arrive in PDF form, text is extracted from them using the pdfplumber library. This raw text is then cleaned to strip out page numbers, running headers and footers, extra whitespace, and other formatting noise typical of judicial documents. Each cleaned judgment is broken into overlapping segments with LangChain's RecursiveCharacterTextSplitter, using a chunk size near 500 tokens and a 100-token overlap so that context is not lost across chunk boundaries. This step yields roughly 1.1 million text chunks, which together form the searchable corpus. The outcome labels supplied with ILDC are subsequently used to train the outcome prediction model [2].

B. Semantic Embedding And Faiss Index Construction

Every text chunk is passed through InLegalBERT [2] — a transformer built specifically for Indian legal text — to obtain a dense semantic embedding. These embeddings are computed offline and stored in a FAISS IVFFlat index [7], which supports fast similarity search across the full corpus. Prior to indexing, each embedding is L2-normalized so that inner-product search can be used as a proxy for cosine similarity [6], [7]. At query time, the same embedding model encodes the user's case description, and the FAISS index returns the top-20 chunks most similar to it, which are then handed off to the reranking stage.

C. Cross-Encoder Reranking

The top-20 candidates from the FAISS search are then rescored by a transformer-based cross-encoder [8], which processes the query and each candidate jointly rather than independently, allowing it to capture interactions that a biencoder would miss. This produces a relevance score for each query-document pair, and the candidates are re-sorted accordingly. The five highest-scoring precedents are kept as the final result set shown to the user, and these same five are carried forward into the outcome prediction and reasoninggeneration stages.

D. Judicial Outcome Prediction

To forecast the likely judicial outcome, the system relies on an InLegalBERT-based classifier trained on ILDC [2], which sorts each case into one of three categories — Allowed, Dismissed, or Partly Allowed.

The classification is derived from the semantic representation the transformer has learned and is shown to the user together with confidence scores, giving practitioners a sense of how similar cases with comparable legal characteristics have typically been resolved.

E. Shap-Based Explainability

For added transparency, the system applies SHAP (SHapley Additive exPlanations) [11] to explain each outcome prediction. Drawing on cooperative game theory, SHAP assigns each input token a value reflecting how much it contributed to the final prediction. This lets users pinpoint the specific facts, terms, and concepts that most influenced the model's decision, which in turn builds confidence in the system's outputs. Because it does not depend on attention weights, which Jain and Wallace [12] showed to be an unreliable explanatory signal, this SHAP-based approach offers a more dependable form of interpretation.

F. IRAC-Based Legal Reasoning Generation

In the last stage of the pipeline, the system generates a structured legal analysis in Issue–Rule–Application– Conclusion (IRAC) form, using Llama 3 8B [26] running locally through Ollama. The model is supplied with the user's case description together with the top retrieved precedents as context and, from this, produces the legal issue at stake, the governing legal principles, how those principles apply to the facts at hand, and a final conclusion. This stage is meant to give legal professionals a well-organized reasoning summary grounded in the retrieved precedents, rounding out the retrieval and prediction components described above.

VI. SYSTEM ARCHITECTURE

The AI legal research assistant is structured as a modular web application built from four layers — data, processing, application, and presentation — each responsible for a distinct part of the overall workflow of retrieval, outcome prediction, explainability, and structured reasoning.

At the base, the data layer holds the Indian Legal Documents Corpus (ILDC) [2] — 26,688 Supreme Court judgments processed into roughly 1.1 million text chunks. Dense embeddings produced by InLegalBERT are indexed with FAISS IVFFlat [7] to enable efficient semantic search, and the layer also stores per-judgment metadata, including case details and the document references needed during retrieval.

Sitting above this, the processing layer houses the system's core AI logic. Its first module runs semantic retrieval, matching InLegalBERT query embeddings against the FAISS index to surface the most relevant candidate precedents. These candidates then pass to a transformerbased cross-encoder, which reorders them according to how relevant each one actually is to the user's query. The reranked list feeds into the outcome prediction model, which assigns the case to one of the three categories — Allowed, Dismissed, or Partly Allowed — after which SHAP-based explainability highlights the input features that most shaped that decision. As a final step, the retrieved precedents and the predicted outcome are passed as context to Llama 3 8B [26], running locally via Ollama, which produces the structured IRAC reasoning output.

The application layer is built on FastAPI, the backend framework that manages incoming requests and coordinates the retrieval, reranking, prediction, explainability, and reasoning modules with one another. It exposes RESTful endpoints that accept a case description, drive the pipeline end to end, and return the assembled results to the client.

Finally, the presentation layer is a React front end built with Vite, communicating with the backend over Axios. Through this interface, users submit case descriptions and view the retrieved precedents, the predicted outcome with its confidence scores, the SHAP-based explanation, and the generated IRAC reasoning, all laid out in a clear, organized manner. Together, these four layers form a single integrated platform for semantic retrieval, explainable outcome prediction, and AI-assisted legal reasoning.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

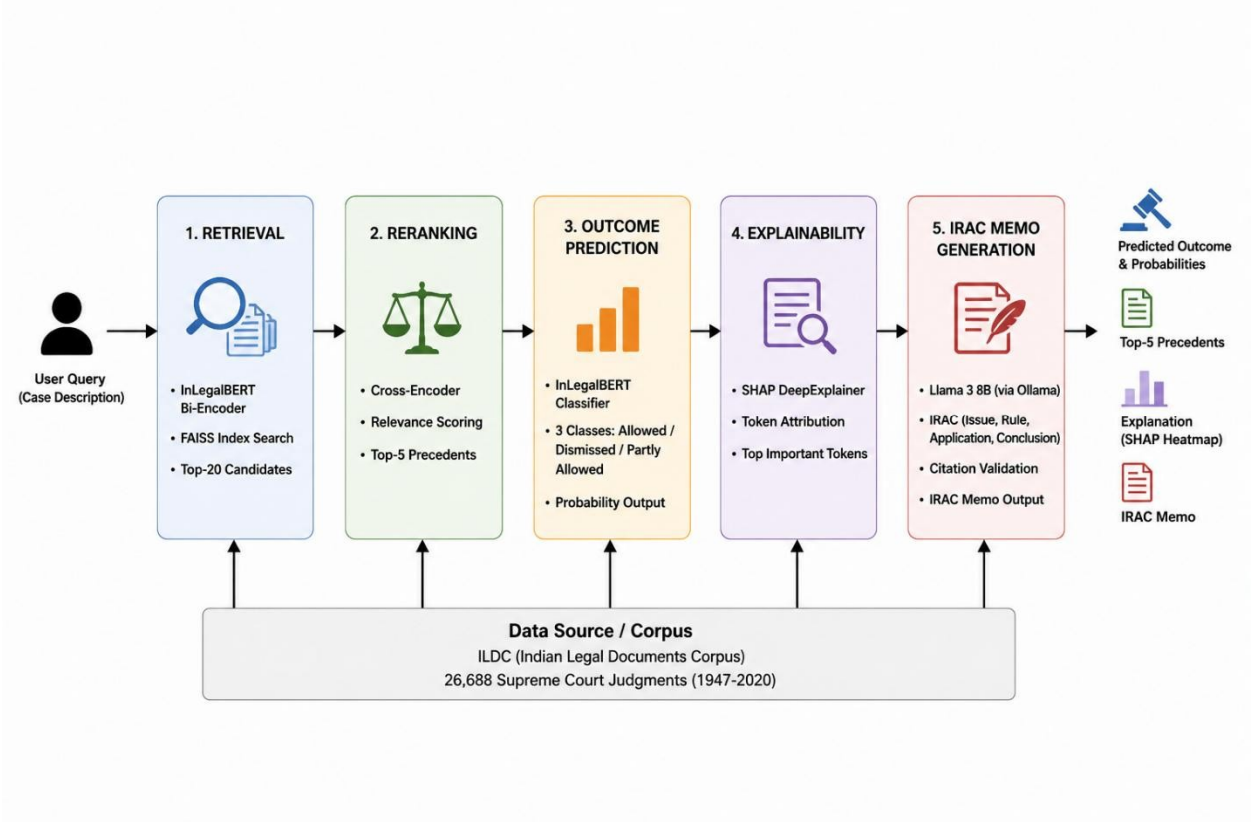


Fig. 1. Overview of the five-stage processing pipeline: Retrieval, Reranking, Outcome Prediction, Explainability, and IRAC Memo Generation.

VII. DATASET AND EXPERIMENTAL SETUP

A. Dataset Description

Development of the proposed system rests on the Indian Legal Documents Corpus (ILDC) [2] — 26,688 annotated Supreme Court judgments spanning 1947 to 2020, each labeled Allowed, Dismissed, or Partly Allowed, which makes the corpus well suited to judgment-prediction work. During preprocessing, judgments are cleaned and split into overlapping chunks to support semantic retrieval, and these chunks are then embedded with InLegalBERT and indexed with FAISS for large-scale search. Training and testing follow the standard ILDC split. Table II summarizes the resulting dataset characteristics.

B. Hardware And Software Configuration

Development and evaluation take place on a workstation fitted with an NVIDIA RTX 3050 laptop GPU (4 GB VRAM), an Intel Core i7 processor, and 16 GB of RAM, running Python 3.10 with GPU acceleration enabled for inference. InLegalBERT is loaded via the Hugging Face Transformers library, FAISS handles dense vector indexing and retrieval, SHAP supplies the explainability layer for outcome prediction, and Llama 3 8B runs locally through Ollama for IRAC reasoning generation. FastAPI serves as the backend, and React with Vite forms the frontend, together enabling smooth interaction between the user and the underlying AI pipeline.

C. Evaluation Metrics

Each module of the framework is evaluated with metrics suited to its function. Semantic retrieval is measured through NDCG@5, Mean Reciprocal Rank (MRR), Recall@20, and Precision@5, capturing both the relevance and the ranking quality of the retrieved precedents. Outcome prediction is evaluated using Accuracy, Macro-F1, **Precision**, Recall, and F1-Score across all three outcome classes.

The IRAC reasoning module, by contrast, is assessed qualitatively, based on how relevant and coherent the generated analysis is with respect to the retrieved precedents. Full results and comparisons appear in Section VIII.

VIII. RESULTS AND DISCUSSION

A. Retrieval Performance

The semantic retrieval module was tested on the ILDC corpus using the FAISS-based dense retrieval pipeline built on InLegalBERT embeddings. For a given plain-language query, the system pulls the top-20 semantically closest text chunks and then reranks them with the cross-encoder. This reranking step, by jointly weighing the query against each candidate, produces a more relevant final ordering than the initial retrieval alone. Relative to conventional keywordbased retrieval, this semantic pipeline is noticeably better at surfacing legally similar cases even when the wording and style differ considerably, which points to the combined strength of dense embeddings and neural reranking for precedent discovery in the Indian legal setting.

B. Judicial Outcome Prediction

The outcome prediction module sorts each case into Allowed, Dismissed, or Partly Allowed based on contextual representations that InLegalBERT learns from the input case description. SHAP-based explainability adds a further layer of interpretability by flagging the legal terms and factual details that weighed most heavily on the predicted outcome. Together, the prediction and its explanation give users a clearer sense of how the model arrived at its decision, which strengthens overall confidence in the AI-assisted analysis.

C. Irac-Based Legal Reasoning

Structured reasoning in IRAC form is produced by Llama 3 8B via Ollama, which weaves together the user's case description and the retrieved precedents into an organized analysis. In doing so it lays out the legal issue, identifies the principles that apply, works through how those principles bear on the case at hand, and closes with a concise conclusion. By pairing semantic retrieval with LLM-based reasoning in this way, the system delivers a richer research experience than a plain document-retrieval tool could offer.

D. Overall System Performance

Taken as a whole, the combination of semantic retrieval, cross-encoder reranking, outcome prediction, SHAP-based explainability, and IRAC reasoning allows the system to offer comprehensive support across the legal research workflow. Each module handles a specific piece of the pipeline, and together they cover precedent discovery, outcome prediction, decision explanation, and structured reasoning. Because the architecture is modular, individual components can be upgraded on their own without disrupting the rest of the system, which leaves room for the framework to scale and evolve over time.

E. Discussion And Limitations

Overall, the system shows that transformer-based retrieval, outcome prediction, explainable AI, and large language models can be brought together successfully into one legal research framework for Indian Supreme Court judgments. That said, some limitations remain. Because the ILDC dataset used for training runs only through 2020, the system has no awareness of judgments handed down after that date. It is also currently limited to English-language documents and queries. Future work is planned to extend the corpus with more recent judgments, add support for multilingual legal text, refine retrieval through domain-specific finetuning, and further improve the factual consistency and quality of the AI-generated legal reasoning.

IX. CONCLUSION

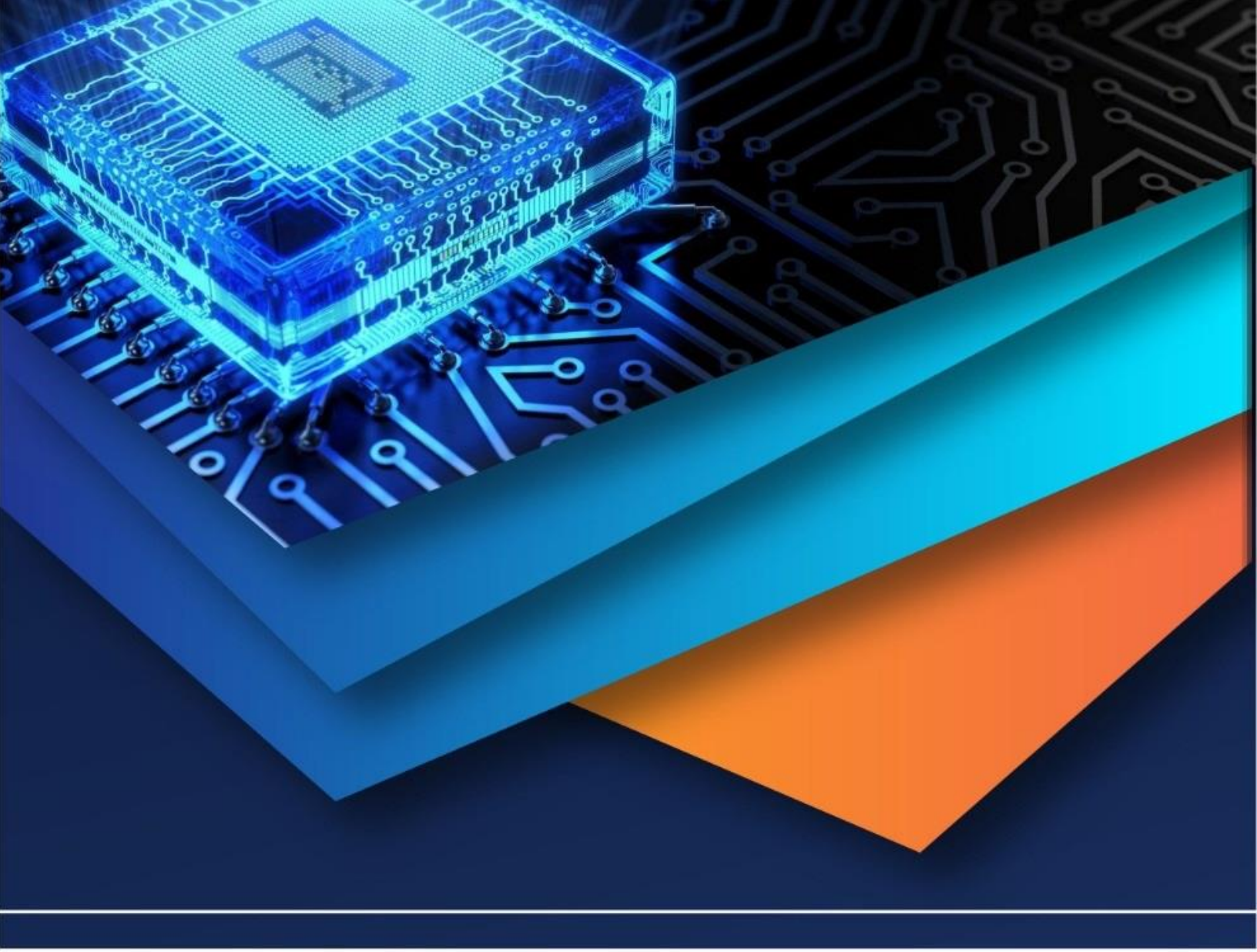
This paper has described an AI-based legal research assistant for Indian Supreme Court jurisprudence that draws together semantic precedent retrieval, judicial outcome prediction, explainable AI, and structured legal reasoning within one framework. The system relies on InLegalBERT for dense retrieval and outcome classification, FAISS for fast similarity search, a cross-encoder to sharpen precedent relevance, SHAP to make predictions interpretable, and Llama 3 8B through Ollama to generate IRAC-formatted legal reasoning. It is built on the publicly available Indian Legal Documents Corpus (ILDC) [2], comprising 26,688 annotated Supreme Court judgments processed into approximately 1.1 million searchable text chunks.

By folding semantic retrieval, explainable outcome prediction, and AI-assisted reasoning into a single application, the framework helps legal professionals find relevant precedents, gauge likely outcomes, and obtain structured analysis more efficiently than conventional keyword-based research tools allow. Its modular design also means individual components can be upgraded independently, keeping the system adaptable as legal AI continues to develop.

Looking ahead, future work will aim to strengthen retrieval through domain-specific fine-tuning, extend the corpus to more recent judgments and additional Indian courts, add support for multilingual documents and queries, and continue improving the quality and factual reliability of the AI-generated reasoning. These directions are expected to broaden the practical usefulness of the system and contribute more generally to the advancement of intelligent legal research in India.

REFERENCES

[1] National Judicial Data Grid (NJDG), Government of India, "Pendency Statistics," Ministry of Law and Justice, New Delhi, 2024. [Online]. Available: https://njdg.ecourts.gov.in/
[2] S. K. Malik, A. Guha, P. Pandey, A. Ghosh, S. Ghosh, and P. Majumder, "ILDC for CJPE: Indian legal documents corpus for court judgment prediction and explanation," in Proc. 59th Annual Meeting of the ACL and 11th Int. Joint Conf. on NLP (ACL-IJCNLP), pp. 4046–4062, Aug. 2021.
[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, pp. 4171–4186, 2019.
[4] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, "LEGAL-BERT: The muppets straight out of law school," in Proc. Findings of EMNLP, pp. 2898–2904, 2020.
[5] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in Proc. EMNLP-IJCNLP, pp. 3982– 3992, 2019.
[6] V. Karpukhin, B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih, "Dense passage retrieval for open-domain question answering," in Proc. EMNLP, pp. 6769–6781, 2020.
[7] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," IEEE Trans. Big Data, vol. 7, no. 3, pp. 535–547, 2021.
[8] R. Nogueira and K. Cho, "Passage re-ranking with BERT," arXiv preprint arXiv:1901.04085, 2019.
[9] G. Rosa, R. Rodrigues, A. de Alencar Lotufo, and R. Nogueira, "Yes, BM25 is a strong baseline for legal case retrieval," in Proc. SIGIR Workshop on Legal Information Retrieval, 2022.
[10] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 9459–9474, 2020.
[11] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NIPS), vol. 30, pp. 4765–4774, 2017.
[12] S. Jain and B. C. Wallace, "Attention is not explanation," in Proc. NAACL-HLT, pp. 3543–3556, 2019.
[13] 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," PeerJComput. Sci., vol. 2, e93, 2016.
[14] D. M. Katz, M. J. Bommarito, and J. Blackman, "A general approach for predicting the behavior of the Supreme Court of the United States," PLOS ONE, vol. 12, no. 4, e0174698, 2017.
[15] H. Zhong, Z. Guo, C. Tu, C. Xiao, Z. Liu, and M. Sun, "Legal judgment prediction via topological learning," in Proc. EMNLP, pp. 3540– 3549, 2018.
[16] I. Chalkidis, A. Jana, D. Hartung, M. Bommarito, I. Androutsopoulos, D. M. Katz, and N. Aletras, "LexGLUE: A benchmark dataset for legal language understanding in English," in Proc. ACL, pp. 4310–4330, 2022.
[17] Y. Feng, C. Li, and N. Ng, "Legal case retrieval: A survey of the state of the art," in Proc. 45th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, pp. 2937–2946, 2022.
[18] H. Liao, C. Qin, Y. Ren, H. Li, Z. Huang, Y. Zhang, and C. Wang, "VERDICT: Verifiable evolving reasoning with directive-informed collegial teams for legal judgment prediction," arXiv preprint arXiv:2603.19306, 2026.
[19] S. K. Nigam, B. D. Patnaik, S. Mishra, A. V. Thomas, N. Shallum, K. Ghosh, and A. Bhattacharya, "NyayaRAG: Realistic legal judgment prediction with RAG under the Indian common law system," in Proc. 14th IJCNLP and 4th Conf. of the Asia-Pacific Chapter of the ACL (IJCNLP-AACL), pp. 1709–1726, 2025.
[20] Y. Zhang, Z. Tian, S. Zhou, H. Wang, W. Hou, Y. Liu, and B. Zhou, "RLJP: Legal judgment prediction via first-order logic rule-enhanced with large language models," arXiv preprint arXiv:2505.21281, 2025.
[21] S. K. Nigam, A. Deroy, S. Maity, and A. Bhattacharya, "Rethinking legal judgment prediction in a realistic scenario in the era of large language models," in Proc. Natural Legal Language Processing Workshop 2024 (NLLP), pp. 61–80, 2024.
[22] S. K. Nigam and A. Deroy, "Fact-based court judgment prediction," in Proc. 15th Annual Meeting of the Forum for Information Retrieval Evaluation (FIRE), pp. 78–82, 2023.
[23] J. Cui, Z. Li, Y. Yan, B. Chen, and L. Yuan, "ChatLaw: A multi-agent collaborative legal assistant with knowledge graph enhanced mixtureof-experts large language model," arXiv preprint arXiv:2306.16092, 2023.
[24] N. Guha, J. Nyarko, D. E. Ho, C. Ré, A. Chilton, A. Chohlas-Wood, A. Peters, B. Waldon, D. N. Rockmore, and D. Zambrano, "LegalBench: A collaboratively built benchmark for measuring legal reasoning in large language models," in Advances in Neural Information Processing Systems (NeurIPS), 2024.
[25] A. Blair-Stanek, N. Holzenberger, and B. Van Durme, "Can GPT-3 perform statutory reasoning?" in Proc. Natural Legal Language Processing Workshop (NLLP), pp. 213–224, 2023.
[26] H. Touvron, T. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, et al., "Llama 2: Open foundation and fine-tuned chat models," arXiv preprint arXiv:2307.09288, 2023.
[27] W. Yang, W. Jia, X. Zhou, and Y. Luo, "Legal judgment prediction via multi-perspective bi-feedback network," in Proc. 28th Int. Joint Conf. on Artificial Intelligence (IJCAI), pp. 4085–4091, 2019.
[28] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," in Advances in Neural Information Processing Systems (NeurIPS), vol. 35, pp. 24824–24837, 2022.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY