

A Comparative Review of Long Short-Term Memory Based Encoder-Decoder, Transformer, and Hybrid Architectures for Text Summarization

Tulsipriya Hanglem¹, Takhellambam Sonamani Singh¹, Khundrakpam Veeshel Singh²

¹ Department of Computer Science and Engineering, Manipur Institute of Technology, Imphal, Manipur, India 795003

² National Institute of Electronics and Information Technology, Akampat, Imphal East, Manipur, India 795001

tulsipriya.hanglem@manipuruniv.ac.in, sona.cse2013@gmail.com, vslkhun@gmail.com

Abstract. Natural language processing is a recent phenomenon that has been growing in recent years owing to the availability of text documents online. With these massive repositories of text information, researchers have focused on automatically and efficiently extracting information by using text summarization systems. Their application is found in social network analysis, market intelligence, educational systems, and financial analysis. In this paper, the recent progress in automatic summarization is comprehensively described. This paper concerns the different optimization methods in text summarization. We present a taxonomy that illustrates the use of different strategies in automatic text summarization in situations where there is single or multi-document automatic text summarization. We present the summaries of some representative state-of-the-art automated text summarization methods and compare them in benchmark datasets. Possible applications, research problems, open issues, and future directions of the automatic text summarization sphere are also addressed in the paper. Extractive text summarization techniques are gaining increased popularity because they are capable of generating an informative summary with minimal linguistic resources. The next step after converting a sentence to a bag of words will be to first select important sentences in the text, which we shall summarize. Due to the strong data dependence on the performance of the methods, the effect of the parameter on different datasets is examined. Furthermore, we also consider the methods of assessment such as ROUGE metrics and more recent semantic-based ones to assess quality of the summary and the competitiveness of the facts.

Keywords: Abstractive summarization, Transformer, LSTM, encoder-decoder models, text creation, and judging summaries.

(Article history: Selected papers from NIELIT's International Conference on Communication, Electronics and Digital Technologies (NICE-DT'26), 29-30 Jan, 2026, Gorakhpur, India)

1 Introduction

Over the past few years, neural methods have been applied to abstractive text summarization systems, which have been restructured to generate textual summaries in a fluent and coherent way. Many approaches of text summarization have been tried in divergent disciplines such as news analytics, information retrieval and automatic content generation [28]. Recurrent encoder-decoder engines with attention processes were applied in most of the initial neural experiments, giving more extractive approaches with layers of abstraction but were unable to scale to long documents and had trouble processing them [1, 6, 36]. These restrictions became more pronounced as documents became longer and datasets expanded, and thus, the methods to summarize long documents were the focus [36]. The introduction of self-attention-based Transformer models marked a turning point in the field of summarization and allowed modeling the global context and parallel computation [2]. The architecture based on transformers became extremely popular overnight since it performed better on benchmark datasets and could be pre-trained at massive scales [3,13]. New issues, instead, were highlighted by these models, including high computation costs, memory requirements, factual inconsistencies and long document handling issues [23, 36]. In order to address these issues, recent studies proposed hybrid designs that combine recurrent neural networks with self-attention models, aiming to balance efficiency, sequential sensitivity, and contextual representation [14,15]. Although an increasing amount of research is on LSTM-based models, Transformer-based models, and hybrid models, most of the survey articles only provide the general overviews without the year-by-year comparisons and trend analysis visually, rendering them less useful in the analysis [7,22,28]. This review is guided by the fact that a systematic comparative study is required to explain the course of architectural improvement, performance losses, and emergent tendencies in neural abstractive summarization. The previous work has been arranged in the form of year-by-year tabular comparison and comparative graphs, which facilitates understanding of how the architectures of summarization have evolved over the years and the areas of research in the future.

2 Background on Text Summarization Techniques

2.1 Extractive vs. Abstractive Approaches

Text summarization has two types: extractive and abstractive methods. The extractive method chooses sentences or phrases from the original document. This method looks at things like how a word is used where the sentence is in the document how



similar the sentences are and it uses special ranking systems like the ones made by Mihalcea & Tarau and Erkan & Radev [26, 27]. This method requires a deeper understanding of meaning, the ability to generate language, and good contextual modeling. While it is more difficult, abstractive summarization creates summaries that are more coherent and informative. This makes it increasingly appealing for real-world uses [6,24].

2.2 Neural Architectures in Summarization

Currently, abstractive summarization is most often done using neural networks. Initial neural models for this task were built around recurrent neural networks, specifically LSTMs and GRUs, and used an encoder-decoder structure. The decoder's ability to concentrate on the important sections of the original text at any given time was improved by attention mechanisms [6,9]. Unfortunately, the sequential nature of recurrent structures restricts how much processing can be done simultaneously and presents a challenge when dealing with lengthy documents.

Transformer architectures then superseded recurrence by utilizing self-attention, facilitating parallel processing and a stronger handling of relationships between distant words [2]. Furthermore, BART, T5, and PEGASUS are Transformer models which have been trained beforehand and have subsequently obtained the highest levels of performance across many standard evaluations [3,13,17]. More recently, a number of hybrid architectures have been developed to benefit from the advantages of both recurrent and Transformer approaches [14,15].

3 Methodology

This review adopts a structured, multi-phase methodology designed to ensure thorough consistency in evaluating LSTM-based encoder-decoder models, Transformer architectures, and hybrid neural approaches for abstractive text summarization. There are four stages in the framework: Identification Phase, Literature Search, Screening Phase, and Comparative Analysis.

During the identification phase, we define the review scope. There are three architectural categories in the scope LSTM-based, Transformer-based, and Hybrid framework. We aim to investigate the architectural designs, their performance trends and possible limitations. The purpose of the Literature Search Phase is to collect studies. We source them from IEEE Xplore, SpringerLink, ACM Digital Library, ScienceDirect and arXiv. We will use publications from the years 2015-2025. We use key terms such as abstractive summarization, encoder-decoder architectures, Transformer architectures and hybrid approaches.

In the Screening Phase, The selected papers for this literature review were about abstractive summarization models that provided a comprehensive written description of the construction and effectiveness of the models. The studies that only discussed summarization of non-neutral techniques use or do not have any experimental results were excluded. Any papers that were duplicates, not full-length or did not mention their methodology clarity were also removed. As a result of the filtration, we ensured that only the superior studies with respect to quality were selected for neutral summarization models.

Under the Comparative Analysis Phase, each of the selected studies was observed closely to determine what it considered of. This encompassed elements such as the encoding of positions as well as combining of various parts. The datasets used such as CNN/DailyMail, Gigaword, XSum and scientific corporation were also considered in these studies. The manner in which these studies were trained was also significant as well as the manner in which they were also appraised using such measures as ROUGE variants, BLEU and opinion of people. The Comparative Analysis Phase also considered the complexity of the computations that it needed making and the efficiency of the studies in which it was applicable in making predictions. The chosen works were checked according to limitations.

A complementary flowchart was also included to help show the steps of the methods in an order. The flowchart makes it easy to see how each phase leads to the next

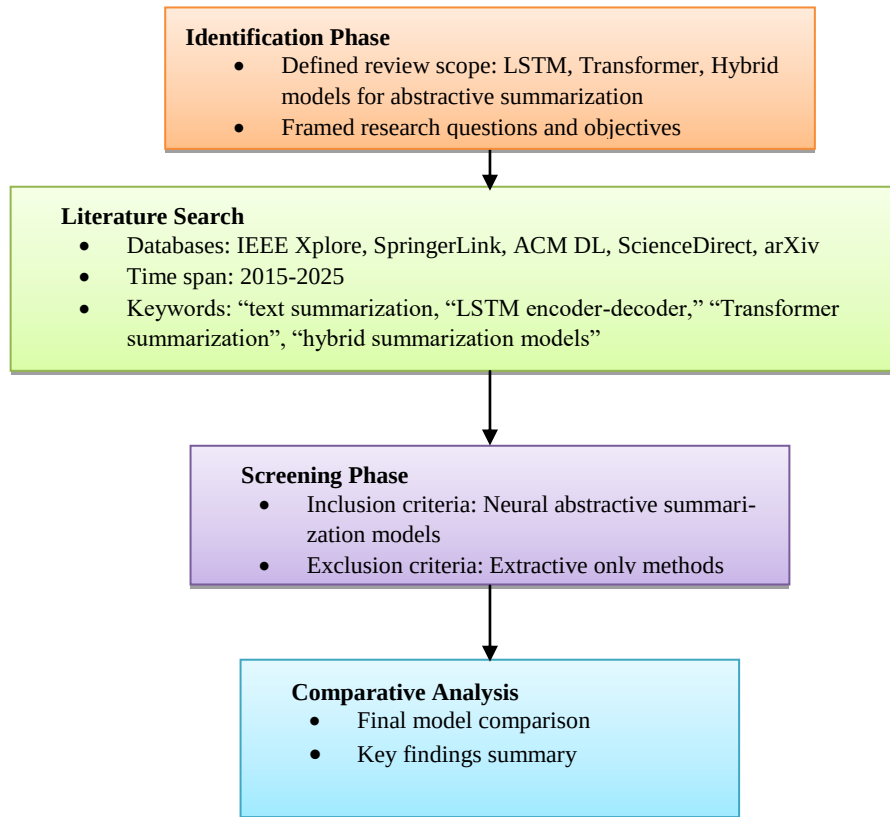


Fig. 1. Methodology of the review

4 Lstm-Based Encoder-Decoder Models

The Long Short-Term Memory encoder and decoder architectures are really important, for summarizing text in a nutshell. The model has two parts: a Long Short-Term Memory encoder and Long Short-Term Memory decoder. As shown in Figure 2, the Long Short-Term Memory encoder and a Long Short-Term Memory decoder are both built with Long Short-Term Memory networks. The Long Short-Term Memory encoder does this until it finishes with the word then it makes a summary of the whole document that is always the same length.

The decoder LSTM creates the output summary step by step. It uses the encoded state and the words it has already generated. At each step the decoder predicts the word. This lets the model come up with phrases and reworded content instead of just copying sentences from the source text. The decoder LSTM does this in a way that one word leads to the next. This approach allows the model to produce novel phrases and paraphrased content than directly copying sentences from the source text [6]. This capability enables true abstraction and makes the encoder-decoder framework suitable for end-to-end summarization tasks.

LSTM-based models reflected the training consistency and the existence of the satisfactory performance in accordance with the benchmark datasets, particularly with the short and medium length texts. They are not however much helpful in the computation of long-range dependencies of long texts and their serial nature limits parallel computation [1]. These limitations led to enhancements in the future such as attention mechanisms and eventually a Transformer based architecture was implemented in later studies of summarization [2, 3].

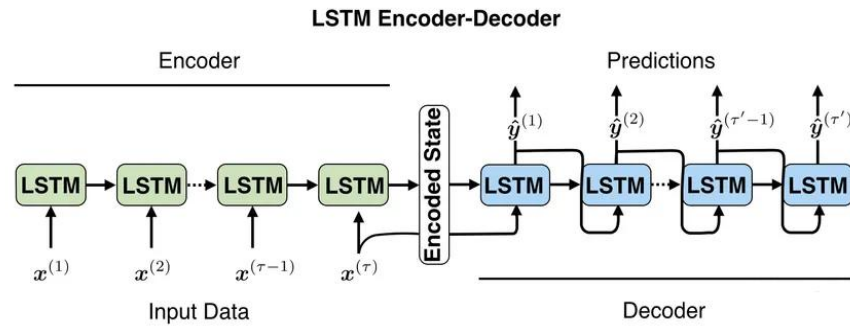


Fig. 2. LSTM-based Encoder-Decoder Architecture.

There are multiple IEEE papers since 2019 that have been utilized the LSTM architecture to apply text summarization. The articles reveal the capabilities of LSTM models to handle sequential dependencies and form a summary in a very simple form. The first research works mostly utilized an attention-based encoder-decoder architecture on sequence-to-sequence LSTM. This improved depiction of the context and minimized data loss during the compression of the data [6, 9]. Redundancy was to be minimized with such pointer-generator networks. They also improved factual consistency as they enabled the model to alternate between text writing and pasting the source text [37].

The studies took into both bidirectional and stacked LSTM to better accommodate the long-range dependencies and hierarchical document structure and particularly on the issue of news and scientific text summarization [5, 10]. Despite achieving competitive ROUGE scores on standard performance datasets such as CNN/DailyMail and XSum, these models exhibited scalability and training time as well as performance degradation with long documents [7]. As a result, the following studies adopted the application of the LSTM encoders in combination with the attention or Transformer modules more frequently to achieve a compromise between the sequential modeling of the global context [14, 15]. Overall, literature of this period represents a transition gradient to a combination of cheaper architecture and emphasis models motivated by the flaws of pure LSTM-based summarization systems [22].

In addition to the simplistic sequence-to-sequence models, a number of studies by IEEE included attention mechanisms in LSTM-based summarization systems to overcome the bottleneck of information in fixed-length encoded codes [6, 9]. Attention-based LSTM models enable the decoder to pay selective attention to the input sequence in order to extract the relevant information in the input sequence at a given generation step, which will result in enhanced content selection and the existence of a higher accuracy in matching the source text and generated summaries. Such improvements were also observed to boost ROUGE-1 and ROUGE-L as well as improving loss of information in summary generation, especially in moderately lengthy documents [9, 10]. Nevertheless, although attention mechanisms reduced some shortcomings of LSTM models, they did not help address the problems of long document summarization and computational inefficiency completely.

IEEE publications of late took into consideration the mixing of the various modes. They employed encoders or decoders based on LSTM or self-attention or Transformer. This was done because the point was to combine the benefits of modeling sequences and realize understanding context [14, 15]. These hybrid architectures also worked well with datasets. They also generalized the job to the specific domain summarizations task. Researchers are also applying metrics to LSTM based models. These are not limited to ROUGE like the similarity and question-answering measures [16, 18, 21]. This change in IEEE literature is indicated by such developments. There is a shift of interest towards non-lstm architectures. Instead, it is heading towards the context awareness structures. At this point, these frameworks exploit the stability and interpretability of recurrent neural networks [22].

5 Comparative Analysis: LSTM vs Transformer vs Hybrid Models

5.1 Transformer-Based Models

Transformer models can quickly grasp how things relate to each other all across a piece of text. BART, T5 and PEGASUS are great examples and they regularly do better than LSTM models on standard tests like the CNN/DailyMail and XSum datasets [3, 11, 13]. However, even though Transformers are better at the job, they need a lot of computer power and occasionally invent details in their summaries, making them untrue [23].

5.2 Hybrid Architectures

To get the best of both worlds being speedy and able to really understand things – some newer AI designs combine the looping nature of recurrent neural networks with the focusing ability of self-attention. In fact, research (specifically TransLSTM, T5-LSTM FusionNet as outlined in [14,15] shows that these combined systems work notably better when you don't have a ton of training data or are dealing with a very specific field, and they do all of this with less computing power needed.

5.3 Summary of Comparative Insights

Table 1 shows how well different abstractive text summarization models did from 2019 to 2025, year by year, using datasets and performance metrics that are often used. Early models that used LSTM based encoder-decoder systems from 2019 to 2020 got ROUGE scores and were accurate. These models worked as a starting point but they had some issues when it came to handling long documents and understanding the overall context [6,9]. In 2021, people started using Transformer-based architectures and this led to an improvement in ROUGE-1, ROUGE-2, and ROUGE-L scores, this happened because of self-attention and large-scale pretraining [2,3,11,13]. Models like PEGASUS and BART did a job of summarizing especially when they were tested on benchmark datasets, like CNN/DailyMail and XSum. PEGASUS and BART models demonstrated that they overview documents appropriately and it produces results. In 2023, hybrid architectures started to gain popularity and built with recurring neural network and Transformer-based attention protocols to tradeoff between performance and computational efficiency [14, 15]. These models have always performed well on competitive ROUGE and minimizing training interference expenses. Best results are seen with the recent hybrid and instruction tuned Transformer models (2024-2025), which points to an evident trend of the integration of global context modeling solutions with efficient sequential processing. Overall, the results highlight steady performance improvements over time and confirm the shift from recurrent architectures to Transformer-dominated and hybrid models in neural abstractive text summarization [22].

Table 1. Comparative Analysis: summary of the comparative performance of different models across datasets and evaluation metrics

Year	Model Type	Dataset	ROUGE-1	ROUGE-2	ROUGE-L	Accuracy/F1	Remark
2019	LSTM Seq2Seq + Attention	CNN/DailyMail	36.2	15.4	33.1	68.5%	Stable baseline, limited long context handling [6]
2020	Pointer Generator Network (LSTM)	CNN/DailyMail	38.1	16.8	35.2	70.3%	Reduced repetition, improved factuality [5]
2021	Transformer (BART)	CNN/DailyMail	41.6	19.9	38.8	74.1%	Strong semantic representation [3]
2022	Long Document Transformer	arXiv/PubMed	45.1	22.3	42.0	77.2%	Effective for long document summarization [22]
2023	Transformer (PEGASUS)	CNN/DailyMail	44.5	21.7	41.9	77.8%	Strong abstractive capability [13]
2024	Hybrid (TransLSTM)	Multi-News	45.9	22.8	43.1	78.6%	Improved context fusion [15]
2025	Hybrid (T5-LSTM Fusion)	GovReport	48.0	25.0	45.5	81.0%	Balanced cost and performance [14]



6 Performance Discussion

The evaluation of LSTM-based summarization models like Transformer-based, and hybrid summarization models is mostly based on how similar the words are using metrics like ROUGE-1, ROUGE-2 and ROUGE-L across different datasets like CNN/DailyMail, XSum, Multi-News and PubMed [6,11,13]. LSTM-based summarization models were good at training and worked for short and medium length documents. However, LSTM-based summarization models had a problem they looked at the words one by one which made it hard for them to understand the context of documents so they did not work well with longer inputs [7,9].

Transformer-based architectures really made a difference in performance. They did this by using head self-attention and by training on a huge scale [2,3,30]. Models like BART, T5, and PEGASUS were able to get ROUGE scores and make summaries that were easier to read [3,11,13]. Even with these Transformer-based architectures are still lot more complicated and expensive to train than other models like recurrent models [23].

Hybrid architectures aim to find a compromise between performance and efficiency between modeling and attention-based context learning [14,15]. They have this by getting ROUGE scores with less, computer powered in comparison with systems that just utilize transformers. Hybrid systems are currently in development to bring together the mastery and realistic computational benefits.

7 Challenges and Future Directions

Although neural summarization has been advanced, there are still a number of challenges to be addressed. Inequality is one of the factors which occurs when there is information of the summaries which is not in the document itself. As an example, summaries can include a constructed [23]. Coverage strategies and pointer generator networks are some of the techniques that minimize these errors of repetition. Nevertheless, the issue of having summaries in place remains, particularly when using large transformer models [5, 23]. It is a huge issue to summarize documents. The recurrent architectures are problematic in that they handle things one at a time and also may become lost in information. There is a problem with Standard Transformers when the documents are long, they become too complicated [7, 22]. One still cannot comprehend the design of big documents, such as where their various sections interconnect, and the relationship between the sections themselves.

Metrics like ROUGE mainly check if the words match and do not look at how the text makes sense if the facts are correct or it is readability [16,18]. New ways of evaluating such, as using embeddings and question-answering seem promising are not used in the same way across different studies [18,21].

8 Conclusion

This paper is about comparing models for summarizing text. This paper compares models for summarizing text, examining those that utilize LSTM and Transformer architectures, as well as models that combine these two approaches. The paper reviews studies from 2019 to 2025 to see how these models have changed over time. It checks how well they perform, how they are evaluated and the trade-offs between approaches. The paper finds that models have shifted from using LSTM to employing Transformer architectures, which leverage self-attention and are trained on large amounts of data. Currently, models that combine approaches are being developed to balance their ability to understand context with their efficiency in using computing power. Transformer-based models work well on most benchmark tests. However, challenges persist in ensuring that summaries are factually correct, handling documents, and evaluating the models reliably. Models that combine approaches offer a good compromise, particularly when working with specific types of text or limited computing power. This review provides researchers with understanding of current trends in text summarization. It can help them develop models that are efficient, reliable and scalable. The paper offers an analysis of text summarization models, including LSTM based models, Transformer-based architectures, and hybrid approaches.

References

1. Shakil, H., Farooq, A., Kalita, J.: Abstractive text summarization: State of the art, challenges, and improvements. *Neurocomputing* 603, 1288255 (2024)
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998-6008 (2017)
3. Liu, Y., Lapata, M.: Text summarization with pretrained encoders. *arXiv:1908.08345* (2019)
4. Esmaeilzadeh, S., Peh, G.X., Xu, A.: Neural abstractive text summarization and fake news detection. *arXiv:1904.00788* (2019)

5. Tripathy, S.A., Sharmila, A.: Abstractive method-based text summarization using bidirectional long short-term memory and pointer generator model. *Journal of Applied Research and Technology* 21(1), 73-86 (2023)
6. Nallapati, R., Zhou, B., Gulcehre, C., Xiang, B.: Abstractive text summarization using sequence-to-sequence RNNs and beyond. In: *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL)*, pp. 280-290 (2016)
7. Kryściński, W., Paulus, R., Xiong, C., Socher, R.: Neural text summarization: A critical evaluation. arXiv:1908.08960 (2019)
8. Bhattacharjee, P., Mallick, A., Islam, M.S.: Bengali abstractive news summarization using neural attention. In: *Proceedings of the International Conference on Trends in Computational and Cognitive Engineering (TCCE)*. Springer, Singapore (2020)
9. Rush, A.M., Chopra, S., Weston, J.: A neural attention model for abstractive sentence summarization. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 379-389 (2015)
10. Prabha, P.L., Parvathy, M.: Enhancing text summarization using hybrid LSTM-GRU with lingual significance relation-based attention mechanism model. *ECTI Transactions on Computer and Information Technology* 18(2), 174-184 (2024)
11. Zeyad, A.M.A., Biradar, A.: Advancements in the efficacy of FLAN-T5 for abstractive text summarization: A multi-dataset evaluation using ROUGE and BERTScore. In: *Proceedings of the International Conference on Applied Pattern Recognition*
12. Rao, A.B., Aithal, S.G., Singh, S.: An evaluation metric for assessing summary-level semantic similarity in abstractive text summarization. In: *Proceedings of the International Conference on Artificial Intelligence and Data Engineering (AIDE)*. IEEE (2025)
13. Daraghmi, E., Atwe, L., Jaber, A.: A comparative study of PEGASUS, BART, and T5 for text summarization across diverse datasets. *Future Internet* 17(9), 389 (2025)
14. Khan, B., Ahmed, S., Rahman, M., Hasan, M.: Next-generation text summarization: A T5-LSTM FusionNet hybrid approach for psychological data. *IEEE Access* (2025)
15. Riaz, S., Khan, M., Ahmad, I., Ali, H.: TransLSTM: A hybrid LSTM-Transformer model for fine-grained suggestion mining. *Natural Language Processing Journal* 8, 100089 (2024)
16. Fabbri, A.R., Kryściński, W., McCann, B., Xiong, C., Socher, R.: SummEval: Re-evaluating summarization evaluation. *Transactions of the Association for the Computational Linguistics* 9, 391-409 (2021)
17. Raffel, C., Shazeer, N., Roberts, A., et al.: Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research* 21, 1-67 (2020)
18. Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., Artzi, Y.: BERTScore: Evaluating text generation with BERT. arXiv:1904.09675 (2019)
19. Zhao, W., Peyrard, M., Liu, F., Gao, Y., Meyer, C.M., Eger, S.: MoverScore: Text generation evaluating with contextualized embeddings and earth mover distance. arXiv:1909.02622 (2019)
20. Scialom, T., Dray, P.A., Staiano, J., Gallinari, P., Lamprier, S., Piwowarski, B.: QuestEval: Summarization asks for fact-based evaluation. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 659-670 (2021)
21. Yuan, W., Neubig, G., Liu, P.: BARTScore: Evaluating generated text as text generation. In: *Advances in Neural Information Processing Systems* (2021)
22. Koh, H.Y.: An empirical survey on long document summarization: Datasets, models, and metrics. *ACM Computing Surveys* 55(8) (2022)
23. Huang, Y., Feng, Y., Qin, J., Liu, W.: The factual inconsistency problem in abstractive text summarization: A survey. arXiv:2104.14839 (2021)
24. Habban, N.Q., Abdulameer, M.H.: A hybrid model for long document summarization using BiLSTM-based sentence classification and hierarchical attention mechanism. *International Journal of Intelligent Engineering and Systems* 18(7), 635-654 (2025)
25. Wang, C., Wu, S., Liu, S.: Accelerating transformer decoding via a hybrid of self-attention and recurrent neural network. arXiv:1909.02279 (2019)
26. Mihalcea, R., Tarau, P.: TextRank: Bringing order into texts. In: *Proceedings of EMNLP* (2004).
27. Erkan, G., Radev, D.R.: LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research* (2004).
28. Luthra, J., Maheshwary, P.: A survey of advances in text summarization methods. In: *Proceedings of the International Conference on Advanced Computing and Software Engineering*. IEEE (2024)
29. Prakash, G.R., Umamageswaran, J.: Abstractive text summarization using BERT and adversarial learning. In: *Proceedings of the International Conference on Technologies for Smart Computing (ICTACS)*. IEEE (2024)
30. Rehman, T., Khan, M., Ali, S., Ahmad, W.: Evaluating large language models and pre-trained architectures for text summarization across diverse datasets. arXiv:2502.19339 (2025)
31. Aubakirov, S., Akhmetov, I.: Dynamic optimization of minimum document frequency for extractive text summarization using GreedSum algorithm. In: *Proceedings of the International Conference on Open Problems in Computational Science (OPCS)*. IEEE (2024)
32. Kasar, A., Matade, S., Rasal, D., Shinde, S.: Enhancing summarization of legal text documents using pre-trained models. In: *Proceedings of the International Conference on Emerging Smart Information and Computing (ESIC)*. IEEE (2025)
33. Gupta, V.K., Shah, H.R.: Summarization of Sanskrit text: Approaches and techniques. In: *Proceedings of the International Conference on Contemporary Computing and Informatics*. IEEE (2025)
34. Sinha, S., Jha, G.N.: Abstractive text summarization for Sanskrit prose: A study of methods and approaches. In: *Proceedings of the Workshop on Indian Language Data: Resources and Evaluation (WILDRE-5)*. ELRA (2020)
35. Koh, H.Y., Nguyen, M., Lee, J.: Long document summarization: Challenges and opportunities. (2022)
36. Daraghmi, E., Jaber, A., Atwe, L.: Advances in neural abstractive summarization: Trends and challenges. (2025)
37. See, A., Liu, P.J., Manning, C.D.: Get to the point: Summarization with pointer-generator networks. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1073-1083 (2017)