Journal of Soft Computing Exploration

Homepage: https://shmpublisher.com/index.php/joscex

p-ISSN: 2746-7686
e-ISSN: 2746-0991

.

Efficient hierarchical summarization of long legal documents using a lightweight transformer and divide and conquer strategy

Muhammad Zhafran Ammar[1], Ricky Eka Putra[2], Yuni Yamasari[3]
[1,2,3] Department of Informatics Engineering, Universitas Negeri Surabaya, Indonesia

Article Info

Article history:

Received Marh 09, 2026
Revised March 26, 2026
Accepted April 28, 2026

Keywords:

Legal document
Summarization
Natural language processing
DistilBART
Divide-and-conquer strategy

ABSTRACT

This research addresses the challenges of summarizing long and complex legal documents, which often exceed the input length limitations of transformer-based models and contain intricate legal reasoning structures. The purpose of this study is to develop an efficient and scalable summarization framework that preserves semantic fidelity and structural coherence in judicial summaries. To achieve this objective, a hybrid summarization pipeline is proposed by integrating a Bidirectional Encoder Representations from Transformers (BERT)-based extractive model with a hierarchical abstractive model based on Distilled Bidirectional and Auto-Regressive Transformers (DistilBART), combined with a Divide-and-Conquer strategy. The proposed method partitions long legal documents into smaller segments, processes each segment independently, and reconstructs them into a coherent final summary. Experiments were conducted on the Indian Legal Case Summarization dataset and evaluated using Recall-Oriented Understudy for Gisting Evaluation (ROUGE), BERTScore, and Cosine Similarity to assess both lexical overlap and semantic similarity. The results show that the hierarchical DistilBART model outperforms the extractive baseline, achieving a ROUGE-1 score of 0.3802 and a Cosine Similarity of 0.6917. These findings demonstrate that the proposed framework provides an effective solution for long-document summarization in the legal domain.

Corresponding Author:

Ricky Eka Putra,
Department of Informatics Engineering,
Universitas Negeri Surabaya,
Jl. Ketintang Wiyata, Ketintang, Kec. Gayungan, Surabaya, Jawa Timur 60231
Email: rickyeka@unesa.ac.id
https://doi.org/10.52465/joscex.v7i2.5

1. INTRODUCTION

The rapid digitalization of the global legal sector has precipitated an unprecedented surge in the volume of judicial documents, case laws, and regulatory filings. This massive influx of digital data creates a significant bottleneck for legal professionals who must process and analyze extensive texts to extract pertinent information [1]. Consequently, automated text summarization has emerged as a critical technology to mitigate information overload by providing condensed versions of long documents while preserving essential legal reasoning and factual findings [2]. However, summarizing long-form legal documents presents unique challenges that distinguish it from general domain summarization. Legal texts are characterized by specialized terminology, complex discourse structures, and extreme lengths that frequently exceed the processing capacities of standard computational models [3]. Furthermore, these documents possess an intricate rhetorical structure that includes distinct segments such as facts, issues, arguments, precedents, statutes, and final rulings [3], [4]. Large-scale datasets like EUR-Lex-Sum highlight the severity of this scale, as documents in this domain often encompass tens of thousands of tokens, making manual review nearly impossible [5].

In recent years, the field of legal summarization has transitioned from traditional extractive methods to more sophisticated abstractive approaches. Initial state of the art solutions relied on standard Transformer-based models such as BERT, Legal-BERT, and Pegasus, which demonstrated high performance in capturing semantic nuances in both general and legal domains [6], [7]. Despite their success, these models are fundamentally hindered by the quadratic complexity of their self-attention mechanisms, which typically limits input capacity to 512 or 1,024 tokens [8]. To circumvent this length constraint, researchers developed efficient Transformer variants like Longformer and Legal-LED, which utilize sparse or windowed attention to process sequences of up to 16,384 tokens [9], [10]. Models such as LongT5 and Legal-Pegasus further extended these capabilities through specialized pre-training but still face significant constraints when deployed on consumer-grade hardware [11], [12]. Furthermore, a divide-and-conquer strategy has been proposed to partition long documents into manageable chunks that are summarized independently before being aggregated into a coherent final output [13], [14]. Other approaches have explored hierarchical attention layers and semantic self-segmentation to reflect the natural structure of legal documents and maintain global context [15], [16]. While chunking methods combined with short-document summarizers like BART have shown promising results in legal case evaluations, they often struggle with maintaining inter-segment consistency and global semantic flow [17].

Despite these advancements, existing methodologies exhibit significant limitations that constitute a critical research gap. While efficient Transformers like Longformer and LED extend the input window, they still require substantial memory and computational power, which may be inaccessible to many legal practitioners in resource-constrained environments [18], [19]. Moreover, these models often struggle with the "lost in the middle" phenomenon or a lack of global coherence when the source text is exceptionally long [3], [9]. Input truncation, a common practice even with extended limits, discards content that may be vital for a complete summary and leads to destructive semantic loss, particularly in low-resource legal settings [16]. Many current systems still revert to truncation when a document exceeds the extended limit, potentially losing vital information in a domain where every detail can alter a legal outcome long [3], [20]. Although the divide-and-conquer approach is promising, many implementations do not fully utilize lightweight architectures designed for efficiency, instead relying on heavy models that negate the speed benefits of the strategy [16]. Additionally, the synergy between memory-efficient techniques and structured hierarchical chunking specifically for the legal domain remains under-explored [21].

While some researchers have focused on modifying existing heavy architectures, such as replacing attention with the Hartley transform for long-context legal tasks [19], there is limited research regarding the integration of truly lightweight Transformer models with a hierarchical divide-and-conquer strategy tailored for legal structures. Most current solutions are either computationally expensive or lose the hierarchical relationship between different parts of a legal case, failing to adequately represent the rhetorical roles across segments [17]. Therefore, this research intends to develop an efficient hierarchical summarization framework using a lightweight Transformer architecture paired with a divide-and-conquer strategy. The uniqueness of this paper lies in its ability to process extremely long legal documents through a multi-stage hierarchical process that significantly reduces the computational footprint while maintaining the structural integrity and semantic flow of the summary. This involves semantic chunking aligned with rhetorical roles, lightweight encoding with sparse mechanisms, and iterative aggregation to ensure no information is lost due to truncation.

The objective of this research is to design a lightweight Transformer variant optimized for legal texts and to implement a hierarchical divide-and-conquer pipeline that avoids information loss from truncation. Furthermore, the study aims to evaluate model efficiency in terms of memory usage and inference speed on standard hardware, as well as assess accuracy against established benchmarks such as ROUGE, BERTScore, and domain-specific metrics on datasets like IN-Abs, UK-Abs, and CLSum [17], [21]. By integrating these strategies, this research is expected to provide a more inclusive and efficient solution for future legal text analytics systems that can be deployed effectively even with limited computational resources [7], [22], [23].

2. METHOD

The methodology of this research is designed to address the challenges of summarizing long and complex legal documents by comparing a sentence-level extractive baseline with a hierarchical abstractive approach. This section presents the systematic procedures employed to preprocess legal documents, implement transformer-based models, and evaluate the generated summaries against expert-written headnotes. To ensure systematic analysis and reproducibility, the research is conducted through a structured pipeline.

To ensure a systematic analysis and reproducibility, the research is conducted through a structured pipeline consisting of several stages. The overall workflow of the proposed methodology is illustrated in Figure 1.

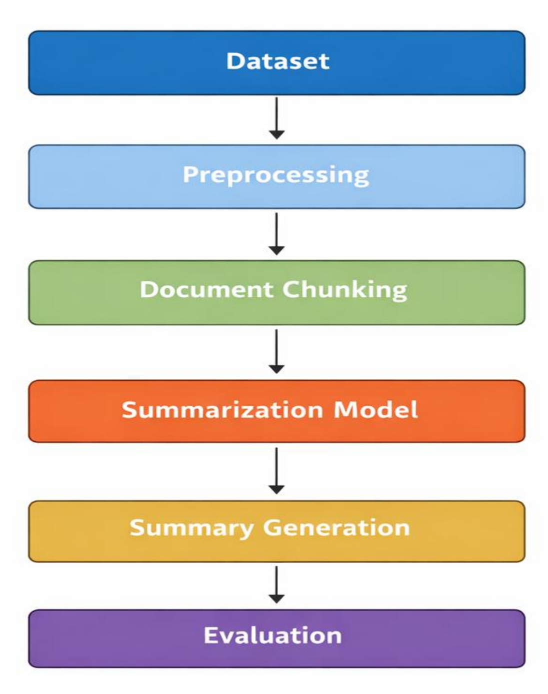

Figure 1. Research stages

The process begins with Data Acquisition, which involves retrieving the Indian Legal Case Summarization dataset that provides long-form judicial judgments paired with expert headnotes [21]. Following this, Text Preprocessing cleans the raw legal texts by removing noise, normalizing formatting, and segmenting the text into individual sentences to prepare it for numerical representation [22]. To address the input limitations of transformer models, Document Chunking partitions the preprocessed text into smaller, non-overlapping chunks of approximately 1,024 tokens each [14].

The Summarization Model Implementation encompasses two approaches: the Extractive Baseline, where sentences are transformed into embeddings using BERT to identify the most salient information through semantic matching [23]; and the Abstractive Model, which employs a lightweight DistilBART model to generate fluent, synthesized partial summaries for each document chunk [17]. Summary Generation then aggregates these partial summaries chronologically to reconstruct a comprehensive final summary that covers the entire judgment [13]. Finally, Performance Evaluation quantifies the quality of the final summaries using a multi-metric approach, including ROUGE for lexical overlap, BERTScore for token-level semantic alignment, and Cosine Similarity for document-level vector similarity [24].

Dataset

This research utilizes the IN-Abs dataset, a specialized benchmark designed for the abstractive summarization of long legal documents, as introduced by Shukla et al. at the 2022 AACL workshop on Natural Legal Language Processing [17]. The corpus is curated from the Legal Information Institute of India and comprises comprehensive judicial judgments from the Supreme Court of India [12], [17]. These documents present a distinct challenge for conventional NLP models due to their extreme length, dense legal vocabulary, and complex hierarchical structures [7], [25].

The dataset provides pairs of original court judgments and expert-authored "headnotes," which serve as the gold standard for evaluating abstractive summarization performance [17], [25]. The fundamental attributes and descriptions of the fields within the IN-Abs dataset are detailed in Table 1.

Tabel 1. Dataset attributes and descriptions

Field	Description
Case Judgment	The full text of the original judicial ruling, encompassing factual backgrounds, arguments, and legal ratios [8], [26].
Summary	An abstractive headnote manually written by legal experts to capture the core essence and legal principles of the case [26], [31].
Document Length	The total word count of the original judgment, which reaches an average of 4,368 to 4,782 words [8].
Summary Length	The total word count of the expert summary, with an average length ranging between 839 and 932 words [8].

In terms of quantitative distribution, the complete IN-Abs corpus contains a total of 7,130 document-summary pairs [7], [17]. This collection is formally partitioned into a training set consisting of 7,030 pairs used for model fine-tuning and a standardized test set consisting of 100 pairs used for performance benchmarking [7], [17]. For the purposes of this study, we specifically employ the 100-document test set to evaluate our proposed model. This subset is recognized in legal NLP literature as a rigorous benchmark for testing the efficiency of summarization systems when handling documents that far exceed the token limitations of standard Transformer architectures [7], [25].

Text Processing

Raw legal documents are characterized by idiosyncratic formatting, complex syntactic structures, and significant noise, all of which can impede the performance of Transformer-based models. In this study, the preprocessing stage is critical for cleaning and transforming the raw text into a structured representation suitable for hierarchical summarization [24]. The initial phase involves rigorous sentence segmentation, which requires specialized handling of legal-specific abbreviations such as "v." or "no." to prevent erroneous segmentation and preserve the integrity of legal citations. Following segmentation, noise removal is performed to eliminate non-textual artifacts, including page numbers, headers, and formatting remnants from the document conversion process. To further refine the input, extremely short or uninformative sentences that fall below a predefined minimum word threshold are filtered out, ensuring that the model focuses on content-rich sequences [24].

The cleaned text is subsequently tokenized using the specific tokenizer associated with the selected Transformer architecture. Maintaining strict consistency between the preprocessing pipeline and the model's pre-trained vocabulary is essential to ensure that the input is represented accurately without significant information loss. Given that legal documents in the IN-Abs dataset typically span 4,000 to 5,000 words, they far exceed the input capacity of standard Transformer models. To address this, a divide-and-conquer chunking strategy is implemented, wherein the document is organized into fixed-length segments, typically restricted to 512 or 1,024 tokens. This strategy often incorporates a sliding window or overlap between consecutive chunks to maintain contextual coherence and mitigate information loss at segment boundaries. Such a systematic approach is necessary to avoid lead bias, which is the tendency of models to prioritize earlier sections of a text, and ensures that the entirety of the long legal document is processed without truncation [16]. Collectively, these preprocessing steps produce a clean, segmented, and tokenized output that serves as the foundation for the subsequent hierarchical summarization stage.

BERT Extractive Summarization

The baseline extractive summarization model is implemented using the BERT-base-uncased architecture, which utilizes a bidirectional Transformer-based approach to generate deep contextualized representations [26]. To derive a fixed-dimensional embedding for each sentence S_i, the model applies a mean pooling strategy over the token-level outputs rather than relying exclusively on the [CLS] token representation. This method ensures that the resulting 768-dimensional dense vector captures the collective semantic information of all tokens within the sentence [27]. This transformation allows the textual components of long legal documents to be represented as a sequence of dense semantic vectors suitable for mathematical comparison.

A global document representation D is subsequently established by calculating the arithmetic mean of all sentence embeddings within the document [28]. This document centroid serves as a proxy for the overarching thematic context, representing the aggregate semantic information of the entire text. The importance of an individual sentence is then quantified by measuring its semantic proximity to this document centroid using Cosine Similarity [24]. The importance score for each sentence is determined by the following formulation:

$$score_i = \frac{S_i \cdot D}{\|S_i\|\|D\|} \quad (1)$$

Where:
$score_i$: The importance score of the i-th sentence.
S_i: The embedding vector of the i-th sentence
D: The global document embedding (centroid)

In this equation, $score_i$ denotes the importance weight of the i-th sentence, while S_i and D represent the sentence embedding and the document centroid vector, respectively. Following the calculation of these scores, all sentences are ranked in descending order. The final extractive summary is formed by selecting the top-ranked sentences until a specific length constraint or a predefined k number of sentences is reached [29]. To maintain the simplicity and interpretability of this baseline, no additional redundancy reduction techniques, such as Maximal Marginal Relevance, are applied during the selection process. This unsupervised approach provides a computationally efficient and transparent baseline for long document summarization, establishing a fundamental performance benchmark for the hierarchical abstractive model discussed in the subsequent section.

DistilBART

DistilBART is a distilled variant of the Bidirectional and Auto-Regressive Transformers model, specifically optimized for sequence-to-sequence tasks such as text summarization [1]. It utilizes a standard Transformer architecture that integrates a bidirectional encoder with a left-to-right autoregressive decoder [30]. The distillation process involves training a smaller student model to emulate the behavior of a larger teacher model, resulting in a lightweight architecture that maintains high performance while significantly reducing computational latency during inference [1]. This efficiency is particularly advantageous for processing large-scale legal datasets like IN-Abs, where processing speed is as critical as output quality.

The model functions by mapping an input sequence X to a latent representation through the encoder, which the decoder then utilizes to generate a target sequence Yin a token-by-token manner [30]. The primary training objective of DistilBART is to maximize the conditional log-likelihood of the generated summary, which is mathematically expressed as follows:

$$L = \sum_{t=1}^{T} logP(y_t | y_{<t}, X) \quad (2)$$

Where:
X: The input document or text chunk
y_t: The token generated at time step t.
$y_{<t}$: The sequence of tokens generated before time step t.

In this formulation, X represents the input document or text chunk, y_t denotes the token generated at time step t, and $y_{<t}$ signifies the sequence of tokens generated prior to time step t [30]. Despite its effectiveness, the standard DistilBART architecture is constrained by a fixed input token limit, typically 1,024 tokens. This limitation is problematic for legal documents in the IN-Abs dataset, which often contain between 4,000 and 5,000 words, as direct processing would result in the loss of critical information due to truncation.

To overcome these constraints, this study implements a hierarchical divide-and-conquer strategy [14]. In the divide stage, the long legal document is partitioned into smaller, manageable chunks of approximately 1,024 tokens each. During the conquer stage, each individual chunk is summarized independently by the DistilBART model to generate a localized partial summary [15]. Finally, in the combine stage, these partial summaries are concatenated in their original sequence to produce a comprehensive final summary that captures the entire progression of the judgment, from the initial facts to the final ruling [13]. This hierarchical approach ensures broad document coverage and prevents the lead bias often found in standard Transformer applications. Unlike the extractive BERT method which focuses on sentence selection, this DistilBART-based framework generates fluent, abstractive summaries that synthesize the semantic essence of the legal text, serving as the primary proposed methodology for this research.

.

Evaluation Metrics

The performance of the summarization models is quantified using three distinct evaluation frameworks to ensure a robust assessment of both lexical precision and semantic fidelity. The first metric employed is ROUGE, which stands for Recall-Oriented Understudy for Gisting Evaluation, and it serves as the standard for measuring n-gram overlap between generated outputs and reference summaries [11]. This research specifically utilizes ROUGE-1, ROUGE-2, and ROUGE-L to assess unigram overlap, bigram overlap, and the longest common subsequence, respectively [11]. To prioritize the coverage of information from the original legal text, the ROUGE-N recall score is calculated using the following equation [31]:

$$ROUGE\text{-}N=\frac{\sum_{S\in\{Ref\}}\sum_{gram_n\in S}Count_{match}(gram_n)}{\sum_{S\in\{Ref\}}\sum_{gram_n\in S}Count(gram_n)} \quad (3)$$

In this formulation, $Count_{match}$ represents the number of n-grams that co-occur in both the generated and reference summaries, while Count signifies the total number of n-grams present within the reference summary.

To address the limitations of surface-level lexical matching, BERTScore is utilized to evaluate the semantic similarity between summaries by leveraging contextual embeddings from pre-trained Transformer models [11]. This metric performs greedy matching between candidate and reference tokens based on cosine similarity, which allows the evaluation to recognize synonyms and paraphrased content that ROUGE might otherwise penalize [3]. The BERTScore F_{BERT} is derived from the precision and recall of token alignments as follows [31]:

$$F_{BERT}=2\cdot\frac{P_{BERT}\cdot R_{BERT}}{P_{BERT}+R_{BERT}} \quad (4)$$

Within this equation, P_{BERT} and R_{BERT} represent the precision and recall scores calculated using the cosine similarity of the contextual embedding vectors for each token pair. A higher F_{BERT} value indicates greater semantic alignment between the model output and the ground truth.

Furthermore, Cosine Similarity is implemented to provide a holistic measure of semantic alignment at the document level [24]. Unlike token-level metrics, this approach evaluates the global thematic consistency of the summary. Both the final generated summary and the reference headnote are transformed into single high-dimensional vectors, denoted as V_{gen} and V_{ref}, using a Sentence-BERT encoder [26]. The degree of similarity is determined by the cosine of the angle between these two vectors through the following calculation [24]:

$$CosineSimilarity=\frac{V_{gen}\cdot V_{ref}}{|V_{gen}||V_{ref}|} \quad (5)$$

In this context, a score approaching 1 indicates that the generated summary is semantically identical to the reference, whereas a score near 0 suggests a lack of semantic correlation. These three metrics collectively provide a comprehensive evaluation framework. By integrating ROUGE for structural overlap, BERTScore for token-level semantic understanding, and Cosine Similarity for global contextual preservation, the study ensures an objective analysis of the summarization model's performance [17].

3. RESULTS AND DISCUSSIONS

The experimental results demonstrate the efficacy of the proposed zero-shot hierarchical framework in addressing the inherent challenges of long legal document summarization. At the onset of the pipeline, the preprocessing stage proved instrumental in transforming raw judicial texts into a structured format suitable for Transformer-based processing. By utilizing regular expression-based sentence segmentation, the system successfully maintained the integrity of complex legal citations and abbreviations, such as "v." or "no.", which frequently cause erroneous breaks in standard sentence splitters. The implementation of a length-based filtering threshold was particularly effective; the removal of segments with fewer than 10 characters successfully eliminated non-substantive administrative artifacts and formatting noise without compromising the legal narrative. This lightweight preparation facilitated a robust divide and conquer strategy, where documents averaging between 4,000 and 5,000 words were organized into discrete chunks with a heuristic limit of approximately 512 tokens. This structural organization is vital for long-sequence modeling, as it ensures that the entirety of the judicial discourse is captured and processed in a sequential manner, thereby preserving the logical flow of the legal reasoning from initial facts to final rulings [14]. By avoiding the common pitfall of input truncation, which often destroys the semantic integrity of a legal case, the hierarchical approach ensures that valuable information in the latter portions of a judgment is not ignored [16].

The quantitative performance of the zero-shot framework provides clear evidence of the superiority of the hierarchical abstractive approach over the extractive baseline. The evaluation utilized ROUGE for lexical precision, BERTScore for contextual token-level alignment, and Cosine Similarity for global document-level vector similarity. These metrics collectively highlight the models' ability to synthesize legal information from the IN-Abs dataset, as summarized in Table 2 and further illustrated in Figure 2 through comparative performance visualizations.

Tabel 2. Performance evaluation results

Model Implementation	Rouge-1	Rouge-2	Rouge-L	BERTScore F1	Cosine Similarity
BERT Baseline	0.3417	0.1690	0.1999	0.8332	0.6534
DistilBART + Divide-and-Conquer	**0.3802**	**0.1865**	**0.2128**	**0.8352**	**0.6917**

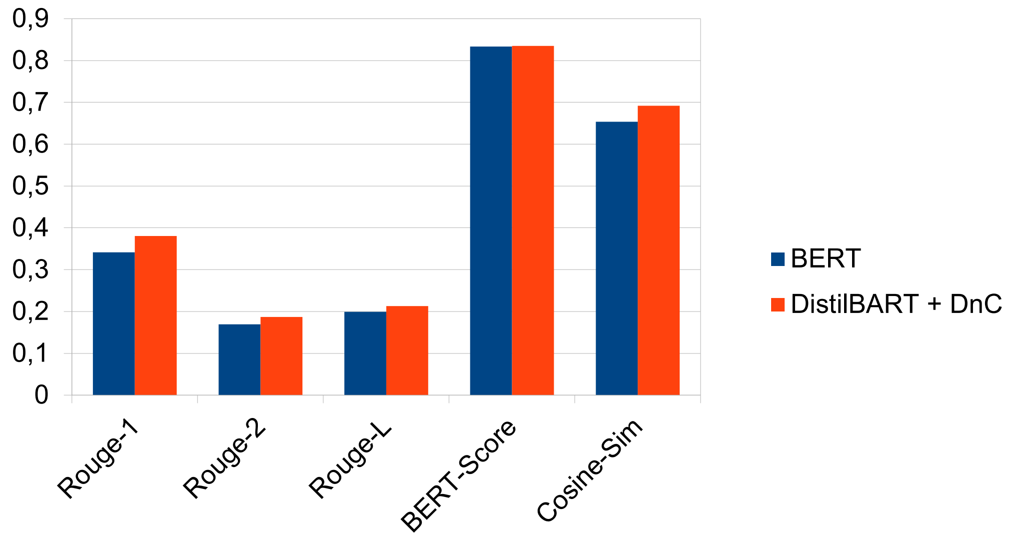

Figure 2. Comparison evaluation results

As illustrated in Figure 2, the grouped bar chart comparing ROUGE scores reveals a consistent upward trend for the DistilBART implementation. The improvement in ROUGE-1 from 0.3417 to 0.3802 suggests that the abstractive model is more capable of capturing the primary keywords and legal entities essential to a concise headnote. The gain in ROUGE-L specifically reflects a better preservation of the longest common subsequence, indicating that the model maintains the structural order of the judicial argument more effectively than simple sentence selection. Furthermore, the increase in Cosine Similarity to 0.6917, a notable 5.8% gain over the baseline, demonstrates that the hierarchical strategy achieves a much higher degree of global semantic alignment with the reference summaries. These performance gains are particularly remarkable given that they were achieved in a strictly zero-shot setting, without any parameter optimization or fine-tuning on legal-specific corpora. The results suggest that the inherent architecture of DistilBART, combined with a structured processing framework, can effectively generalize to the specialized domain of law by leveraging its pre-trained linguistic knowledge [1].

The underperformance of the BERT extractive baseline can be attributed to its fundamental limitation as a sentence selection mechanism. In long legal documents, an extractive approach often results in a disjointed collection of sentences that lack the cohesion and logical transitions necessary for an effective legal summary. This lack of discourse coherence is especially problematic in the legal domain, where the summary must synthesize complex arguments rather than merely repeat isolated statements. In contrast, the DistilBART model leverages an encoder-decoder architecture that facilitates semantic rewriting, allowing the system to condense and paraphrase information from multiple sentences within a chunk into a singular, fluid narrative. This ability to perform abstractive synthesis is crucial for generating summaries that mirror the professional quality of expert-written headnotes, which often require a high degree of linguistic abstraction and conceptual merging.

Furthermore, the effectiveness of the divide and conquer strategy is demonstrated by its ability to mitigate the information loss typically caused by the lead bias observed in standard Transformers. Standard models often prioritize the beginning of a document due to fixed context windows, which is detrimental for legal texts where the final judgment and ratio decidendi are often located at the end of the file [16]. By processing each chunk independently and subsequently merging the partial summaries, the proposed framework ensures that the critical legal findings situated throughout the document are accurately represented in the final output. The implementation of an additional filtering step to skip chunks with fewer than 50 tokens further refined the output quality, ensuring that the final summary remained focused on high-density legal content rather than procedural filler. This hierarchical approach successfully preserves the global context by ensuring each segment of the judgment contributes to the final synthesis.

The practical and scientific implications of this methodology are significant for the legal technology domain. The zero-shot nature of the framework means that high-quality summaries can be generated without the need for prohibitively expensive and time-consuming human-annotated datasets. This makes the system highly scalable and efficient for legal practitioners who must navigate vast volumes of judicial documentation daily. By utilizing a lightweight Transformer model like DistilBART within a modular processing strategy, the framework offers a generalizable solution that can be deployed on standard hardware without the computational overhead required by larger language models [1]. Furthermore, the comparative performance across different architectures can be visualized through bar charts or comparative plots to provide a more intuitive representation of these metric variances. Ultimately, the findings confirm that a hierarchical abstractive approach is superior to traditional extractive methods for long legal document summarization, providing a robust pathway for future developments in automated judicial assistance [17].

4. CONCLUSION

This study has successfully demonstrated that a hierarchical abstractive framework, integrating a divide and conquer strategy with the DistilBART architecture, provides a robust and efficient solution for the summarization of long legal documents. By overcoming the structural limitations of standard Transformer models, the research confirms that semantic synthesis through a structured pipeline is superior to traditional sentence-level extractive methods. The implementation of document chunking effectively mitigated the lead bias common in judicial texts, ensuring that critical legal reasoning and final rulings located at the end of a judgment are accurately preserved within the generated headnotes.

The empirical findings indicate that the hierarchical DistilBART approach significantly outperforms the BERT extractive baseline in terms of both lexical accuracy and contextual alignment. These results validate the viability of lightweight models in highly specialized domains, proving that they possess sufficient linguistic knowledge to capture nuanced legal logic even in a strictly zero-shot setting. Consequently, the proposed framework offers a scalable and computationally efficient alternative for real-world legal technology applications where processing speed, resource optimization, and semantic fidelity are paramount.

Future research should explore the integration of domain-specific pre-trained models and multi-step refinement strategies to further enhance the precision of abstractive generation. Additionally, evaluating the performance of this hierarchical framework across diverse legal systems and other long-form document types, such as scientific articles or regulatory filings, would establish its broader scientific and practical applicability.

REFERENCES

[1] M. F. Kabir, S. A. Mitu, S. Sultana, B. Hossain, R. Islam, and K. R. Ahmed, "LegalSummNet: A Transformer-Based Model for Effective Legal Case Summarization," *Int. J. Adv. Comput. Sci. Appl.*, 2025, [Online]. Available: https://api.semanticscholar.org/CorpusID:281858404

[2] V. Shevchuk, "Enhancing Text Summarization in a Limited Data and Long Sample Setting," *Res. Sq.*, 2024, doi: 10.21203/rs.3.rs-4535506/v1.

[3] D. Mohan and L. Nair, "Deep Learning-Based Semantic Segmentation for Legal Texts: Unveiling Rhetorical Roles in Legal Case Documents," *E3S Web Conf.*, vol. 529, May 2024, doi: 10.1051/e3sconf/202452904019.

[4] V. Malik *et al.*, *Semantic Segmentation of Legal Documents via Rhetorical Roles*. 2021. doi: 10.48550/arXiv.2112.01836.

[5] D. Aumiller, A. Chouhan, and M. Gertz, "EUR-Lex-Sum: A Multi- and Cross-lingual Dataset for Long-form Summarization in the Legal Domain," in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Y. Goldberg, Z. Kozareva, and Y. Zhang, Eds., Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 7626–7639. doi: 10.18653/v1/2022.emnlp-main.519.

[6] S. Dadas, M. Kozłowski, R. Poświata, M. Perełkiewicz, M. Białas, and M. Grębowiec, "A support system for the detection of abusive clauses in B2C contracts," *Artif. Intell. Law*, vol. 33, no. 4, pp. 913–951, 2025, doi: 10.1007/s10506-024-09408-8.

[7] A. Deroy, K. Ghosh, and S. Ghosh, "How Ready are Pre-trained Abstractive Models and LLMs for Legal Case Judgement Summarization?," *ArXiv*, vol. abs/2306.0, 2023, [Online]. Available: https://api.semanticscholar.org/CorpusID:259064225

[8] D. Mamakas, P. Tsotsi, I. Androutsopoulos, and I. Chalkidis, "Processing Long Legal Documents with Pre-trained Transformers: Modding LegalBERT and Longformer," *ArXiv*, vol. abs/2211.0, 2022, [Online]. Available: https://api.semanticscholar.org/CorpusID:253254835

[9] P. G. Devi, A. Agrawal, A. Dubey, and C. R. Chowdary, "Long document summarization using page specific target text alignment and distilling page importance," *ArXiv*, vol. abs/2509.1, 2025, [Online]. Available: https://api.semanticscholar.org/CorpusID:281420751

[10] J. Niklaus and D. Giofré, "BudgetLongformer: Can we Cheaply Pretrain a SotA Legal Language Model From Scratch," *Second Work. Effic. Nat. Lang. Speech Process.*, 2022.

[11] A. Deroy, K. Ghosh, and S. Ghosh, "Applicability of large language models and generative models for legal case judgement summarization," *Artif. Intell. Law*, vol. 33, no. 4, pp. 1007–1050, 2025, doi: 10.1007/s10506-024-09411-z.

[12] D. Datta, S. Soni, R. Mukherjee, and S. Ghosh, *MILDSum: A Novel Benchmark Dataset for Multilingual Summarization of Indian Legal Case Judgments*. 2023. doi: 10.18653/v1/2023.emnlp-main.321.

[13] M. Sie, R. Beek, M. Bots, S. Brinkkemper, and A. Gatt, *Summarizing Long Regulatory Documents with a Multi-Step Pipeline*. 2024. doi: 10.18653/v1/2024.nllp-1.2.

[14] A. Gidiotis and G. Tsoumakas, "A Divide-and-Conquer Approach to the Summarization of Long Documents," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 3029–3040, 2020, [Online]. Available: https://api.semanticscholar.org/CorpusID:219425367

[15] B. Pang, E. Nijkamp, W. Kryscinski, S. Savarese, Y. Zhou, and C. Xiong, *Long Document Summarization with Top-down and Bottom-up Inference*. 2023. doi: 10.18653/v1/2023.findings-eacl.94.

[16] G. Moro and L. Ragazzi, "Semantic Self-Segmentation for Abstractive Summarization of Long Documents in Low-Resource Regimes," *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 10 SE-AAAI Technical Track on Speech and Natural Language Processing, pp. 11085–11093, Jun. 2022, doi: 10.1609/aaai.v36i10.21357.

[17] A. Shukla *et al.*, "Legal Case Document Summarization: Extractive and Abstractive Methods and their Evaluation," in *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Y. He, H. Ji, S. Li, Y. Liu, and C.-H. Chang, Eds., Online only: Association for Computational Linguistics, Nov. 2022, pp. 1048–1064. doi: 10.18653/v1/2022.aacl-main.77.

[18] N. Aletras, I. Chalkidis, L. Barrett, C. Goanț\ua, and D. Preoțiuc-Pietro, Eds., "Proceedings of the Natural Legal Language Processing Workshop 2022," Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics, Dec. 2022. doi: 10.18653/v1/2022.nllp-1.0.

[19] D. Giofr'e and S. Ghantasala, "Legal-HNet: Mixing Legal Long-Context Tokens with Hartley Transform," *ArXiv*, vol. abs/2311.0, 2023, [Online]. Available: https://api.semanticscholar.org/CorpusID:265066982

[20] C. M. Greco and A. Tagarelli, "Bringing order into the realm of Transformer-based language models for artificial intelligence and law," *Artif. Intell. Law*, vol. 32, no. 4, pp. 863–1010, 2024, doi: 10.1007/s10506-023-09374-7.

[21] S. Liu, J. Cao, Y. Li, R. Yang, and Z. Wen, "Low-resource court judgment summarization for common law systems," *Inf. Process. Manag.*, vol. 61, no. 5, p. 103796, 2024, doi: https://doi.org/10.1016/j.ipm.2024.103796.

[22] S. Ghosh, M. Dutta, and T. Das, *Indian Legal Text Summarization: A Text Normalisation-based Approach*. 2022. doi: 10.48550/arXiv.2206.06238.

[23] S. Ghosh, D. Verma, B. Ganesan, P. Bindal, V. Kumar, and V. Bhatnagar, "Human Centered AI for Indian Legal Text Analytics," *ArXiv*, vol. abs/2403.1, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:268512835

[24] J. Sihombing, A. Arnita, S. Idrus, and D. Niska, "Implementation of text summarization on indonesian scientific articles using textrank algorithm with TF-IDF web-based," *J. Soft Comput. Explor.*, vol. 5, pp. 310–319, Dec. 2024, doi: 10.52465/joscex.v5i3.475.

[25] S. Sharma and P. Singh, *Domain-Specific Summarization: Optimizing InLegalBERT for Indian Judgment Reports*. 2023. doi: 10.21203/rs.3.rs-3792484/v1.

[26] J. Ju, M. Liu, L. Gao, and S. Pan, "Monash-Summ@{L}ong{S}umm 20 {S}ci{S}umm{P}ip: An Unsupervised Scientific Paper Summarization Pipeline," in *Proceedings of the First Workshop on Scholarly Document Processing*, M. K. Chandrasekaran, A. de Waard, G. Feigenblat, D. Freitag, T. Ghosal, E. Hovy, P. Knoth, D. Konopnicki, P. Mayr, R. M. Patton, and M. Shmueli-Scheuer, Eds., Online: Association for Computational Linguistics, Nov. 2020, pp. 318–327. doi: 10.18653/v1/2020.sdp-1.37.

[27] M. Zhong, P. Liu, Y. Chen, D. Wang, X. Qiu, and X. Huang, "Extractive Summarization as Text Matching," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds., Online: Association for Computational Linguistics, Jul. 2020, pp. 6197–6208. doi: 10.18653/v1/2020.acl-main.552.

[28] Y. Dong, A. Mircea, and J. C. K. Cheung, "Discourse-Aware Unsupervised Summarization for Long Scientific Documents," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, P. Merlo, J. Tiedemann, and R. Tsarfaty, Eds., Online: Association for Computational Linguistics, Apr. 2021, pp. 1089–1102. doi: 10.18653/v1/2021.eacl-main.93.

[29] M. Munaf, H. Afzal, K. Mahmood, and N. Iltaf, "Low Resource Summarization using Pre-trained Language Models," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 23, Jul. 2024, doi: 10.1145/3675780.

[30] P. Mehta, "From Extractive to Abstractive Summarization: A Journey," in *Proceedings of the {ACL} 2016 Student Research Workshop*, H. He, T. Lei, and W. Roberts, Eds., Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 100–106. doi: 10.18653/v1/P16-3015.

[31] H. Nguyen, H. Chen, L. Pobbathi, and J. Ding, "A Comparative Study of Quality Evaluation Methods for Text Summarization," *ArXiv*, vol. abs/2407.0, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:270869494