Text Summarization using different Methods for Deep Learning

Fatima Shaker[1], *Mustafa* M. Abd Zaid[2*], *Ahmed* Ali Talib Al-khazaali[3] and *Zahraa* Ali Al-khzaali[4]

[1]Islamic University, najaf,Iraq.
[2]Islamic University, najaf,Iraq.
[3]University of Alkafeelty of Alkafeel, najaf,Iraq.
[4]Altinbas University, Istanbul.

Abstract. In the current era of rapid information expansion, text summarization has become vital for comprehending textual material. Physically condensing large textual volumes is challenging for humans, especially considering the vast amount of text content available online. Text summarization is an active field of research that focuses on summarizing large texts into shorter versions while retaining important information. The writing can be categorized as either extractive or abstractive regarding its summary. Extractive summarizing methods function by determining the significance of individual sentences within a text and selecting them to form a summary. This approach relies solely on sentences extracted from the original text. Abstractive summarization methods aim to rephrase significant information. Text summarization can be achieved by many deep learning methodologies, including fuzzy logic, Convolutional Neural Networks (CNN), transformers, neural networks, and reinforcement learning. Over the past three years, there has been a modest shift in the research focus on text summarizing. Current developments aim to increase the efficiency of text summarization and attain optimal accuracy. This study aims to examine the various methods of using deep learning for text summarization and identify the current deep learning developments.

1 Introduction

Historically, individuals have manually condensed textual documents for an extended duration. Nevertheless, the huge quantity of data available on the internet, including articles, news, blogs, and large written material, is causing consumers to struggle with effectively managing the vast amount of information. Text summary is essential for tackling this issue and has acquired importance.

The significance of a text summary lies in its ability to efficiently extract crucial information from a lengthy text within a short timeframe. It enables rapid and effortless access to the most significant facts while addressing any concerns about the needs. Essential for the summary evaluation [1]. Summary of the text The extracted data is gathered and organized into a succinct report by selecting the pertinent and necessary information from the original text [2]. The text summary may be classified as either Extraction or Abstract summarization. An extractive summary is a summary that selects and organises the essential sentences from a text in a predetermined sequence [2]. The user's text is [3]. Extractive summarization involves identifying sentence properties, assigning scores to these sentences, and generating a summary based on these values. The outcome of this summary is based on the qualities gathered from the phrases of the document [4]. The extracted summary has a primary limitation. Regardless of its major relevance, when a sentence is picked for the summary, it will include all the information and knowledge it contains [3].

*Corresponding author: mustafamajeed2014@gmail.com

Abstract summarization aims to generate new sentences by initially constructing a summary, which is then categorized into two classes: semantic-based approaches and structured-based methods. A significant limitation of abstractive summarization is its inherent issue of relevance, where the intended meanings of the newly generated sentences may not align with those intended in the original document text. Occasionally, the summary generated may select less pertinent secondary information from the document. Abstractive summarization is primarily used for small documents, making it difficult to apply these techniques to larger texts [3] [2].

This essay sought to study the many text summarization strategies utilized in recent years. To conduct a comprehensive study, many researchers were engaged to provide a concise analysis of the most effective and preferred ways, along with the underlying reasons.

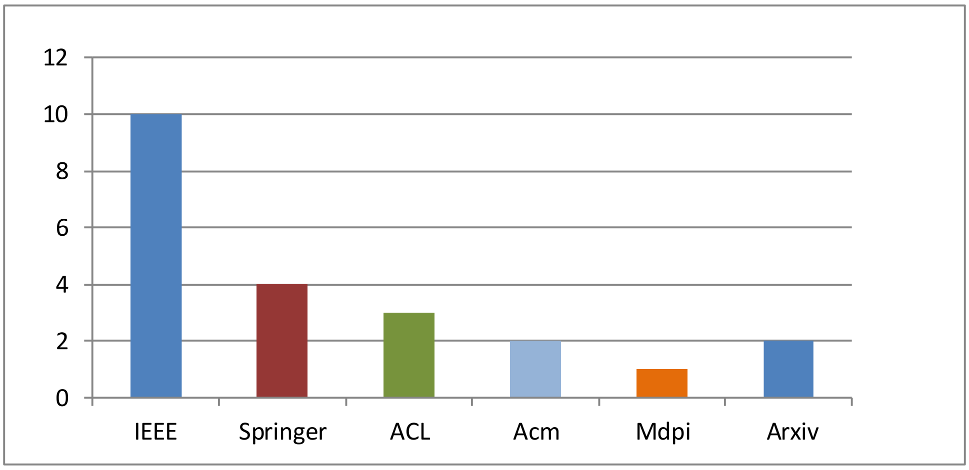

Fig. 1. A section of scholarly articles obtained from different websites**.**

Figure 1 explains the source from which The papers that are to be analyzed have been chosen.. Research articles from reputable sources such as Springer, IEEE, MDBI, ACL, arXiv, and ACM library have been collected.

2 Related work

Nikhil and Samidha[2] have introduced a methodology for summarising text by extracting important information. This work focuses on a single document and combines two processes to select important sentences and generate a meaningful summary. The techniques employed for sentence selection from the text include the Restricted Boltzmann Machine and Fuzzy Logic. Initially, fuzzy logic and the restricted (Boltzmann) machine provided two summaries for each separate document. Subsequently, the outcomes of both summaries are merged and subjected to distinct processes to generate the last summary from the given document. The outcome demonstrated that the recommended design technique effectively addresses the issue of document overload by offering a brief and effective summary. This work utilized ten English-language documents from a news item dataset that was provided in kaggle. In the method provided, the ROUG matrices achieved an F-measure of 0.84, precision of 0.88, and recall of 0.80 [2].

Shengli Song et al[5]. introduced the ATSDL (Abstract et al.) system that uses LSTM-CNN to produce new phrases by analyzing smaller segments in more detail, including conceptual words. The ATSDL process has two parts. The first part involves extracting sentences from the original text. Within this step, three sub-processes are employed for extraction of phrase: phrase acquisition, refining phrases, and combining phrases. The second step involves generating text summaries using the LSTM-CNN model. The experimental results for the ROUGE metric show that the R-1 score is 34.9 and the R-2 score is 17.8 for both the CNN and daily mail datasets [5].

Afsaneh et al.[6] suggested utilizing deep learning techniques, specifically a deep neural auto-encoder and deep belief network, to produce an extractive multi-document summary. The construction of these frameworks is based on the DUC2007 dataset, and their validation is performed using the ROUGE metrics. After normalizing the data and feature extraction step, the feature-sentence matrix was inputted into the neural networks to estimate the sentences value by using network output scores. Overall, the autoencoder network exhibits superior performance compared to the DBN. The autoencoder achieves a R-2 score of 0.088, whereas the DBN has a mean R-2 score of 0.0899, indicating superior performance compared to other systems. In addition, the R-1 values derived from these systems are suitable and precise [6].

Wen and Giuseppe [7] have presented a novel neural single-document extraction summary model for long documents. This model combines the current topic's local scope with the entire document's global reach. The dataset used for this study consists of scientific publications from arXiv and PubMed. The model was then evaluated and compared to previous research using both abstractive and extractive models based on R-1, R-2, and METEOR scores. The initial model for text summarization is LSTM-minus, a variant of embedded learned text spans. This proposed model incorporates fundamental components. A sentence encoder, document encoder, and sentence classifier are used in the sentence encoder. The sentence encoder's primary function is transforming word embedding sequences into a fixed-length vector. Document Encoder generates its output using bi-directional gated recurrent units (GRU). Each statement consists of two hidden states: one forward and one backward. Once the sentence representation, topic segment, and document have been retrieved, three factors are combined to make a final prediction, pi, on whether the sentence should be included in the text summary. The arXiv dataset was evaluated using the attentive context model, which achieved the following results: R-1: 43.58, R-2: 17.37, R-L: 29.30, and METEOR: 21.71. The PubMed dataset was evaluated using the attentive context model, with the results as follows: R-1: 44.81, R-2: 19.74, R-L: 31.48, and METEOR: 20.83 [7].

A model has been suggested by Rajeev Kumar Singh et al.[8] that aims to summarise news items and generate precise and accurate headlines. This new method, known as SHEG, was introduced by researchers. This approach has surpassed existing frameworks and is a unique model that generates a concise summary and a relevant headline. The CNN/Daily Mail, Gigaword, and Newsroom datasets were utilized to learn, verify, and validate this model. The extractive mechanism employs Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to execute attention at both the word-level and sentence-level. The pointer-generator network is trained using an innovative controlled actor-critical (CAC) model. This training balances variance and bias in the reinforced abstractive mechanism [8].

Table 1 provides a brief overview of the studied articles and their publishing year, the methods used, the datasets utilized, notes, and the rough metric.

Table 1. Synopsis of an Analyzed Articles

Reference	Year	Methods Used	Dataset	Notes	ROUGE matrices
[9]	2020	memory network and non-local network.	CNN/ Daily Mail .	The proposed model exhibits greater efficacy than the baseline model.	R-1=5.6%, R-2=5.3%, and R-L=6.2%.

https://doi.org/10.1051/bioconf/20249700074

[10]	2020	Proposed an HH-ATS	CNN/ Daily Mail and Gigaword	the proposed model outperforms the baseline approaches of the current state-of-the-art in terms of ROUGE matrix scores.	The CNN/Daily Mail dataset achieved a R-1= 43.16, R-2 20.32, and a R-L =39.14. On the Gigaword dataset, the R-1 =38.43, the R-2 = 19.75, and the R-L= 36.11.
[11]	2020	The BERT model and K-means clustering.	CNN/ Daily Mail	The current methodology's negative aspect is that a few sentences cannot fully summarise the document's context.	R-1(F1) = 41.4, R-2(F1) = 17.9, and The result of the subtraction of F1 from R-L =37.9.
[12]	2020	BERT model	CNN/ Highlights News.	Providing a better summary for a topic with high overlap in the news has been seen. Real-time information is used by the device to produce dynamic results.	R-1=41.72, R-2=19.39, and R-L = 38.76.
[13]	2020	BERTSUMEXT	Data in the textual clinical	improvement all nine ROUGE criteria.	Recall=2.09, precision=0.8, and F measure = 1.44.

[8]	2020	Word embedded, CAC ,(GRU) RNNs to mechanism they use Bi- LSTM encoder	NEWSROOM , CNN and Gigaword	-heart models provide an abstract summary and headline.	NEWSROOM for SHEGE Abstract summary outcomes to abstractive and extractive is R-1=28.2, R-2=14.7, and R-L= 25.9 Results on NEWSROOM headline R-1=13.81, R-2=7.94, and R-L= 11.42 Comparative outcomes of various headline creation models on Gigaword for R-1=31.82, R-2=13.2, and R-L=28.80.

 https://doi.org/10.1051/bioconf/20249700074

					The CNN/Daily Mail presented a comparison of various headline creation framework models with R-1=40.67, R-2=17.74, and R-L=36.69. Result Comparative of different extractive frameworks on CNN/ Daily Mail for R-1=42.5, R-2=17.6, and R-L=35.6.
[14]	2019	Sequence two Sequence model. Convolutional Neural Networks model.	DUC and GigaWord	In addition, they use a copy technique to remove out-of-vocabulary (OOV) terms from the original text.	DUC corpus to R-1=29.74, R-2=9.85, and R-L=25.81 Gigaword to R-1=37.95, R-2=18.64, R-L=35.11.
[6]	2019	Neural Network Autoencoder and Deep Belief Network(DBN).	DUC2007 Dataset	Overall, the autoencoder network exhibits superior performance when pitted against DBN.	For deep belief network, the result was(R-1 = 0.391 and R-2 = 0.089). while the autoencoder was (R-1=0.398 and R-2=0.092).
[7].	2019	LSTM-minus, and (GRU)	arXiv and PubMed	It takes into account both the immediate context within the current subject and the broader context of the whole document. It is used for extensive documents.	For the arXiv adatabse using attentive context model are R-1=43.58, R-2=17.37, R-L=29.30, and METEOR=21.71. The concat model achieves scores of R-1=43.62, R-2=17.36, R-L=29.14, and METEOR =21.78. For PubMed dataset using the attentive context model are as follows: R-1= 44.81, R-2 =19.74, R-L= 31.48, and METEOR = 20.83. The concat model achieved an R-1=44.85, R-2 = 19.70, R-L =31.43, and METEOR = 20.83.
[15].	2019	It used LSTM and Gated Recurrent Units based RNN.	Bengali news	The suggested model has certain limitations, such as the need for a substantial quantity of training data, lengthy training time, and expensive hardware expenses.	average F1 scores is - for R-1=0.63, R-2=0.59 and R-3=0.56.

[16].	2019	Seq2Seq model equipped with attention mechanisms, Scheduled-Sampling.	CNN/ Daily Mail , Arabic news andSaudi newspapers	Results of this study demonstrate the feasibility of applying identical structures in different languages by describing the dataset using word-embedding techniques.	Regarding the English dataset are R-1=6.48, R-2=0.76,and R-L 05.31. The Pointer-Generator model achieved scores of R-1=30.84, R-2 9.82, and R-L=21.39. The Scheduled-Sampling technique is R-1=34.18, R-2=12.47, and R-L= 23.68 respectively.
[5].	2018	LSTM-CNN	Daily Mailand CNN	There are two processes in ATSDL: the first is to extract sentences from the source phrases. 2. uses deep learning approaches to generate text summaries.	Applying the Policy-Gradient algorithm, R-1=, 33.26, R-2=, 10.97, and R-3=, 23.28.

[4].	2018	CNN. The phrase selection is determined by the ranking of sentences using the ILP technique.	DUC2007 Dataset	This model has been developed to attain competitive efficacy when compared to state-of-the-art approaches, while having a relatively basic convolutional neural network (CNN) architecture.	The results show that for mv-cnn, the R-1= 40.92 and the R-2 = 9.11. While CNN-ILP, R-1 = 39.68 and R-2 = 10.26.

[17].	2018	(SCST) for Seq2seq model	the Gigaword , DUC-2004, and LCSTS Datasets	The suggested model outperforms state-of-the-art techniques on several benchmark datasets.Moreover, this framework has the capability to produce summaries that exhibit enhanced structure, logical flow, and a wider range of content.	High ROUGE scores for datasets 1- Convert Gigaword (R-1=36.92 , R-2=18.29 , and R-L=34.58). The internal test set of Gigaword achieved R-1=46.92, R-2=24.83, and R-L=44.04. 3- DUC-2004 (R-1= 31.15, R-2=10.87, and R-L=27.68). 4- LCSTS (R-1=39.93/45.12 R-2=21.58/33.08 and R-L=37.92/42.68), The left side represents word-level ROUGE, while the right side represents character-level ROUGE.
[18]	2019	Extractive summarization using deep learning approaches for multi-document queries.	DUC2005, DUC 2007	A comparison was conducted on six deep learning techniques, revealing that Bi-LSTM emerged as the most superior option.	High degree for LSTM is (R-1=43.53, R-2= 11.40 and R-L= 18.67)
[19]	2018	Graphs-based	DUC-2002	The main contribution of this work is the investigation of the impact of a well-established algorithm for social network analysis, which allows for reliable analysis of enormous networks.	The Euclidean method's score is (0.3833 For R-1= 0.5581)
[20]	2023	Reinforcement Learning	DUC-2004 CNN/ DailyMail	Suggested an abstractive summarization model using the joint-attention process and pre-existing information.	ROUGE For datasets 1- CNN/DailyMail (R-1=34.11 , R-2=12.78 , and R-L=22.5). 2- DUC-2004(R-1=26.55 , R-2=7.06 , and R-L=22.05).

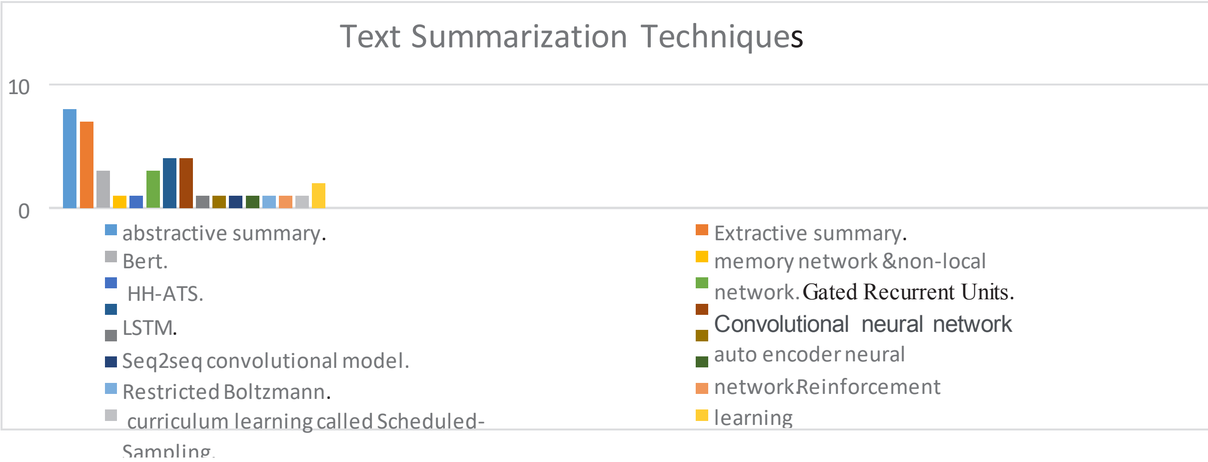

Fig. 2. Various techniques that the research articles employ

In Figure (2) appropriately describes the spectrum of summary text approaches employed in the 16 examined publications, including Boltzmann Machine , Gated Recurrent Units., Fuzzy Logic, and reinforcement learning.

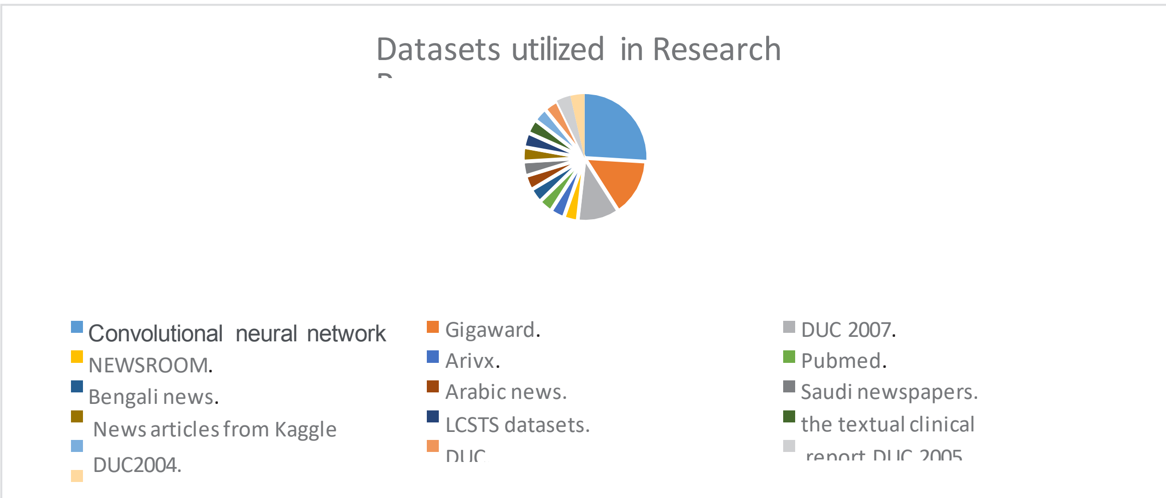

Fig .3. Various of the datasets that is used in papers

Figure 3 illustrates the frequency of datasets used in the analyzed papers. Scientists have used data from many sources, including internet networks, social media postings, reputable news outlets such as the Daily Mail CNN, and datasets like Gigaword and DUC 7. Additional information may be obtained by referring to Figure 3 .

3 The fundamental structure of the text summary system

Pre-processing, identifying the most crucial information, and concatenating the information for summary production are the fundamental procedures that all ATS systems rely on. Pre-processing may differ from solution to solution, although it often includes some of the same processes that are widely used in other NLP applications [21][22].

1. Sentencization: The division of the document into individual sentences.

2. Tokenization is the act of dividing the text into a list of terms, standardising terms (i.e., placing all words in lowercase and removing accents), and eliminating unwanted information from the text (such as commas, hyphens, periods, HTML elements, etc.).

3. Elimination of stopwords: The process of getting rid of words that are widely used in any language but don't contribute much to the text, such conjunctions, articles, prepositions, and pronouns.

4. The removal of low-frequency terms: This involves getting rid of uncommon or misspelt words.

5 .Stemming: While stemming removes a word's end or beginning, it also considers a list of frequently occurring prefixes and suffixes that are present in inflected words.

4 Text Summarization Operations

Text summary operations spilt into one of the following categories:

1- Single-phrase operations

2- Operations on several sentences. One sentence may be the subject of a single-sentence operation, and at least two sentences can be the subject of a multi-sentence operation[23].

Text summarization operations can also be classified as either:

1- A single, atomic operation that cannot be split up into many actions (like word insertion and deletion).

2-A complicated process that can be broken down into smaller operations (word substitution, word reordering, sentence merging, etc.).

Figure 4 displays the text summary operations categorized by number of sentences or single sentences. A source document can be transformed into a summary document by using certain methods alone, sequentially, or in parallel. [24]

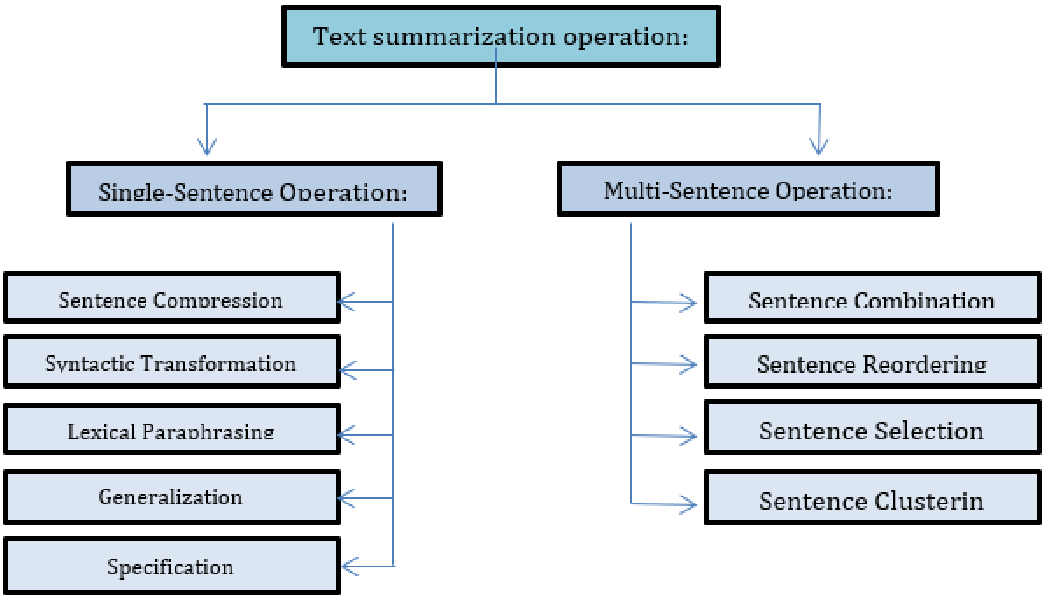

Fig. 4. Text summarization operations

5 CONCLUSION AND FUTURE WORK

The importance of the text-summarizing process has grown in recent years, mostly because of the vast volume of information accessible on the web. The examination of the approaches resulted in a variety of results. The Recursive Neural Network (RNN) was the most often used deep learning strategy. Several methodologies used Long Short-Term Memory (LSTM) to address the issue of gradient vanishing, but other methodologies utilized transformers and GRU. Abstractive summarization has also been accomplished using the sequence-to-sequence paradigm. This study presents many ways for generating summary text, including transformers such as Fuzzy Logic, BERT, Restricted Boltzmann Machine, LSTM-CNN, Word Embedding, and Recursive Neural Network(RNN). The Efficiency of the summary has been evaluated using R-1, R-2, and R-L. Fig 2 illustrates the various deep learning methodologies, including extractive and abstract methods. Table 1 demonstrates that the most favorable outcomes for extractive and abstract summaries are achieved when employing transformers. These transformers effectively handle challenges associated with lengthy documents and improve the accuracy of the summarization process. Over the last years, the extractive approach has remained highly sought after due to its simplicity compared to the abstract method.

Additionally, the capacity to mix methods is still accessible, resulting in an appealing outcome. In this essay, we examine the different text summarization patterns. The data received from 16 research publications were collated and presented tabularly, as shown in Fig 1. By analyzing the findings of the reviewed publications, it is feasible to provide further recommendations in the future and expand the rang of future studies to enhance text summarization techniques and improve its effectiveness.

References

1. A. Bader, A. M. El, and B. Menai, “Automatic Arabic text summarization : a survey,” Artif. Intell. Rev., 2015, doi: 10.1007/s10462-015-9442-x.
2. N. S. Shirwandkar and S. Kulkarni, “Extractive Text Summarization Using Deep Learning,” Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018, pp. 1–5, 2018, doi: 10.1109/ICCUBEA.2018.8697465.
3. S. K. K and S. Mathew, “SURVEY OF SCIENTIFIC DOCUMENT,” vol. 21, no. 2, pp. 141–177, 2020.
4. S. Charitha, N. B. Chittaragi, and S. G. Koolagudi, “Extractive Document Summarization Using a Supervised Learning Approach,” 2018 IEEE Distrib. Comput. VLSI, Electr. Circuits Robot. Discov. 2018 - Proc., pp. 7–12, 2019, doi: 10.1109/DISCOVER.2018.8674133.

5. S. Song, H. Huang, and T. Ruan, "Abstractive text summarization using LSTM-CNN based deep learning," Multimed. Tools Appl., vol. 78, no. 1, pp.857–875, 2019, doi: 10.1007/s11042-018-5749-3.
6. A. Rezaei, S. Dami, and P. Daneshjoo, "Multi-Document Extractive Text Summarization via Deep Learning Approach," 2019 IEEE 5th Conf.
7. Knowl. Based Eng. Innov. KBEI 2019, pp. 680–685, 2019, doi: 10.1109/KBEI.2019.8735084.
8. W. Xiao and G. Carenini, "Extractive summarization of long documents by combining global and local context," EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf., pp. 3011–3021, 2020, doi: 10.18653/v1/d19-1298.
9. R. K. Singh, S. Khetarpaul, R. Gorantla, and S. G. Allada, "SHEG: summarization and headline generation of news articles using deep learning,"Neural Comput. Appl., 2020, doi: 10.1007/s00521-020-05188-9.
10. Z. Hao and B. Xue, "2020 5th Asia-Pacific Conference on Intelligent Robot Systems, ACIRS 2020," 2020 5th Asia-Pacific Conf. Intell. Robot Syst. ACIRS 2020, pp. 163–167, 2020.
11. M. Yang, C. Li, Y. Shen, Q. Wu, Z. Zhao, and X. Chen, "Hierarchical Human-Like Deep Neural Networks for Abstractive Text Summarization," IEEE Trans. Neural Networks Learn. Syst., pp. 1–14, 2020, doi: 10.1109/tnnls.2020.3008037.
12. A. Srikanth, A. S. Umasankar, S. Thanu, and S. J. Nirmala, "Extractive text summarization using dynamic clustering and co-reference on BERT,"
13. Proc. 2020 Int. Conf. Comput. Commun. Secur. ICCCS 2020, pp. 0–4, 2020, doi: 10.1109/ICCCS49678.2020.9277220.
14. M. Ramina, N. Darnay, C. Ludbe, and A. Dhruv, "Topic level summary generation using BERT induced Abstractive Summarization Model," Proc.
15. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020, no. Iciccs, pp. 747–752, 2020, doi: 10.1109/ICICCS48265.2020.9120997.
16. P. Vinod, S. Safar, D. Mathew, P. Venugopal, L. M. Joly, and J. George, "Fine-tuning the BERTSUMEXT model for clinical report summarization,"
17. 2020 Int. Conf. Emerg. Technol. INCET 2020, pp. 1–7, 2020, doi: 10.1109/INCET49848.2020.9154087.
18. Y. Zhang, D. Li, Y. Wang, and Y. Fang, "applied sciences Abstract Text Summarization with a Convolutional Seq2seq Model," 2019, doi: 10.3390/app9081665.
19. A. Al Munzir, "Text analysis for Bengali Text Summarization using Deep Learning," 2019 10th Int. Conf. Comput. Commun. Netw. Technol., pp. 1– 6, 2019.
20. A. M. Zaki, M. I. Khalil, and H. M. Abbas, "Deep architectures for abstractive text summarization in multiple languages," Proc. - ICCES 2019 2019 14th Int. Conf. Comput. Eng. Syst., pp. 22–27, 2019, doi: 10.1109/ICCES48960.2019.9068171.
21. L. Wang, J. Yao, Y. Tao, L. Zhong, W. Liu, and Q. Du, "A reinforced topic-aware convolutional sequence-to-sequence model for abstractive textsummarization," arXiv, pp. 4453–4460, 2018.
22. S. Yao, "A Comparative Study of Deep Learning Approaches for Query-Focused Extractive," 2019 IEEE 2nd Int. Conf. Inf. Comput. Technol., pp.153–157, 2019.
23. G. L. De La Peña Sarracén and P. Rosso, "Automatic text summarization based on betweenness centrality," ACM Int. Conf. Proceeding Ser., vol. Part F137707, 2018, doi: 10.1145/3230599.3230611.
24. Li, Y.; Huang, Y.; Huang, W.; Yu, J.; Huang, Z. An Abstractive Summarization Model Based on Joint-Attention Mechanism and a Priori Knowledge. Appl. Sci. **2023**, 13, 4610. https://doi.org/10.3390/app13074610
25. Denny MJ, Spirling A. Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Political Analysis*. 2018;26(2):168-189. doi:10.1017/pan.2017.44
26. Cajueiro, D. O., Nery, A. G., Tavares, I., De Melo, M. K., Reis, S. A. D., Weigang, L., & Celestino, V. R. (2023). A comprehensive review of automatic text summarization techniques: method, data, evaluation and coding. *arXiv preprint arXiv:2301.03403*.
27. Belkebir, R., & Guessoum, A. (2018). TALAA-ATSF: a global operation-based arabic text summarization framework. *Intelligent Natural Language Processing: Trends and Applications*, 435-459.
28. El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). Automatic text summarization: A comprehensive survey. *Expert systems with applications*, *165*, 113679.