International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

Enhancing the Generation of Legal Case Headnotes with Expanded Abstractive Summarization Models

V Anjali
Research Scholar, Computer Science Department, VISTAS
Dr. N. Shyamala Devi
Assistant Professor, Department of Information Technology, VISTAS

Abstract

Legal case headnotes play a vital role in briefing about the cases. Automation of headnotes summarization is very crucial because of the legal language used in the case briefing. There are many transformer-based models like PEGASUS, BART & T5 for summarizing the legal texts. The above models are refined to a specific legal dataset which will investigate a combined approach of strategy which includes extractive summarization models to enhance the quality of summarizing the legal head notes which also holds in line with the limitation of the abstractive summarization approaches. When this method of approach is applied a summary system of legal headnotes will be implemented. By offering strong tools for legal text summarizing, this study is expected to advance the field of legal technology, expedite legal research procedures, and increase accessibility to legal information

.

Keywords: *Legal case summarization, headnotes, abstractive summarization, extractive summarization, transformer models, PEGASUS, BART, T5, legal NLP, hybrid summarization, legal datasets, fine-tuning, legal technology, legal research tools*

I. Introduction :

The data's which is in textual format are widely available with many numbers of volumes which is very crucial to summarize the main content available, especially in legal texts there are many lengthy texts like case laws, legal case articles will be available. To make this concise the headnotes provide the overall facts of the summaries available under each case. Abstractive summarization is an interesting approach when it is compared with the extractive summarization approach. This work includes a legal document specific summarization system of datasets available. With headnotes, this will integrate metrics which is evaluated using ROUGE metrics.

II. Literature Survey

Here, the methodologies like Head note based summarization, Rhetorical Status, Extractive Summarization, and Abstractive Summarization Will be examined

a. Extractive Summarization

This summarization model focuses on examine the important sentences from a document and from that it will create a summary. This model is widely used because of its simplicity, but its lacks in missing of the content

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

- **LexRank :** This ranking is based on the graph based algorithms. The graph shows the similarity of the sentence available in the statements.
- **TextRank :** This is similar to the above Lex Rank approach of making a graph of similar sentences available.
- **Supervised Extractive Models:** This model uses the pretrained transformer on the large datasets to identify the similar or relevant sentences available.

b. Abstractive Summarization

Abstractive summarization captures the essence of the text which are original.

Pointer-Generator Networks : This give the generation of new content from the existing content of combing both extractive and generative capabilities .

- **Pre-Trained Transformers**: PEGASUS (2020), BART (2020), and T5 (2020) are fine tuned for the domain specific datasets.

c. Headnote-Based Summarization

This summarization models only focuses on the summarizing the headnotes by capturing the key facts, issues and judgement available in the legal document.

- **Legal-Specific Fine-Tuning**: mT5 and Legal-BERT models will be fine-tuned for working the legal dataset summarization tasks
- **Hybrid Techniques**: This technique combines the extractive summarization and abstractive summarization techniques to identify only the important critical sections.

d. Semantic Relationship-Based Summarization

The model produces the main context of the summaries

- **Graph-Based Approaches:** Concept of graph summarization mapping of detailing the relationships entities and semantic connectivity will be captured in this approach.
    - **Neural Semantic Summarization:** Pre-trained language models with semantic embeddings (e.g., BERT, RoBERTa) enable capturing deeper contextual and relational information

Analysis of Summarization Techniques in Legal Applications

Aspect	Extractive Summarization	Abstractive Summarization	Headnote-Based Summarization	Semantic Relationship-Based Summarization
Definition	Selects the most relevant sentences	Generates new sentences that	Focuses on summarizing legal	Summarizes based on relationships

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

Aspect	Extractive Summarization	Abstractive Summarization	Headnote-Based Summarization	Semantic Relationship-Based Summarization
	from the text.	convey the essence.	cases (facts, issues, judgments).	between concepts.
Examples	TextRank, LexRank, BERTSUM	PEGASUS, BART, T5	Legal-BERT, mT5, Swiss Legal Dataset Summarizer	Concept Graph Summarization, Semantic Graph Models
Strengths	Retains original phrasing and factual accuracy.	Produces more coherent and human-like summaries.	Tailored to the legal domain, captures key legal aspects.	Captures contextual and relational nuances effectively.
Weaknesses	May lack coherence and miss broader context.	Risk of factual inaccuracies (hallucination).	Requires domain-specific training datasets.	Complex to implement and computationally intensive.
Legal Application	Case search result previews, document indexing.	Headnote generation, summarizing judgments.	Automated headnotes for legal documents.	Argument mining, legal principle summarization.
State-of-the-Art Models	BERTSUM, SummaRuNNer	PEGASUS, GPT-3, T5	mT5, Legal-BERT, Hybrid Legal Models	Semantic-Aware Transformers
Accuracy Scores	Extractive models maintain ~0.85 factual accuracy in legal summaries but lack contextual depth.	Abstractive models achieve 0.75–0.80 for fluency and coherence, with domain adaptation improving factual scores.	Headnote-specific methods often score 0.85+ on domain relevance and legal consistency.	Semantic summarization reaches 0.70–0.80, emphasizing contextual relations but needing improvement in coherence.

Table 1: ***Analysis of Summarization Techniques***

Current abstractive summarization models struggle with domain-specific challenges in legal texts, such as:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

1. Maintaining the formal structure and terminology of legal language.
2. Preserving critical details, as legal judgments often involve nuanced reasoning.
3. Reducing redundancy without losing essential case facts.

There is a need to fine-tune and adapt abstractive techniques to create concise, accurate, and legally coherent headnotes.

III. PROPOSED STATEMENT

The main facts, legal precepts, and judicial reasoning of a case are succinctly summarized in legal case headnotes. For the general public, scholars, and legal experts to rapidly understand the essential elements of court rulings, these summaries are essential. Nonetheless, there are a number of difficulties in abstractively summarizing legal papers, particularly headnotes. The complexity of legal terminology is frequently too much for current models to manage, resulting in summaries that are erroneous, lacking, or unduly simplistic.

IV. OBJECTIVE OF THE PROPOSED RESEARCH

To develop a legal domain specific summarization model which will be fine tuned from the existing transformer based models like PEGASUS, BART to advance the models ability to understand and generate the summaries which gradually improve the accuracy of summarization and it will be tailored by creating a evaluation framework using the automated metrics ROUGE, BLUE and also it requires human evaluation based on legal accuracy

VII. Methodology

Data Collection and Preprocessing

Datasets: Legal datasets that are openly accessible will be used.

Ex: Indian Kanoon: A dataset of Indian legal cases.

Pre processing: To extract the pertinent case specifics (facts, legal principles, and judgments), the data will undergo pre processing and cleaning)

Model Selection

- **Transformer Models**: BART and PEGASUS, T5 transformer-based models will be used for performing the tasks of summarizing
- **Domain-Specific Fine-Tuning**

Model Training and Testing

To improve the legal knowledge understanding the models are pre trained and to hybrid approach will be implemented. Using an extractive model to identify the most relevant passages, then apply an abstractive summarizer to generate a concise and coherent summary.

Evaluation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

- **Automated Metrics**: Use of ROUGE and BLEU scores to measure the performance of the models based on precision, recall, and fluency.
- **Human Evaluation:** Evaluation criteria will focus on whether the generated headnotes retain the legal integrity of the case and summarize the key legal aspects effectively.

IX: Conclusion

By using the transformer-based model and hybrid model, the field of summarization has advanced significantly. The difficulties of abstractively summarizing legal case headnotes are addressed in this paper. The goal is to provide summaries that are both legally accurate and succinct by utilizing sophisticated natural language processing models designed for legal materials. These solutions can address the fundamental issues in legal case summarization by utilizing sophisticated models and evaluation frameworks, improving the training datasets, and incorporating domain-specific information. The findings of this study will greatly advance legal technology advancements and offer more effective means of dealing with legal documents, which will be advantageous to both the general public and legal experts.

References:

1. Kanithi Purna Chandu “Text Summarization Using Natural Language Processing”,International Journal of Research Publication and Reviews, Vol 3, no 11, pp 649-655,.
2. Dr. Rashmi Sharma, Shivam Chaudhary, Sejal Tyagi “TEXT SUMMARIZER USING NLP NATURAL LANGUAGE PROCESSING”,International Research Journal of Modernization in Engineering Technology and Science
3. Natalie Schluter, "The limits of automatic summarisation according to ROUGE," Valencia,2017.
4. Ravi, K., & Chandra, S **“Leveraging Legal Text Data with BERT for Predictive Legal Outcome Modeling,”** *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 12, pp. 4829-4843, December 2023.
5. Xie, S., & Zhang, Y **A Hybrid Deep Learning Model for Legal Text Analysis and Document Retrieval,”** *IEEE Access*, vol. 12, pp. 58214-58225, August 2023..
6. Kumari, R., & Jain, P. **Multi-Label Legal Document Classification Using Transformer Networks,”** *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 4, pp. 1194-1205, April 2024.
7. **Jain, R., & Garg, V. “Application of NLP for Legal Analytics: A Systematic Review and Future Directions,”** *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 7, pp. 1893–1904, July 2023
8. **Yuan, X., & Zhuang, F**.“A Transformer-Based Model for Legal Text Summarization and Document Recommendation,” IEEE Transactions on Artificial Intelligence, vol. 8, no. 3, pp. 592–605, March 2024.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

9. López, F. J., Silva, A. P., & Sánchez, M. "Automatic Legal Document Summarization using Deep Learning Models," IEEE Transactions on Artificial Intelligence, vol. 1, no. 3, pp. 213–220, 2019.

10. Torres, F. A., & Silva, G. C. A. P. "Legal Text Classification Using Deep Learning Techniques," IEEE Transactions on Artificial Intelligence, vol. 6, no. 2, pp. 235–245, 2020

11. Mahar, S., Zafar, S., & Nishat, K. (2021). "Headnote Prediction Using Machine Learning," *The International Arab Journal of Information Technology*, 18(5), 2021

12. Suryawanshi, X., Naikwadi, V., & Patil, S. (2023). "ABSTRACTIVE SUMMARIZATION OF INDIAN LEGAL DOCUMENTS USING T5 & QLoRA," *International Education and Research Journal*, 9(5), 2023

13. Anand, D., & Wagh, R. (2019). "Effective Deep Learning Approaches for Summarization of Legal Texts," *International Journal of Computer Applications*, 178(7), 2019.

14. Gode Swamy Rao, V., Katrahalli, S., Bhat, D., & Arora, T. (2024). "A Comprehensive Tool for Legal Document Interpretation and Summarization using Large Language Models," *International Journal of Communication Networks and Information Security (IJCNIS)*, 16(4), 818–824.

15. Althammer, J., et al. (2023). "Bringing order into the realm of Transformer-based language models for artificial intelligence and law," *Artificial Intelligence and Law*, 31(2), 237–258.

16. Cohan, A., Dernoncourt, F., Kim, D. S., Bui, T., Kim, S., Chang, W., & Goharian, N. (2018). "A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents," *arXiv preprint arXiv:1804.05685*.

17. Liu, Y., & Lapata, M. (2019). "Hierarchical Transformers for Multi-Document Summarization," *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5070–5081

18. Pandya, V. (2019). "Automatic Text Summarization of Legal Cases: A Hybrid Approach," *arXiv preprint arXiv:1908.09119*

19. Bhattacharya, P., Poddar, S., Rudra, K., Ghosh, K., & Ghosh, S. (2021). "Incorporating Domain Knowledge for Extractive Summarization of Legal Case Documents," arXiv preprint arXiv:2106.15876.

20. Kore, R. C., Ray, P., Lade, P., & Nerurkar, A. (2020). "Legal Document Summarization Using NLP and ML Techniques," *International Journal of Engineering and Computer Science*, 9(5), 25039–25046

21. Takale, S., Thorat, S. A., & Sajjan, R. S. (2022). "An Automatic Legal Document Summarization and Search Using Hybrid System," *ResearchGate*.

22. Ghosh, S., Dutta, M., & Das, T. (2022). "Indian Legal Text Summarization: A Text Normalisation-based Approach," *arXiv preprint arXiv:2206.06238*.