International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 12
***Article Received:** 25 July 2021 **Revised:** 12 September 2021 **Accepted:** 30 November 2021*

A Survey of Text Summarization Systems for Indian Languages

Vipan,

Department of Computer Applications, Sikh National College, Banga, Punjab, India.

Abstract

Text summarization plays a crucial role in information retrieval, particularly in the context of Indian languages, where linguistic diversity and resource scarcity pose unique challenges. This survey explores the current landscape of text summarization techniques applied to several major Indian languages, including Hindi, Tamil, Marathi, Punjabi, Bengali, and Kannada. The paper provides a comprehensive review of both extractive and abstractive summarization methods, highlighting language-specific strategies, challenges, and progress in the field. It discusses key issues such as the lack of large-scale annotated corpora, the complexity of handling linguistic variations, and the difficulty of processing code-mixed and informal language. Furthermore, the paper outlines future research directions, including the integration of multilingual models, the development of large-scale datasets, and the need for domain-specific summarization tools. The findings emphasize the importance of creating culturally and regionally sensitive summarization systems that are tailored to the unique characteristics of Indian languages.

Keywords: Text Summarization, Indian Languages, Extractive Summarization, Abstractive Summarization, Natural Language Processing (NLP), Indian Languages Resources.

1. Introduction

In the digital age, the volume of textual data generated and consumed daily has increased exponentially. Text summarization—an automatic process of condensing long pieces of text into shorter versions while retaining the essential information—has become an indispensable tool for managing and making sense of this vast information overload[3]. While significant progress has been made in text summarization for widely spoken languages such as English, the development of similar systems for Indian languages is still emerging and poses unique challenges[4].

India is a linguistically diverse nation with 22 officially recognized languages and hundreds of dialects spoken across its vast geography[1]. This diversity underscores the need for robust and inclusive text summarization systems that cater to various linguistic communities. Such systems can greatly enhance access to information, promote digital inclusion, and support multilingual content generation, especially in domains like news media, education, e-governance, and social media monitoring[4],[5].

Moreover, most Indian languages are low-resource in the context of Natural Language Processing (NLP), lacking large-scale annotated corpora, linguistic tools, and computational resources[2]. This scarcity intensifies the need for innovative summarization techniques tailored to the linguistic and structural features of these languages. Effective summarization in Indian languages can also play a crucial role in bridging the digital divide and fostering equitable information dissemination in a multilingual society[1],[2].

By focusing on text summarization for Indian languages, This not only contribute to the global advancement of NLP technologies but also address critical local needs, empowering millions of native language speakers to access and interact with digital content more efficiently[3],[4].

The primary objective of this survey is to present a comprehensive overview of the current state-of-the-art in automatic text summarization systems developed specifically for Indian languages. This includes an in-depth analysis of existing techniques, methodologies, tools, and datasets that have been employed across different linguistic contexts within India[6],[7]. The survey aims to categorize and compare various summarization approaches—both extractive and abstractive—based on their performance, language adaptability, resource requirements, and real-world applicability[7],[8].

The scope of this work spans multiple Indian languages, such as Hindi, Punjabi, Bengali, Tamil, Telugu, Malayalam, Marathi, and others, while highlighting language-specific challenges and innovations[9]. This paper focuses on peer-reviewed academic publications, publicly available systems, and research prototypes developed over the past two decades[6,8]. Furthermore, the survey aims to identify existing gaps in research and offer future directions that could foster inclusive, resource-efficient, and context-aware summarization systems tailored to the linguistic diversity of India[9],[10].

Overview of Text Summarization Techniques and Their Significance

Text summarization is a core area of Natural Language Processing (NLP) that focuses on generating concise and meaningful summaries from large volumes of textual data. The increasing prevalence of digital content across various sectors—news media, education, government documentation, and healthcare—has amplified the need for efficient summarization systems. These systems are particularly valuable in multilingual societies like India, where the ability to quickly access critical information in diverse native languages is essential for inclusive information dissemination [6], [9].

Text summarization techniques are broadly categorized into extractive and abstractive methods.

Extractive summarization identifies and selects important sentences or phrases directly from the source text to create the summary. It relies on features such as sentence position, term frequency-inverse document frequency (TF-IDF), keyword density, and similarity measures. Due to its relative simplicity and language independence, extractive summarization has been widely adopted, particularly for low-resource languages like many of those spoken in India [6], [7].

Abstractive summarization, by contrast, involves understanding the semantics of the source content and generating novel sentences that encapsulate the core ideas. This approach mimics human-like summarization but requires advanced machine learning models such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks. Despite its effectiveness in producing coherent and fluent summaries, abstractive summarization is computationally intensive and highly data-dependent, making it challenging for Indian languages with limited annotated corpora [8], [10].

Additionally, hybrid techniques that integrate extractive and abstractive strategies are gaining traction. These methods aim to combine the efficiency of extractive models with the semantic depth of abstractive models, striking a balance between performance and linguistic richness [7].

The significance of text summarization lies in its ability to enhance content comprehension, reduce information overload, and facilitate timely decision-making. In the Indian context, where information is produced and consumed in multiple languages and dialects, summarization systems can support broader access to information, bridge literacy gaps, and enable multilingual content delivery [6], [9], [10].

Challenges and Opportunities Specific to Indian Languages

Developing effective text summarization systems for Indian languages involves navigating a range of linguistic and technological challenges, primarily due to the vast diversity and complexity of these languages. India officially recognizes 22 scheduled languages and hosts hundreds of regional dialects, each characterized by unique scripts, grammatical structures, and syntactic rules [1], [4]. This diversity presents a significant challenge in designing generalized summarization systems that can work seamlessly across multiple languages and domains.

A core obstacle is the scarcity of annotated corpora and linguistic tools. Most Indian languages are considered low-resource in the context of Natural Language Processing (NLP), lacking the large-scale, high-quality datasets necessary for training data-intensive models—especially those used in abstractive summarization [2], [5], [10]. In addition, many languages still do not have access to essential NLP components such as part-of-speech (POS) taggers, named entity recognizers, and parsers, all of which are foundational for both extractive and abstractive summarization systems [3], [6].

The script diversity and morphological richness of Indian languages further complicate the summarization process. Languages such as Hindi, Punjabi, Tamil, Telugu, Malayalam, and Bengali employ different writing systems and exhibit agglutinative or highly inflectional morphological properties, making tasks like tokenization, lemmatization, and sentence segmentation more difficult compared to English [4], [6], [7].

Additionally, the phenomenon of code-mixing and code-switching, where blend of English with regional languages (especially on social media and digital platforms), introduces syntactic irregularities and hybrid vocabulary. This disrupts standard processing pipelines and reduces the accuracy of summarization systems trained on monolingual corpora [7], [9].

Despite these challenges, there are considerable opportunities for advancement. Increased academic interest are fostering the development of linguistic resources and open-access corpora for Indian languages [1], [5]. Recent advances in multilingual embeddings and transformer-based models like mBERT and IndicBERT allow for transfer learning across languages, making it feasible to build summarization systems for under-resourced languages using cross-lingual knowledge [8], [10].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 12
***Article Received:** 25 July 2021 **Revised:** 12 September 2021 **Accepted:** 30 November 2021*

Furthermore, the societal impact of multilingual summarization in India is substantial. By enabling quick access to essential information in healthcare, education, and governance, especially in rural and non-English-speaking communities, these systems promote inclusivity and digital empowerment [4], [9]. They also pave the way for building intelligent voice assistants, educational tools, and content recommendation engines tailored to native language users.

In conclusion, while Indian languages present numerous linguistic, computational, and resource-based hurdles, they also offer rich opportunities for innovation and societal transformation. Addressing these challenges can catalyze the development of equitable and context-aware NLP systems that reflect India's linguistic plurality.

Literature review of Text Summarization Systems for Indian languages

Tamil			
S.No.	**Researchers/Authors**	**Technology**	**Year**
1	Thevatheepan Priyadharshan, Sagara Sumathipala[11]	Restricted Boltzmann Machine (RBM)	2018
2	Syed Sabir Mohamed, Shanmugasundaram Hariharan[12]	Centroid Approach	2016
3	Sankar K, Vijay Sundar Ram R, Sobha Lalitha Devi[13]	Scoring of sentences based on World frequency analysis, world positional & string pattern based weight calculation	2011
4	J. Jai Hari Raju, P. Indhu Reka, K.K Nandavi, Dr. Madhan Karky[14]	Just Framework proposed, Humanness score as evaluator for summary	2011
5	M. Banu, C. Karthika, P. Sudarmani , T.V. Geetha[15]	Sub Graph, Language-Neutral syntax, Subject Oriented Predicate(SOP) is used to create semantic graph	2007

Marathi			
S.No.	**Researchers/Authors**	**Technology**	**Year**
1	Apurva D. Dhawale, Sonali B. Kulkarni, Vaishali M. Kumbhakarna[16]	LDA(Latent Dirichlet Allocation), LNS (Label Induction Grouping), SVM (Support Vector Machine)	2020
2	Anishka Chaudhari, Akash Dole, Deepali Kadam[17]	Recurrent Neural Network (RNN)	2019
3	Yogeshwari V. Rathod[18]	Text-rank algorithm , Page-rank algorithm	2018
4	Vaishali V. Sarwadnya, Sheetal S. Sonawane[19]	Graph based model, Text Ranking	2018
5	Shraddha A. Narhari, Rajashree Shedge[20]	Label Induction Grouping (LINGO), Single Value Decomposition (SVD), Principle Component analysis (PCA)	2017
6	Pooja Bolaj, Sharvari Govilkar[21]	Supervised Learning Methods(Naïve Bayes (NB), Modified K Nearest Neighbor (MKNN) and Support Vector Machine (SVM))	2016
7	N.Dangre, A. Bodke, A. Date, S. Rungta, S.S. Pathak[22]	Stop word removal & stemming, Clustring(K-means, KNN, SVM), Ranking algorithm	2016

Hindi			
S.No.	**Researchers/Authors**	**Technology used**	**Year**
1	Sakshee Vijay, Vartika Rai, Sorabh Gupta, Anshuman Vijayvargia, Dipti Misra Sharma[23]	Statistical and linguistic extractes of the text, word and sentence level features.	2017

2	Vipul Dalal, Latesh Malik[24]	Subject-object-verb (SOV) triples, Semantic graph of the document and Particle Swarm Optimization (PSO) algorithm.	2017
3	Swati Sargule, Ramesh M Kagalkar[25]	Bernoulli model of randomness, lexical association, Sentences indexing, word indexing.	2016
4	K. Vimal Kumar & Divakar Yadav[26]	Scoring the sentences based on occurrence of the radix of thematic words.	2015
		Punjabi	
S.No.	**Researchers/Authors**	**Technology**	**Year**
1	Arti Jain, Divakar Yadav, Anuja Arora[27]	Particle swarm optimization (PSO), Calculation within PSO is performed using fitness Function which looks into various statistical and linguistic features of the Punjabi datasets. Data Set – Monolingual Punjabi corpus (30,000 sentences), Punjabi Hindi parallel corpus(1000 sentences)	2021
2	Vishal Gupta, Gurpreet Singh Lehal[28]	Based on Statistical & linguistic features, word boundary identification, Stop word elimination, Punjabi language stemmer, eliminating duplicate sentences	2012
3	Vishal Gupta, Gurpreet Singh Lehal[29]	Sentence length feature, keyword selection (TS-ISF approach), mathematical regression used to estimate text feature weights	2011
		Bengali	
S.No.	**Researchers/Authors**	**Technology**	**Year**
1	Radia Rayan Chowdhury, Mir Tafseer Nayeem, Tahsin Tasnim Mim, Md. Saifur Rahman Chowdhury, Taufiqul Jannat[30]	Graph-based unsupervised, POS tagger+Pre trained Bengali language model	2021
2	Prithwiraj Bhattacharjee , Avi Mallick , Md. Saiful Islam , and Marium-E-Jannat[31]	Sequence to sequence based Long Short-Term Memory (LSTM) network model with attention at encoder-decoder	2020
3	Mahimul Islam, Fariha Nuzhat Majumdar, Asadullahhil Galib, Md Moinul Hoque[32]	Keyword scoring, sentiment analysis, and the interconnection of sentences, text ranking	2020
4	Sheikh Abujar, Abu Kaisar Mohammad Masum, Md. Sanzidul Islam, Fahad Faisal & Syed Akhter Hossain[33]	RNN(Recurrent Neural Network), long short-term memory (LSTM)	2020
5	Md Ashraful Islam Talukder, Sheikh Abujar **et.al**.[34]	Bi-directional RNNs with LSTM in encoding layer and attention model at decoding layer	Jul-19
6	Ratul Sikder, Md. Monowar Hossain, F.M. Rahat Hasan Robi[35]	Mathematical rules and Bengali grammatical rules, text & word tokenization, stemming, stop word removal	2019

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 12
***Article Received:** 25 July 2021 **Revised:** 12 September 2021 **Accepted:** 30 November 2021*

7	Kamal Sarkar[36]	Ranking the candidate summary sentences, TF*IDF, position and sentence length feature	2012
8	Amitava Das, Sivaji Bandyopadhyay[37]	Topic-sentiment model for sentiment identification and aggregation, clustering (k-means) and Document level Theme Relational Graph representation	2010
Kannada			
S.No.	**Researchers/Authors**	**Technology**	**Year**
1	Arpitha Swamy, Srinath S[38]	Summarization based on Clustering(K-means) and The Latent Semantic Analysis, Term-Frequency/Inverse Sentence Frequency (TF/ISF)	2019
2	Arpitha Swamy, Srinath S[39]	Term Frequency-Inverse Sentence Frequency (TF/ISF), Keywords feature, Sentence length and Sentence position.	2019
3	Anusha B. S., Harshitha P, Divya Ramesh, Uma D., Lalithnarayan C.[40]	Rock clustering Algorithm, Naive-Bayes algorithm for classification, word vectorising stemmer, sub-sampling approach	2019
4	Jagadish S Kallimani, Srinivasa K G, Eswara Reddy B[41]	Attribute based Information Extraction (IE) rules and class based Templates, Term Frequency/Inverse Document Frequency (TF/IDF)	2014
5	Varsha R Embar, Surabhi R Deshpande, Vaishnavi A K, Vishakha Jain, Jagadish S Kallimani[42]	Parts of speech tagging and stemming operations, identification of named entities, location and date, usage of abstraction schemes and Information Extraction(IE) rules.	2013
6	Jayashree R , Srikantamurthy K and Basavaraj S Anami[43]	TF (Term Frequency) and Inverse Document Frequency (IDF), Stop word removal	2013
7	Jayashree R , Srikantamurthy K and Basavaraj S Anami[44]	Feed forward neural network architecture	2013
8	Jagadish S. Kallimani, K. G. Srinivasa, Eswara Reddy B.[45]	Keyword Extraction scoring tokenizing, Sentence ranking	2012
9	Jayashree R, Srikanta Murthy K, Basavaraj .S.Anami.[46]	Key words extraction from pre-categorized Kannada documents, GSS (Galavotti, Sebastiani, Simi) coefficients and IDF (Inverse Document Frequency) methods along with TF (Term Frequency) for extracting key words	2012

10	Jagadish S Kallimani, Srinivasa K G, Eswara Reddy B[47]	Keyword Extraction scoring tokenizing, Sentence ranking	2010

Challenges in Text Summarization for Indian Languages

Linguistic Diversity Indian languages belong to different language families—Indo-Aryan, Dravidian, Tibeto-Burman, and Austroasiatic. This diversity results in significant syntactic, morphological, and semantic variations. For instance, word order (SOV vs SVO), complex morphology, and script variation across languages complicate the creation of a single summarization model that works universally. Developing customized models for each language remains a key challenge [48], [49].

Resource Scarcity Many Indian languages lack large annotated corpora, lexical databases, POS taggers, or syntactic parsers required for training summarization models. While efforts like IndicNLP provide foundational tools, high-quality parallel and monolingual datasets are limited compared to English or other global languages [48], [50], [51].

Limited Pre-trained Language Models While models like BERT have revolutionized NLP, their fine-tuned or native counterparts for Indian languages are relatively nascent. Pre-trained models like IndicBERT and MuRIL cover only a subset of major Indian languages, limiting their applicability to low-resource languages [51], [52].

Scarce Multilingual and Cross-lingual Resources In the Indian context, documents often appear in code-mixed forms or with multiple language variants. Existing summarization systems struggle with cross-lingual understanding and summarization, particularly for languages with insufficient alignment data or bilingual corpora [53], [54].

Evaluation and Benchmarking Issues There is a lack of standardized evaluation datasets, ROUGE/NLP evaluation tools tailored for Indian languages. Consequently, comparing results across summarization systems or improving models systematically becomes difficult.

Lack of Annotated Datasets for Abstractive Summarization While extractive summarization has seen modest success, abstractive summarization—requiring deep semantic understanding and rephrasing remains largely unexplored due to the absence of human-curated abstractive summary datasets.

Future Research Directions and Innovations

Future research in Indian language text summarization should focus on multilingual and cross-lingual systems by leveraging multilingual pre-trained models like IndicBERT, enabling transfer learning from high-resource to low-resource languages such as Manipuri or Konkani. The development of large-scale, annotated corpora across diverse Indian languages is essential for training robust summarization models, and partnerships with academic and governmental bodies can facilitate this. Incorporating cultural context, idiomatic expressions, and regional variations into models can improve the relevance and interpretability of summaries. Addressing the scarcity of data through low-resource learning approaches like few-shot and zero-shot learning, coupled with transfer learning, can expand model applicability across underrepresented languages. Another critical direction is handling code-mixed and informal language common on social media and conversational platforms, which necessitates advanced models capable of understanding such linguistic blends. Additionally, domain-specific and real-time summarization tools are needed for sectors like news, education, law, and healthcare, tailored to domain-specific vocabulary and usage.

Conclusion

This survey provides an in-depth exploration of the current state of text summarization systems for Indian languages, with a focus on Hindi, Tamil, Marathi, Punjabi, Bengali, and Kannada. The diverse linguistic landscape of India presents both opportunities and challenges for summarization systems, as each language exhibits unique syntactic, morphological, and semantic characteristics. The significance of this survey lies in its emphasis on the importance of adapting summarization models to accommodate these language-specific features, which is crucial for achieving high-quality, contextually relevant summaries.

The advancement of text summarization systems in Indian languages will play a vital role in enhancing information accessibility, particularly in domains such as education, news, healthcare, and legal services. As research progresses, it will become increasingly important to develop solutions that not only improve the accuracy and efficiency of summaries but also ensure that they are linguistically and culturally appropriate. Moreover, the integration of multilingual and cross-lingual models, coupled with the development of large-scale annotated corpora, will drive future advancements and ensure that the benefits of summarization technologies reach underrepresented languages.

References

[1] A. Das and S. Bandyopadhyay, "Developing Multilingual Lexical Knowledge Base: Indian Languages Perspective," *Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP)*, 2010, pp. 35–42.
[2] P. Bhattacharyya, "Indian Language NLP: Resource Building and Challenges," *Proceedings of the Fifth International Conference on Knowledge Discovery and Information Retrieval (KDIR)*, 2013, pp. IS-11–IS-18.
[3] S. S. Saggion, "Automatic Text Summarization," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–179, 2017, doi: 10.2200/S00762ED1V01Y201703HLT037.
[4] K. R. Rajyaguru and P. Majumder, "A Survey on Text Summarization for Indian Languages," *Proceedings of the 2017 International Conference on Big Data Analytics (BDA)*, Hyderabad, India, 2017, pp. 91–103, doi: 10.1007/978-3-319-72347-8_8.
[5] T. S. Kamath and V. C. Dharwadkar, "Challenges in Indian Language Text Summarization: A Review," *International Journal of Computer Applications*, vol. 975, pp. 8887, 2015.
[6] S. Gupta and D. Saini, "Automatic Text Summarization for Indian Languages: A Review," *International Journal of Computer Applications*, vol. 975, no. 8887, pp. 32–38, 2016.
[7] S. B. Bhandari and S. Patil, "Survey of Extractive and Abstractive Text Summarization Techniques," *Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2020, pp. 433–438, doi: 10.1109/ICICCS48265.2020.9120895.
[8] A. Gupta and S. Sharma, "A Survey on Text Summarization Techniques and Research Challenges," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 4, no. 5, pp. 9022–9028, 2016.
[9] P. Majumder and K. R. Rajyaguru, "Indian Language Summarization: Perspectives and Challenges," *In Proceedings of the Forum for Information Retrieval Evaluation (FIRE)*, 2016, pp. 23–26.
[10] P. Bhattacharyya, "Resource Creation and Evaluation for Indian Languages: Challenges and Perspectives," *Computational Linguistics in the South Asian Context*, Springer, 2015, pp. 19–40, doi: 10.1007/978-3-319-11140-9_2.
[11] Thevatheepan Priyadharshan and Sagara Sumathipala, "Text Summarization for Tamil Online Sports News Using NLP," in Proc. IEEE Conf., Dec. 2018.
[12] Syed Sabir Mohamed and Shanmugasundaram Hariharan, "A Summarizer for Tamil language Using Centroid Approach," Int. J. Inf. Retr. Res., vol. 6, no. 1, pp. 1-14, Jan.–Mar. 2016.
[13] Sankar K, Vijay Sundar Ram R, and Sobha Lalitha Devi, "Text Extraction for an Agglutinative Language," Lang. India, vol. 11, no. 5, May 2011.
[14] J. Jai Hari Raju, P. Indhu Reka, K.K Nandavi, and Dr. Madhan Karky, "Tamil Summary Generation for a Cricket Match," in Proc. INFITT Conf., 2011.
[15] M. Banu, C. Karthika, P. Sudarmani, and T.V. Geetha, "Tamil Document Summarization Using Semantic Graph Method," in Proc. IEEE ICCIMA, Dec. 2007.
[16] A. D. Dhawale, S. B. Kulkarni, and V. M. Kumbhakarna, "Automatic Pre-Processing of Marathi Text for Summarization," International Journal of Engineering and Advanced Technology, vol. 10, no. 1, pp. 566-571, Oct. 2020.
[17] A. Chaudhari, A. Dole, and D. Kadam, "Marathi text summarization using neural networks," International Journal of Advance Research and Development, 2019.
[18] Y. V. Rathod, "Extractive Text Summarization of Marathi News Articles," International Research Journal of Engineering and Technology, vol. 5, no. 7, pp. 1124-1129, July 2018.
[19] V. V. Sarwadnya and S. S. Sonawane, "Marathi Extractive Text Summarizer using Graph Based Model," in Proc. IEEE Conference, Aug. 2018.
[20] S. A. Narhari and R. Shedge, "Text categorization of Marathi documents using modified LINGO," in Proc. IEEE Conference, Dec. 2017.
[21] P. Bolaj and S. Govilkar, "Text Classification for Marathi Documents using Supervised Learning Methods," International Journal of Computer Applications, vol. 155, no. 8, pp. 16-21, Dec. 2016.
[22] N. Dangre, A. Bodke, A. Date, S. Rungta, and S. S. Pathak, "System for Marathi News Clustering," Procedia Computer Science, vol. 89, pp. 465-470, 2016.
[23] S. Vijay, V. Rai, S. Gupta, A. Vijayvargia, and D. M. Sharma, "Extractive Text Summarisation in Hindi," in Proc. IEEE Conference, 2017.
[24] V. Dalal and L. Malik, "Data Clustering Approach for Automatic Text Summarization of Hindi Documents using Particle Swarm Optimization and Semantic Graph," International Journal of Soft Computing and Engineering, vol. 7, no. 3, pp. 22-27, July 2017.
[25] S. Sargule and R. M. Kagalkar, "Strategy for Hindi Text Summarization using Content Based Indexing Approach," International Journal of Computer Sciences and Engineering, vol. 4, no. 9, pp. 1-6, Sept. 2016.
[26] K. V. Kumar and D. Yadav, "An Improvised Extractive Approach to Hindi Text Summarization," in Information Systems Design and Intelligent Applications, Springer, pp. 291-298, Jan. 2015.
[27] A. Jain, D. Yadav, and A. Arora, "Particle Swarm Optimization for Punjabi Text Summarization,"

International Journal of Operations Research and Information Systems, vol. 12, no. 4, pp. 109-122, 2021.
[28] V. Gupta and G. S. Lehal, "Automatic Punjabi Text Extractive Summarization System," in Proc. COLING Conference, Dec. 2012.
[29] V. Gupta and G. S. Lehal, "Feature Selection and Weight Learning for Punjabi Text Summarization," International Journal of Trends and Technology, vol. 1, no. 3, pp. 35-40, 2011.
[30] R. R. Chowdhury, M. T. Nayeem, T. T. Mim, M. S. R. Chowdhury, and T. Jannat, "Unsupervised Abstractive Summarization of Bengali Text Documents," Assoc. for Computational Linguistics, Feb. 2021.
[31] P. Bhattacharjee, A. Mallick, M. S. Islam, and M. E. Jannat, "Bengali Abstractive News Summarization (BANS): A Neural Attention Approach," in Trends in Computational and Cognitive Engineering, Springer, Dec. 2020.
[32] M. Islam, F. N. Majumdar, A. Galib, and M. M. Hoque, "Hybrid Text Summarizer for Bangla Document," International Journal of Computer Vision and Signal Processing, vol. 10, no. 1, pp. 12-17, 2020.
[33] S. Abujar, A. K. M. Masum, M. S. Islam, F. Faisal, and S. A. Hossain, "A Bengali Text Generation Approach in Context of Abstractive Text Summarization Using RNN," in Innovations in Computer Science and Engineering, Springer, Mar. 2020.
[34] M. A. I. Talukder, S. Abujar, et al., "Bengali Abstractive Text Summarization Using Sequence to Sequence RNNs," in Proc. IEEE Conference, July 2019.
[35] R. Sikder, M. M. Hossain, and F. M. R. H. Robi, "Automatic Text Summarization for Bengali Language Including Grammatical Analysis," International Journal of Scientific & Technology Research, vol. 8, no. 6, pp. 987-990, June 2019.
[36] K. Sarkar, "An Approach to Summarizing Bengali News Documents," in Proc. ICACCI, ACM, pp. 1270-1274, 2012.
[37] A. Das and S. Bandyopadhyay, "Topic-based Bengali Opinion Summarization," in Proc. COLING Conference, ACM, Aug. 2010.
[38] A. Swamy and S. Srinath, "A Fused Extractive Summarization Approach for Kannada Text Documents," International Journal of Advanced Science and Technology, vol. 28, no. 16, pp. 368-376, 2019.
[39] A. Swamy and S. Srinath, "Automated Kannada Text Summarization using Sentence Features," International Journal of Recent Technology and Engineering, vol. 8, no. 2, pp. 660-663, July 2019.
[40] A. B. S. Anusha, H. P. Harshitha, D. Ramesh, U. D., and C. Lalithnarayan, "Multi-Classification and Automatic Text Summarization of Kannada News Articles," International Journal of Computer Applications, vol. 181, no. 38, pp. 18-23, Jan. 2019.
[41] J. S. Kallimani, K. G. Srinivasa, and B. E. Reddy, "A Comprehensive Analysis of Guided Abstractive Text Summarization," International Journal of Computer Science, vol. 11, no. 4, pp. 45-50, 2014.
[42] V. R. Embar, S. R. Deshpande, V. A. K., V. Jain, and J. S. Kallimani, "sArAmsha - a Kannada Abstractive Summarizer," in Proc. IEEE Conference, Aug. 2013.
[43] J. R., K. Srikantamurthy, and B. S. Anami, "A New Approach to Summarization in the Kannada Language by Sentence Ranking," Journal of Network and Innovative Computing, vol. 1, no. 1, pp. 123-129, 2013.
[44] J. R., K. Srikantamurthy, and B. S. Anami, "An Artificial Neural Network Approach to Text Document Summarization in the Kannada Language," in Proc. IEEE Conference, Dec. 2013.
[45] J. S. Kallimani, K. G. Srinivasa, and B. E. Reddy, "Summarizing Newspaper Articles: Experiments with Ontology Based, Customized, Extractive Text Summary and Word Scoring," Cybernetics and Information Technologies, vol. 12, no. 2, pp. 85-94, 2012.
[46] J. R., K. Srikantamurthy, and B. S. Anami, "Categorized Text Document Summarization in the Kannada Language by Sentence Ranking," in Proc. IEEE Conference, Nov. 2012.
[47] J. S. Kallimani, K. G. Srinivasa, and B. E. Reddy, "Information Retrieval by Text Summarization for an Indian Regional Language," in Proc. IEEE Conference, 2010.
[48] A. Kunchukuttan, "AI4Bharat IndicNLP Library: NLP Resources for Indic Languages," *GitHub Repository*, 2020.
[49] P. Goyal and A. Bhatia, "Exploring Pretrained Language Models for Indian Languages," in *Proc. Workshop on NLP for Less Resourced Languages (LRL)*, 2021, pp. 1–10.
[50] S. Kakwani et al., "IndicCorp: A Multilingual Corpus for Indic Languages," in *Proc. Findings of ACL: EMNLP*, 2020, pp. 2140–2150.
[51] A. Bhattacharjee et al., "Using BERT for Multilingual Abstractive Summarization in Indic Languages," in *Proc. 1st Workshop on Multilingual Representation Learning*, 2021.
[52] R. Chauhan, P. M. Sharma, and D. S. Roy, "A Comparative Survey of NLP Tools for Indian Languages," *International Journal of Computer Applications*, vol. 178, no. 12, pp. 6–10, 2019.
[53] V. Gupta and G. S. Lehal, "A Survey of Text Summarization Extractive Techniques for Indian Languages," *Journal of Emerging Technologies in Web Intelligence*, vol. 5, no. 3, pp. 263–269, 2013.
[54] A. Ekbal and S. Bandyopadhyay, "Automatic Text Summarization: A Survey of the State of the Art," *Journal of Computer and System Sciences*, vol. 7, no. 1, pp. 1–10, 2015