ijamrsd.com
ISSN: 2583-6404
Journal of Advanced Multidisciplinary Research Studies and Development
Peer Reviewed Referred Journal | Bi-monthly | Open Access

A COMPREHENSIVE REVIEW OF AUTOMATIC TEXT SUMMARIZATION TECHNIQUES FOR NON-INDIAN LANGUAGES

Mahesh S. Gaikwad[1], Dr. Bhausaheb V. Pawar[2], Dr. Rupali H. Patil[3]
[1] *Research Scholar, S.S.V.P.S's L.K. Dr. P.R. Ghogrey Science College, Deopur, Dhule, Maharashtra, India.*
Email: maheshgaikwad13@gmail.com
[2] *Professor, Department of Computer Science, KCE's Institute of Management & Research, Jalgaon, Maharashtra, India.*
[3] *Professor, Department of Computer Science, S.S.V.P.S's L.K. Dr. P.R. Ghogrey Science College, Deopur, Dhule, Maharashtra, India.*
Email: rupalipatilh@gmail.com

Abstract
The enormous and continuously growing volume of data available on the Internet makes it extremely difficult to manually summarize information on a single topic in a few lines. Therefore, an efficient technique is required to condense textual content while preserving its original meaning. Automatic Text Summarization (ATS) has emerged as a solution to this problem. This paper presents a review of previous research on text summarization across several non-Indian languages, including English, Arabic, Italian, French, Chinese, and Swedish. It also examines three major approaches to text summarization: extractive, abstractive, and hybrid. Furthermore, the study discusses specific methods and techniques used for each type of text summarization. This review shows that both extractive and abstractive text summarization techniques are effective, but hybrid approaches especially sequence-to-sequence models with attention and copy mechanisms achieves the highest performance.

Keywords: Text Summarization Approaches, Extractive Summarization, Abstractive Summarization, Hybrid summarization, Text summarization models.

► *Corresponding Author: Mahesh S. Gaikwad*

Introduction

Information and data are constantly increasing on the internet. A method is required to reduce the text while maintaining the original meaning. Simplifying a text into a shorter form without compromising its essential information is called automatic text summarization. The primary goal of text summarization is to help readers find important information easily and save time and effort when reading large documents. Text summarization offers multiple advantages, including decreased reading time, improved and simplified content selection, and greater search accuracy. The aim is to create a clear and easy-to-understand summary that highlights the main points of the document. There has been a great deal of research on automatic text summarization for various languages.

The three primary categories of text summarization are extractive, abstractive and hybrid. Extractive summarization involves selecting important phrases, sentences, or features directly from the original document and combining them to create a concise and coherent summary. The

abstractive text summarization method generates new, semantically meaningful phrases. It can create summaries by modifying or adding words that do not appear in the original document. Abstractive approaches are generally more complex. This technique aims to produce accurate and meaningful summaries by first understanding the content's meaning and then expressing it using appropriate words and phrases. Hybrid text summarization combines both extractive and abstractive methods [1].

Text Summarization:

To generate a summary from a text, it is often necessary to identify and select the most important words, phrases, or sentences. The various approaches and methods of text summarization are described below.

Approaches and Methods of Text Summarization:

Following Figure-1 shows that text summarization approaches are mainly classified into Extractive, Abstractive, and Hybrid methods.

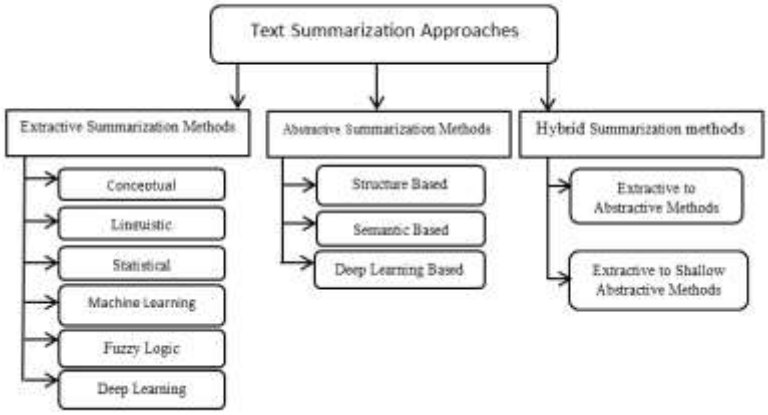

Figure 1: Text Summarization approaches and methods

Text summarization approaches are classified into extractive, abstractive, and hybrid text summarization approaches as follows:

Extractive Text Summarization:

Extractive summarization involves selecting the most important words from a source document and combining them to form a summary. This extraction process follows specific criteria, and no modifications are made to the original text.

The figure-2 illustrates the general workflow of an extractive text summarization process. First, the input document undergoes preprocessing to clean and prepare the text. Then, during the processing stage, sentence representations are created, each sentence is scored, and the most important sentences are selected. Finally, post-processing is performed to organize the selected sentences, resulting in the generation of the final summary.

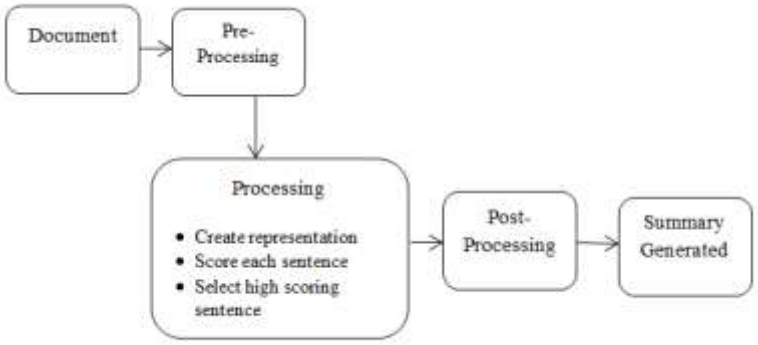

Figure 2: Extractive Text Summarization Process

Different extractive summarization methods are available, as shown in Figure 1. These methods such as -

1) Concept Methods

Concept-based summarization focuses on extracting the main ideas and concepts from a document rather than relying solely on individual words. This method uses external knowledge sources, such as Wikipedia and WordNet, to assess the important sentences based on the concepts they present. The integration of these knowledge bases, the method can evaluate semantic relationships and evaluate importance of sentence more effectively than traditional word-based methods. To model these relationships, graph-based or vector-based models are used to capturing the connections between sentences and their associated concepts. Since knowledge sources like WordNet and Wikipedia cover a broad spectrum of knowledge, concept-based methods are capable of handling a wide variety of topics and produce meaningful summaries of source document.

2) Linguistic Methods:

This method analyzes the relationships between concepts and words to understand meaning and create a summary. Abstractive summarization can be considered as a linguistic method since it involves semantic processing. Linguistic methods are capable to understand meaning of sentences within a document. This method is time-consuming and resource-intensive, requiring large amounts of memory to store linguistic repositories.

3) Statistical Methods:

Statistical methods determine important sentences based on the statistical features of a document like word frequency, sentence position, headings, and other structural attributes without necessarily considering their meaning. The summary is constructed by selecting the sentences with the highest scores.

4) Machine Learning Methods:

In this method, machine learning models are trained on datasets to classification of sentences as important or less important for Text summarization. Sentences are classified as either summary sentences or non-summary sentences. The probability of selecting a sentence for a summary is estimated based on training data and reference summaries. Extracted features are used by an algorithm to score and rank the sentence. Several machine learning methods such as fuzzy logic, support vector machines, Naive Bayes, artificial neural networks, and linear regression have been applied to text summarization. These methods require sufficient training datasets to obtain reliable performance. In certain cases, simple regression model works better than complex classification.

However, creating training data is time consuming because each sentence in the source text must be labeled as summary or non-summary.

5) Fuzzy Logic Based Methods:

Fuzzy logic can be used to assign feature values to words using logical values like 0 and 1, even though these values are not exactly represent real-world conditions [3]. In sentence scoring, the first step is to select a set of features for each sentence. In the next step, the fuzzy logic is applied to calculate a score based on these features. Each sentence is gives a value between 0 and 1, represents the importance of the sentence [4]. However, redundancy among the selected sentences may affect the quality of the summary.

6) Deep Learning Methods:

This method produces summaries by using document-level similarity through embeddings. Chen et al. [5] proposed automatic text summarization for a single document using reinforcement learning combined with a Recurrent Neural Network. The model extracts key sentence features through a sentence-level selective embeddings. Deep learning techniques allow the network to adapt to user preferences and reading content, although understanding the network's decision-making process remains challenging [6]. Mao et al. [7] combined three supervised learning methods with unsupervised learning to produce concise single-page summaries. Recent studies suggest that integrating multiple approaches can improve summary quality by leveraging the strengths of each method [8].

Abstractive Text Summarization:

The goal of abstractive text summarization is to create a concise summary that highlights the key points of a document. Such summaries may include words and phrases that do not appear in the original text.

The figure-3 represents the workflow of an abstractive text summarization system. The input document first undergoes pre-processing, after which the system creates an internal semantic representation of the text and generates a summary during the processing stage. Finally, post-processing refines the output to produce the final coherent summary.

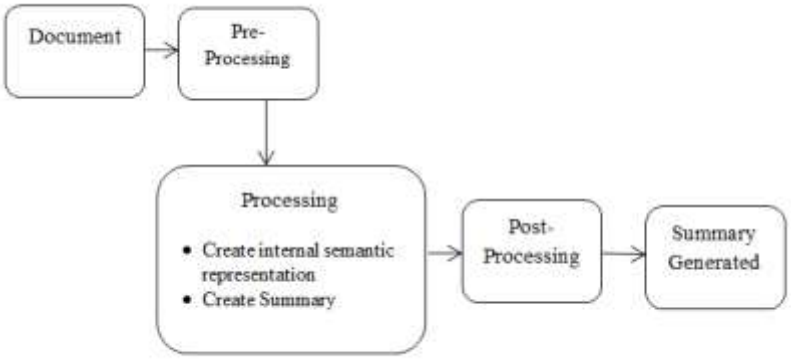

Figure 3: Abstractive Text Summarization Process [1].

Different Abstractive Text Summarization methods are available, as shown in Figure 1. These methods such as -

1) Abstractive Summarization Based on Structure

Different types of Abstractive Summarization Methods Based on Structure are as

i) Templates-Based Method

Templates refer to specific sentence structures that are commonly used in human-generated summaries for particular domains. To create an abstractive summary, the appropriate template slots are filled with information from the original document, taking into account the style of the source text [9]. Linguistic rules and knowledge can be employed to extract text samples that populate these template slots [10].

ii) Rule-based Method

The key steps in this approach involve identifying the concepts and terms in the source text, generating important points based on the document's subject, answering questions by locating these concepts and terms, and then using the answers within predefined patterns to produce the summary [10].

iii) Tree-based Method

The key steps in this approach involve identifying the concepts and terms in the source text, generating important points based on the document's subject, answering questions by locating these concepts and terms, and then using the answers within predefined patterns to produce the summary [10].

iv) Graph-Based Method

In this model, each node represents a word and is connected to other nodes based on its position. Directed edges capture the structure of the sentences. The graph-based approach constructs a graph representing the original content and generates an abstractive summary by exploring and scoring multiple sub paths within the graph.

v) Ontology-Based Method

Ontology-based method uses ontologies to generate abstractive summaries from input documents. A domain-specific ontology can be connected and applied to multiple documents within specialized domains. The summary is created by exploring traversing this mapping to identify and organize the relevant content.

2) Abstractive Summarization method based on Semantics

This method based on semantic representation, such as information items, semantic networks and predicate-argument structures, help captures meaning of input text. The abstractive summary is generated from these representations by selecting sentences most important sentences. Semantic role labeling is used to represent the document content through predicate-argument structures, which are ranked to determine the most important content [10].

3) Abstractive Summarization using Deep Learning

In recent, researchers have used deep sequence-to-sequence learning model to generate abstractive summaries. Recurrent Neural Networks (RNN) with attention mechanisms have achieved notable results in text summarization task. Current research focus to solve issues such as repetitive phrases, unknown vocabulary and other limitations of deep learning models. Some researchers propose integrating different summarization methods to improve the overall quality of the generated summaries [9].

Hybrid Text Summarization

Hybrid text summarization combines both extractive and abstractive approaches. During the pre-processing phase, extractive summarization is used to identify and isolate important sentences. In the summary generation phase, abstractive summarization produces the final summary. The post-processing phase then verifies the validity of the generated sentences, often applying rules to ensure consistency and coherence.

The figure-4 shows the hybrid text summarization process, where both extractive and abstractive methods are combined. The document first undergoes pre-processing, after which important sentences are selected using extractive summarization. These sentences are then refined and rewritten using abstractive summarization to improve coherence and readability. Finally, post-processing produces the final generated summary.

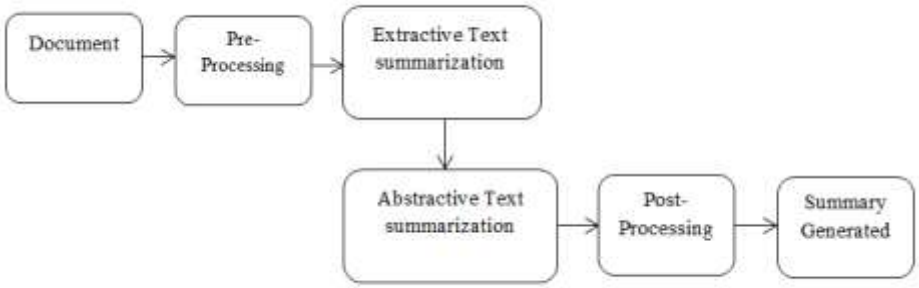

Figure 4: Hybrid Text Summarization Process

Two methods of Hybrid text summarization are as

1) Extractive to Abstractive Method

In this text summarization technique, extractive summarization is first applied to identify key sentences, and the resulting extract is then processed by an abstractive summarization method to generate the final summary. Researchers have proposed a two-phase hybrid approach for long texts, known as EA-LTS.

2) Extractive to Shallow Abstractive Method

In this text summarization technique, extractive summarization is applied first, followed by shallow abstractive methods on the extracted sentences to generate the final summary. Shallow abstractive techniques include synonym replacement, content fusion, and content compression.

Text Summarization Models

The researcher used a variety of proposed architectural models to produce a dataset and summary.

1. TextRank Algorithm [11]

TextRank is an extractive, unsupervised text summarization method. The algorithm begins by splitting the text into individual sentences after combining all content from the articles. Next, vector representations, or word embeddings, are generated for each sentence. The sentence vectors are compared, and the results are stored in a similarity matrix. This matrix is then converted into a graph, and the highest-ranked sentences are selected to form the summary.

2. TF-IDF[12]

In the Term Frequency-Inverse Document Frequency (TF-IDF) model, Term Frequency (TF) represents the number of times a word appears in a document, while Inverse Document Frequency (IDF) indicates how many documents in the collection contain that term.

3. Encoder-decoder model[08]

In an encoder-decoder model, the encoder transforms the input sequence into a one-dimensional hidden vector. The decoder then generates the output sequence from this hidden vector. Given a sequence of words, the model is trained to predict the probability of the target sequence.

4. Recurrent Neural Network [08]

In this model, each step receives input from the output of the previous step. Unlike typical neural networks, where inputs and outputs are independent, predicting the next word in a sequence requires knowledge of the preceding words. Therefore, it is essential to retain the previous words for accurate prediction.

5. Bidirectional recurrent neural networks (BRNN)[08]

When predicting the current state, bidirectional RNNs consider both past and future information, whereas unidirectional RNNs rely solely on past inputs.

6. Long short-term memory (LSTM)[11]

If past conditions influencing the current prediction are not considered, an RNN may fail to accurately estimate the present state. However, capturing dependencies from several steps earlier can be challenging for standard RNNs. LSTM overcome this drawback using three gates as input, output, and forget gate that regulate the flow of information in the network's hidden layers.

7. Gated Recurrent Units (GRUs)[13]

Limitations of standard RNNs is short-term memory. This GRU process information using hidden states and gates that control memory GRU process like the input, output, and forget gates in LSTMs and also the reset and update gates determine what information is retained and how much is forwarded.

8. Attention mechanism and Copy mechanism [14]

The attention mechanism performs the decoder with a weighted context from the encoder, allowing the model to focus on the most important parts of the input text when predicting each word in the output sequence. The copy mechanism, selects specific fragments from the input text and arranges them appropriately in the generated summary.

9. K-Means Clustering with TF-IDF [15]

In this approach, TF-IDF generates a row matrix of phrases. The sentence scores are then used with the K-means algorithm to cluster the phrases, which are subsequently used to create the summary.

Research Methodology

Secondary data for this descriptive study were collected from various sources, including websites, periodicals, publications, magazines, and journals. The data were then analyzed and evaluated to draw relevant conclusions and inferences.

Literature Reviews

Several studies on text summarization have been reported in the literature. A few of them are reviewed here.

English

Aakash Srivatava et al. [12] developed an extractive summarization system based on sentence frequency. The text summarization process involves two steps: tokenization and sentence scoring using TF-IDF. The final summary is generated by selecting the most relevant sentences from the text. The study reported limited results, which were attributed to the low quality of the summaries produced.

Anish Jadhav et al. [8] used a neural network approach for text summarization. A concise and linguistically accurate summary of the entire document was produced by processing the extracted sentences through an encoder-decoder model. The proposed method applies attention to the most relevant sentences, which are then fed into a bidirectional RNN to identify and represent the main concepts of the document.

Rahim Khan et al. [15] proposed an extractive text summarization method that combines TF-IDF and K-Means clustering to generate summaries from textual documents. In their approach, the document is first preprocessed and sentences are represented using TF-IDF weights. The K-Means algorithm is then used to cluster similar sentences, and the most representative sentences from each cluster are selected to form the final summary. The study shows that this method helps in identifying important sentences and reducing redundancy, thereby improving the quality of automatically generated summaries.

Phang et al. [16] proposed **PEGASUS-X,** an extension of the PEGASUS transformer designed for **long-document abstractive summarization.** By extending the context length and using gap-sentence pretraining, it achieves state-of-the-art results on arXiv, PubMed, BigPatent, and GovReport datasets with ROUGE-1 = 49.96, ROUGE-2 = 31.17, and ROUGE-L = 39.45.

Arabic

Wael Etaiwi et al. [17] developed an abstractive text summarization system for the Arabic language using semantic graph embedding. The proposed method involves creating a semantic graph, embedding it, and generating a summary. After the embedding phase, the final summary is produced by a deep neural network using the SemanticGraph2Vec technique.

Samira Lagrini et al. [18] proposed an Arabic text summarization system combining statistical methods with Rhetorical Structure Theory (RST). Phase I involves text segmentation, identifying rhetorical relations, and sentence compression, while Phase II performs preprocessing, sentence scoring and summary creation. The method achieved ROUGE-1 and ROUGE-2 scores of 0.453 and 0.308, with precision values of 0.471 and 0.337.

Molham Al Maleh et al. [14] proposed an abstractive text summarization system using a sequence-to-sequence model with an attention mechanism. This system generates concise summaries and measured the performance of system using Recall-Oriented Understudy for Gisting Evaluation (ROUGE) scores.

Italian

Moreno La Quatra et al. [19] created an Italian text summarization system using a Seq2Seq model with the Bidirectional and Auto-Regressive Transformers for Italian (BART-IT) method. The system is pre-trained on an Italian literature dataset to obtain language-specific features and then fine-tuned on several datasets optimized for abstractive summarization. The results show that BART-IT multilingual models with large parameters, achieving better performance. Based on ROUGE metrics performance of BART-IT model is better than Multilingual Text-to-Text Transfer Transformer (mT5) model.

Landro et al. [20] developed two Italian datasets, Fanpage and IlPost, for abstractive text summarization. The datasets pair news articles with professionally written summaries. This system provides publicly available datasets for further research in Italian summarization.

French

Yongxin Zhou et al. [21] proposed an abstractive text summarization system in French dialogue using the pre-trained BARThez model. The system generates summaries of call center conversations from the DECODA corpus. This system suffer from omissions, hallucinations, and minor grammatical errors.

Gesnouin et al. [22] proposed text summarization system in French language using fine-tuned on over 15,000 French legislative amendment–summary pairs. Performance of the system measured using expert human scoring. This model shows that the quality of the summary good for domain-specific LLMs for complex French legal document.

Chinese

Xuefeng Xi and Zhou Pi et al. [13] proposed text summarization system in Chinese language with Chinese police inquiry record dataset using attention with Seq2Seq model. The system contains global encoding unit (GEU) combine with CNN, attention, and gated units to enhance context representation from the GRU encoder. The word-based embedding model performance is better than the character-based model.

Jiang et al. [23] developed an abstractive text summarization model in Chinese language using multi-layer attention with seq2seq model Evaluated on the LCSTS dataset, it achieved a ROUGE-1 score of 39.51 and ROUGE-L score of 37.25. The model also employs 1×1 convolutional layers to reduce complexity and improve efficiency, This research shows that deeper attention mechanisms can enhance summarization performance for Chinese short texts.

Nepali

Rishi Saran Khanal et al. [11] developed an extractive text summarization for Nepali language using text ranking and LSTM with attention. In this proposed system after preprocessing, a GloVe embedding model was trained on a Nepali news corpus to generate word vectors. As per the ROUGE score it shows that the performance of proposed system is good.

Timalsina, Paudel, and Shahi et al. [24] proposed an abstractive text summarization model in Nepali language using an attention-based LSTM encoder-decoder. This summarization uses a web-scraped Nepali news dataset. It achieved the best results from a single-layer encoder. This study identify the feasibility of neural abstractive summarization for low-resource languages.

Dhakal and Baral et al. [25] developed an abstractive Nepali summarization model by fine-tuning multilingual transformers (mBART, mT5) on a Nepali news headline dataset. Using LoRA and 4-bit quantization, the model enables efficient training while maintaining quality, with evaluations based on ROUGE and human judgments. The results establish a benchmark for transformer-based summarization in low-resource Nepali.

Swedish

Julius Mohsen et al. [26] are developed an abstractive text summarization for Swedish news articles. This summarization uses an encoder-decoder model and BERT and fine-tuned with various datasets. The model produced significantly better summaries on filtered data compared to unfiltered data.

Monsen and Rennes et al. [27] proposed extractive and abstractive summarization for Swedish news articles. Extractive summarization uses BERT and abstractive summarization uses an encoder-decoder model. This text summarization model generate good quality summary.

Portuguese

Lins et al. [28] was proposed abstractive summarization model for Brazilian Portuguese legal documents using contrastive learning with a Seq2Seq encoder-decoder. As per the ROUGE evaluated compared with LLM shows that it creates high quality summaries that preserve the meaning and structure of the legal document.

Russian

Nikolich and Puchkova et al. [29] are proposed abstractive text summarization in Russian language using fine-tuned ruGPT3Small with Russian news dataset. The model produces coherent summaries but scores lower on ROUGE compared to transformer-based model like mBART. Shows that both generative models have potential and limitations for Russian summarization.

Japanese

Iwasaki et al. [30] developed an abstractive text summarization model for Japanese language using a BERT encoder decoder transformer. The Livedoor News Corpus used in the summarization with LSTM baselines on ROUGE, producing coherent summaries while maintaining semantic nuances. Sawahata and Nishino et al. [31] are proposed extractive summarization of Japanese academic papers using topic modeling. This text summarization uses LDA and LSA to determine key topics and select representative sentences, improving thematic coherence and shows that promising results for Japanese academic text summarization.

Table-1 shows that a comprehensive overview of prior research on text summarization for non-Indian languages. It summarizes research conducted by various researchers, information of the languages addressed, datasets used, methodologies and approaches utilized, and the performance of the text summarization system.

Table 1 : Text Summarization comparative statement

Sr. No.	Language of Research and Reference	Dataset(s)	Approach	Method	Result
1	English [12]	-	Extractive	TF-IDF frequency	Short text only
2	English [8]	News	Hybrid	Bi-RNN Encoder–Decoder	High Accuracy
3	English [15]	News headlines	Extractive	K-Means + TF-IDF	BLEU: 0.62
4	English [16]	arXiv, BigPatent, GovReport, PubMed	Abstractive	Transformer with gap-sentence pretraining with extended context architecture for long input	R-1: 49.96 R-2: 31.17 R-L:39.45
5	Arabic [17]	News	Abstractive	Graph based Semantic	R-1: 0.442 R-2: 0.214 R-L: 0.401
6	Arabic [18]	Essex Arabic Summaries Corpus (EASC) News	Hybrid	RST + statistical	Recall 0.453 & 0.308, precision 0.471 & 0.337 for both metrics ROUGE-1 & ROUGE-2. very good performance as per measures
7	Arabic [14]	Headline	Abstractive	Seq2Seq + attention + copy	R-1: 0.413 R-2: 0.192 R-L: 0.381
8	Italian [19]	News & Wikipedia	Abstractive	Seq2seq with BART-IT	R-1: 0.462 R-2: 0.243 R-L: 0.421

					BERTScore: 0.948
9	Italian [20]	News	Abstractive	Seq2Seq Transformer models	R-1:0.40 R-2:0.18 R-L:0.35
10	French [21]	DECODA dialogue	Abstractive	BARThez	R-1:35.45 R-2:16.85 R-L:29.45
11	French [22]	French legislative amendment–summary pairs	Abstractive	Fine-tuning LLaMA-based models	expert scoring (0–20 scale) rather than ROUGE
12	Chinese [13]	Police case inquiry dataset	Abstractive	Neural Seq2Seq + GEU	R-1: 0.489 R-2: 0.312 R-L: 0.457
13	Chinese [23]	**LCSTS (short text)**	Abstractive	Seq2Seq neural architecture with multi-layer attention model	R-1:39.51 R-2: Not reported R-L:37.25
14	Nepali [11]	News	Extractive	LSTM with text ranking	R-1:0.431 R-2: 0.428
15	Nepali [24]	Web-scraped Nepali news articles	Abstractive	Encoder-decoder RNN with **attention** (LSTM cells)	R-1:15.74 R-2:3.29 R-L:15.21
16	Nepali [25]	Web-scraped Nepali news headline corpus	Abstractive	Fine-tuning **multilingual transformer models** with **LoRA** (Low-Rank Adaptation) and **quantization** for efficient training	No
17	Swedish [26]	Swedish news	Abstractive	BERT Encoder–Decoder	R-1:27.54 BERTScore: 17.52
18	Swedish [27]	Filtered Swedish *Dagens Nyheter (DN)* news articles	Extractive & Abstractive	BERT-based sentence selection for extractive and ; Encoder-decoder model used for abstractive summarization	R-1:33.73 R-2: 13.31 R-L: –Nil
19	Portuguese [28]	CLSJUR.BR – Brazilian legal rulings and decisions	Abstractive	Seq-2-Seq encoder-decoder model with contrastive learning, fine-tuned on legal	Model outperforms prior baselines and LLMs for

				text summaries; compared against LLM baselines	Brazilian Portuguese legal text summarization, achieving high-quality abstractive summaries.
20	Russian[29]	**Gazeta** news summarization dataset — Russian news articles with human summaries (63 435 pairs; 90/10 train/test)	Abstractive	**ruGPT3Small** (125 M parameters) fine-tuned with specific prompts and tuned hyperparameters (temperature = 0, beam search tuned, no-repeat n-gram constraint, etc.)	R-1(F1): 11.4 R-2 (F1): 1.4 R-L(F1): 10.0 BLEU: 23.1 BERTScore (F1): 0.89
21	Japanese [30]	Livedoor News Corpus	Abstractive	BERT encoder + Transformer decoder	R-1: 44.31 R-2: 21.09 R-L: 41.64
22	Japanese [31]	Japanese academic papers (scientific articles)	Extractive	LDA, LSA topic modeling	R-L: 34.1

Table 1 shows that the variety of text summarization approaches across non-Indian languages. Text summarization in English language commonly employ both extractive and abstractive methods and obtain high ROUGE scores using neural models. Text summarization in Arabic and European languages use abstractive or hybrid transformer-based methods, while summarization in Chinese, Japanese, and Nepali languages use a mix of extractive and abstractive methods. In general, the performance of text summarization depends on model selection, dataset type, and language characteristics.

Following figure-5 shows that an overview of the research paper publication status for different non-Indian languages.

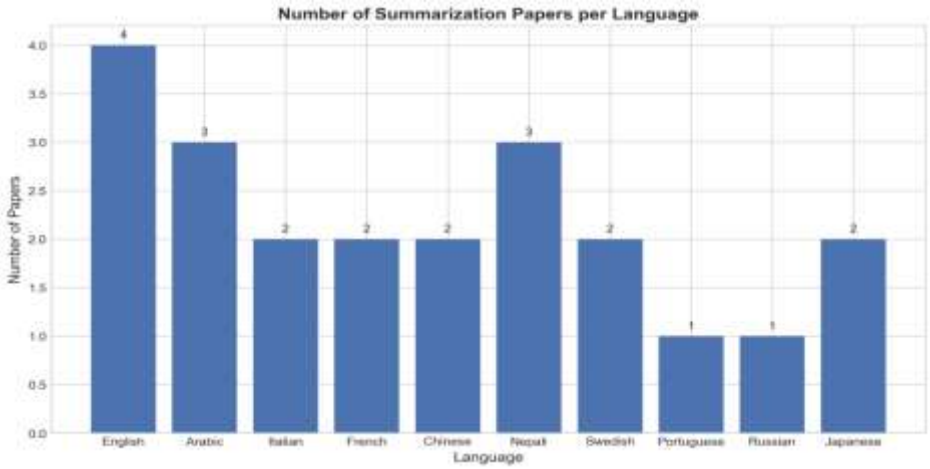

Figure 5: Research Status of various non-Indian languages

The figure-5 shows the distribution of summarization research papers across different non-Indian languages. This distribution is focused on English and selected languages, while some languages remain relatively underexplored in text summarization studies.

Research in text summarization for non-Indian languages has expanded in recent years, as shown in figure-6, which shows year-wise publications across these non-Indian languages.

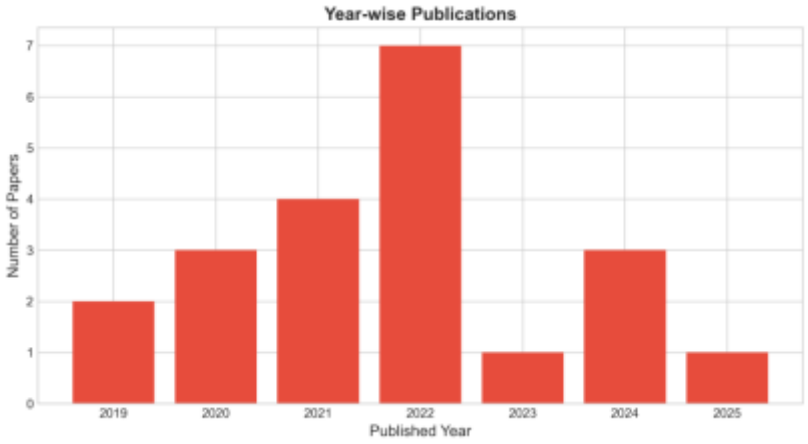

Figure 6: Year Wise Publication for Text Summarization

The figure-6 shows a steady rise in publications peaking in 2022, followed by a decline, indicating fluctuating interest in research activity in recent years.

Following figure-7 shows that a grouped comparison of ROUGE-1, ROUGE-2, and ROUGE-L scores across different languages and studies.

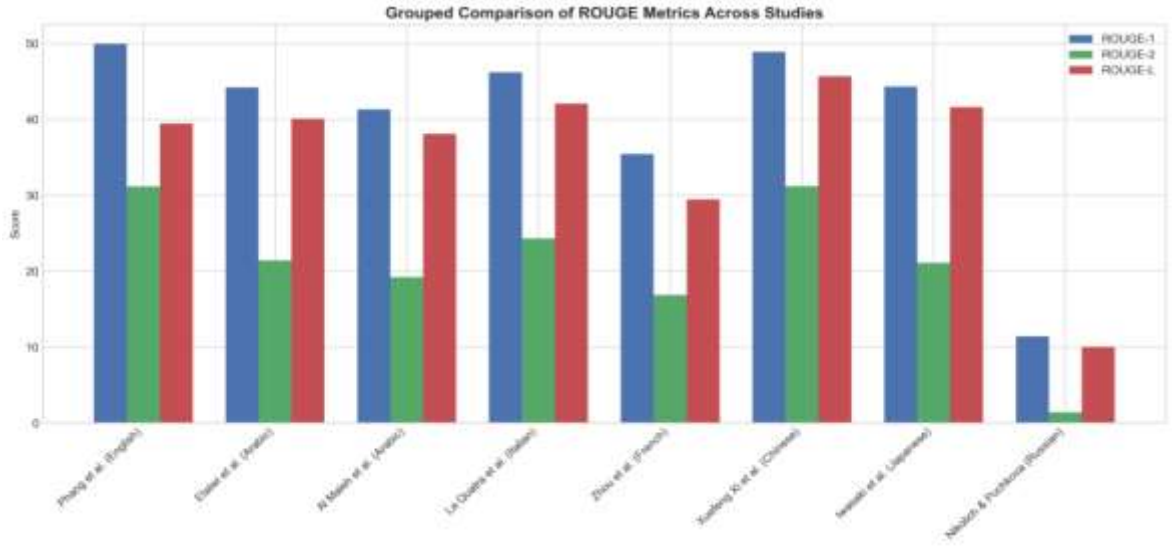

Figure 7: Grouped comparison of ROUGE Metrics across different languages

The figure-7 shows that summarization performance, measured using ROUGE metrics, differs across languages and studies. ROUGE-1 scores are generally the highest, showing that models capture basic content. ROUGE-2 scores are lower, reflecting difficulty in capturing more precise context. ROUGE-L scores lie between ROUGE-1 and ROUGE-2, indicating that moderate ability in preserving sentence structure. As per the figure shows that English and Chinese studies perform best, while Russian shows the weakest results.

Conclusion

Text summarization is a widely studied application in Natural Language Processing (NLP) that aims to create concise summary while preserving meaning. The text summarization various algorithms and techniques are used for foreign languages have been examined in this paper. We determined that a large number of text summarization studies deal with foreign languages. Review shows that TF-IDF, GEU, sequence-to-sequence model, LSTM encoder decoder model, attention and copy mechanism is some of the methods for text summarization. According to this review, both the extractive and abstractive text summarization methods perform well, but integrating the two methods gives better results than using either method independently. The review concludes that more accurate summary is produced when the sequence-to-sequence method is combined with a copy and attention mechanisms.

References

1. Alhojely, S., & Kalita, J. (2020, December). Recent Progress on Text Summarization. *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*. https://doi.org/10.1109/csci51800.2020.00278
2. El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021, March). Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, *165*, 113679. https://doi.org/10.1016/j.eswa.2020.113679
3. Vijay Kumar, N., & Janga Reddy, M. (2019). Factual Instance Tweet Summarization and Opinion Analysis of Sport Competition. *Advances in Intelligent Systems and Computing*, 153–162. https://doi.org/10.1007/978-981-13-3393-4_16

4. Nazari, N. & Mahadavi, M.A.(2019). A survey on automatic text summarization. Journal of AI and Data Mining, 121-135. https://doi.org/10.22044/JADM.2018.6139.1726
5. Chen, L., & Nguyen, M. L. (2019, October). Sentence Selective Neural Extractive Summarization with Reinforcement Learning. *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*. https://doi.org/10.1109/kse.2019.8919490
6. Alami, N., Meknassi, M., & En-nahnahi, N. (2019, June). Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning. *Expert Systems With Applications*, *123*, 195–211. https://doi.org/10.1016/j.eswa.2019.01.037
7. Mao, X., Yang, H., Huang, S., Liu, Y., & Li, R. (2019, November). Extractive summarization using supervised and unsupervised learning. *Expert Systems With Applications*, *133*, 173–181. https://doi.org/10.1016/j.eswa.2019.05.011
8. Jadhav, A., Jain, R., Fernandes, S., & Shaikh, S. (2019, December). Text Summarization using Neural Networks. *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*. https://doi.org/10.1109/icac347590.2019.9036739
9. Kouris, P., Alexandridis, G., & Stafylopatis, A. (2019). Abstractive Text Summarization Based on Deep Learning and Semantic Content Generalization. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/p19-1501
10. Gupta, S., & Gupta, S. K. (2019, May). Abstractive summarization: An overview of the state of the art. *Expert Systems with Applications*, *121*, 49–65. https://doi.org/10.1016/j.eswa.2018.12.011
11. Khanal, R.S., Adhikari, S., & Thapa, S. (2021, Octomber). Extractive Method for Nepali Text Summarization Using Text Ranking and LSTM. Proceedings of 10th IOE Graduate Conference, ISSN: 2350-8914 (Online), 2350-8906 (Print).
12. SRIVASTAVA, CHAUHAN, DAHARWAL, MUKATI, & KAVIMANDAN, S. (2022, July). Text Summarizer Using NLP (Natural Language Processing). *ICONIC RESEARCH AND ENGINEERING JOURNALS*, *6*(2456–8880).
13. Xi, X., Pi, Z., & Zhou, G. (2020, October 6). Global Encoding for Long Chinese Text Summarization. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *19*(6), 1–17. https://doi.org/10.1145/3407911
14. Al-Maleh, M., & Desouki, S. (2020, December). Arabic text summarization using deep learning approach. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00386-7
15. Khan, R., Qian, Y., & Naeem, S. (2019, May 8). Extractive based Text Summarization Using KMeans and TF-IDF. *International Journal of Information Engineering and Electronic Business*, *11*(3), 33–44. https://doi.org/10.5815/ijieeb.2019.03.05
16. Phang, J., Beltagy, I., Peters, M. E., Cohan, A., & Zettlemoyer, L. (2022). *PEGASUS-X: Extreme summarization for long documents* (arXiv:2208.04347). arXiv. https://arxiv.org/abs/2208.04347
17. Etaiwi, W., & Awajan, A. (2022, September 6). SemG-TS: Abstractive Arabic Text Summarization Using Semantic Graph Embedding. *Mathematics*, *10*(18), 3225. https://doi.org/10.3390/math10183225
18. Lagrini, S. (2021). *A New Approach for Arabic Text Summarization*. ACL Anthology. https://aclanthology.org/2021.icnlsp-1.20
19. La Quatra, M., & Cagliero, L. (2022, December 27). BART-IT: An Efficient Sequence-to-Sequence Model for Italian Text Summarization. *Future Internet*, *15*(1), 15. https://doi.org/10.3390/fi15010015

20. Nicola Landro, Ignazio Gallo, Rosa La Grassa, & Enrico Federici. (2022). *Two new datasets for Italian-language abstractive text summarization.* Information, 13(5), 228. https://doi.org/10.3390/info13050228
21. Zhou,Y., Portet, F. & Ringeval, F. (2022, June 20). Effectiveness of French Language Models on Abstractive Dialogue Summarization Task. Processing of 13th Conference on *LREC 2022*, 3571-3581
22. Gesnouin, J., Tannier, Y., Da Silva, C. G., Tapory, H., Brier, C., Simon, H., Rozenberg, R., Woehrel, H., El Yakaabi, M., Binder, T., Marie, G., Caron, E., Nogueira, M., Fontas, T., Puydebois, L., Theophile, M., Morandi, S., Petit, M., Creissac, D., Ennouchy, P., Valetoux, E., Visade, C., Balloux, S., Cortes, E., Devineau, P.-E., Tan, U., Mac Namara, E., & Yang, S. (2024). *LLaMandement: Large Language Models for Summarization of French Legislative Proposals* (arXiv preprint). arXiv. https://doi.org/10.48550/arXiv.2401.16182
23. Jiang, J. (2024). *Chinese Text Summarization Based on Multi-Layer Attention*. In *Proceedings of the 1st International Conference on Data Science and Engineering* (Volume 1: ICDSE), 565–568. SciTePress. https://doi.org/10.5220/0012839000004547
24. Timalsina, B., Paudel, N., & Shahi, T. B. (2022). *Attention based Recurrent Neural Network for Nepali Text Summarization. Journal of Institute of Science and Technology*, *27*(1), 141–148. https://doi.org/10.3126/jist.v27i1.46709
25. Dhakal, P., & Baral, D. S. (2025). *Abstractive Summarization of Low Resourced Nepali Language Using Multilingual Transformers.* In *Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL 2025)* (pp. 124–133). International Committee on Computational Linguistics. https://aclanthology.org/2025.chipsal-1.12/
26. Monsen, J.(2021). Building high-quality datasets for abstractive text summarization-A filtering based method applied on Swedish news articles. Department of Computer and Information Science, Linköping University
27. Monsen, J., & Rennes, E. (2022). *Perceived text quality and readability in extractive and abstractive summaries.* In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, … & S. Piperidis (Eds.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 305–312). European Language Resources Association. https://aclanthology.org/2022.lrec-1.32/
28. Lins, A. A., Carvalho, C. S., Bomfim, F. D. C. J., Bentes, D. C., & Pinheiro, V. (2024). *CLSJUR.BR - A model for abstractive summarization of legal documents in Portuguese language based on contrastive learning.* In P. Gamallo, D. Claro, A. Teixeira, L. Real, M. Garcia, H. G. Oliveira, & R. Amaro (Eds.), *Proceedings of the 16th International Conference on Computational Processing of Portuguese – Vol. 1* (pp. 321–331). Association for Computational Linguistics. https://aclanthology.org/2024.propor-1.33/
29. Nikolich, A., & Puchkova, A. (2021). *Fine-tuning GPT-3 for Russian text summarization* (arXiv:2108.03502). arXiv. https://doi.org/10.48550/arXiv.2108.03502
30. Iwasaki, Y., Yamashita, A., Konno, Y., & Matsubayashi, K. (2020). *Japanese abstractive text summarization using BERT. Advances in Science, Technology and Engineering Systems Journal, 5*(6), 1674–1682. https://doi.org/10.25046/AJ0506199
31. Sawahata, M., & Nishino, T. (2023). *Extractive summarization of Japanese academic papers using topic modeling*. *International Journal of Engineering Education, 39*(2), 759–766. Retrieved from https://doi.org/10.52731/iee.v9.i2.759