



International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 4 (2026)



ijerst.editor@gmail.com
editor@ijerst.com

Research Paper

Semantic Aware Text Summarization Through LaBSE Feature Learning with Omni Score Classification

D.Uma Vishveshwar

Assistant Professor, Department of
Computer Science and Engineering
CMR Engineering College
Hyderabad, Telangana
uma.vishwam@cmrec.ac.in

Datla Gnanpreeth Reddy

UG Student, Department of Computer
Science and Engineering
CMR Engineering College
Hyderabad, Telangana
gnanpreethreddydatla@gmail.com

Mulla Akbar Shehazad

UG Student, Department of Computer
Science and Engineering
CMR Engineering College
Hyderabad, Telangana
mullaakbarshehazad@gmail.com

Mohammad Ayesha Rizwana

UG Student, Department of Computer
Science and Engineering
CMR Engineering College
Hyderabad, Telangana
ayesharizwanamohammad@gmail.com

G.Sumalatha

Associate professor, Department of
Computer Science and Engineering
CMR Engineering College
Hyderabad, Telangana
suma.udurukota@gmail.com

Abstract— The need for text summarization has become a significant task in many real-world applications such as news monitoring, academic literature review, business analytics, legal document analysis, healthcare information management, and digital content recommendation. To address these drawbacks, this study introduces a new text summarization system built on several benchmark datasets, such as BBC and XSUM, to achieve a more general approach and performance for various document types and styles. The framework starts with the Natural Language Processing (NLP) pre-processing phase, which involves cleaning, denoising, and normalizing the textual data and converting it into an appropriate format for further analysis. After preprocessing, the Language agnostic BERT Sentence Encoder (LaBSE) is used to produce high-dimensional semantic vector representations of both sentences that accurately represent the linguistic and textual relationships. Sentiment Named Entity Recognition (SNER) Feature Ranking is proposed to better select the important features of the information, which combines sentiment characteristics and named entity information while identifying and ranking the features. Lastly, the Weighted Omni Score (WOS) Classifier assesses and ranks the extracted content in terms of its informational value, resulting in the production of coherent, concise and contextually appropriate summaries. The proposed framework incorporates the integration of semantic understanding, sentiment appreciation, and intelligent feature prioritization to enhance the relevance of the summaries, the information retained, the readability, and summarization effectiveness in various types of textual datasets.

Keywords— Automated Summarization, Language agnostic BERT Sentence Encoder, Natural Language Processing, Sentiment NER, Weighted Omni Score Classifier

I. INTRODUCTION

With the explosion of digital information, large amounts of textual information are challenging to manage and understand [1]. Studies have shown that there are few companies that have nearly 80% of their information in unstructured formats like news articles, reports, emails, research documents and social media content, and the rate of growth is accelerating yearly [2]. Moreover, with the increasing rate of data generation globally, it is becoming more challenging for individuals and organisations to obtain information that is relevant to them efficiently [3]. Therefore, automatic text summarization has become a vital research direction for effectively condensing lengthy documents [4] into any informative and condensed summary while retaining relevant contextual information.

The most important technology companies and information service providers use text summarization technologies extensively to enhance user productivity and information access. In the realm of enterprise solutions, Google AI's [5] AI summarization feature in Gemini streamlines documents, reports, emails, and research into digestible insights. Likewise, Microsoft Copilot has a feature for summarizing articles, business documents, and workplace messages, and Amazon Web Services offers automated document summarization for enterprise applications. They are now being used in news aggregation, healthcare documentation, legal analysis, academic research, business intelligence and customer support systems where the need to extract information quickly is key.

Despite these developments, summarization is still a time-consuming, labor-intensive and subjective task which can give inconsistent results for different users [8]. Most automatic paradigms have difficulty extracting semantic relationships,

contextual information, sentiments, and significant named entities from complex documents such that information is lost and the quality of the summaries is compromised [9]. Furthermore, with the increasing amount and variety of textual information, more intelligent mechanisms for feature extraction and content prioritization are required.

II. LITERATURE SURVEY

Amarendra et al. [10] proposed a framework of Artificial Intelligence-based NLP-based Chatbot for Document Query Resolution and Text Summarization in Healthcare Environments. Tyagi et al. [11] performed a comparative study on the transformer-based architectures for summarizing legal texts by testing several deep-learning models on complex legal texts. Mushtaq and Veningston [12] introduced a thorough biomedical text summarization system using large language models to boost information extraction from biomedical texts.

To solve this problem, Tang et al. [13] proposed the P2N2S framework which integrates natural language and programming language understanding in code summarization. A methodology was developed based on the use of large language models to produce meaningful summaries of source code, distilling both syntactic and semantic relationships. Patel et al. [14] proposed a legal text summarization model using optimized hybrid architectures and fine-tuned LLaMA-2 network. In the context study, Sharma et al. [15] gave an overview of the automatic text summarization techniques with natural language processing methodology.

Kadhim et al. [16] have proposed an advanced text summarization model that uses NLP techniques and feature-based content selection scoring mechanisms. Verma et al. [17] proposed an abstractive Hindi Text Summarization Model using optimization techniques and Generative Artificial Intelligence (AGI) and Bidirectional Long Short-Term Memory (Bi-LSTM) Classifier. Abadi and Ghasemian [18] designed a three-step fine-tuning process along with reinforcement learning to improve the performance of Persian text summarization with mT5 transformer. In Arabic aspect-based sentiment analysis and abstractive traffic service data summarization, Alotaibi and Nadeem [19] proposed an unsupervised integrated system for these tasks. Santosa et al. [20] explored abstractive news text summarization from Indonesia to obtain a summary that is contextually meaningful and fluent using BART, a transformer architecture.

III. PROPOSED METHODOLOGY

As illustrated in Figure 1, the proposed algorithm is called Semantic Sentiment Guided Weighted Omni Summarization (SSGWOS), which is a new combination of several complementary techniques, which are not extensively investigated in existing text summarization surveys. The textual documents from BBC and XSUM datasets undergo advanced NLP preprocessing in the first step, which involves normalization, token refinement, stop word elimination and contextual cleaning to create textual consistency. Then, LaBSE is used to produce multi-lingual sentence and document level semantic embeddings to capture the depth of contextual relationships between sentences and documents. The proposed framework goes beyond semantic representations by introducing a novel SNER Feature Ranking mechanism that considers sentiment polarity, emotional relevance, entity prominence, and contextual relevance to

weight the features of text that are most informative. The ranked features are subsequently passed through the proposed WOS strategy, a new multi-factor scoring approach that combines semantic relevance, sentiment influence, named entity contribution, contextual coherence and information density in a unified scoring mechanism to produce a dynamic ranking result. Lastly, the top-ranked sentence representations are chosen, and are then arranged in a coherent and information-dense manner to create concise, coherent, and informative summaries. The novelty of the proposed SSGWOS algorithm is resides in the synergistic collaboration of semantic encoding of LaBSE, named entity ranking based on sentiment and omni-dimensional weighted scoring to generate the summary, which preserves the context better than other algorithms, and keeps entities as well, and ranks them based on their semantic relevance, which in turn covers the content better than other algorithms.

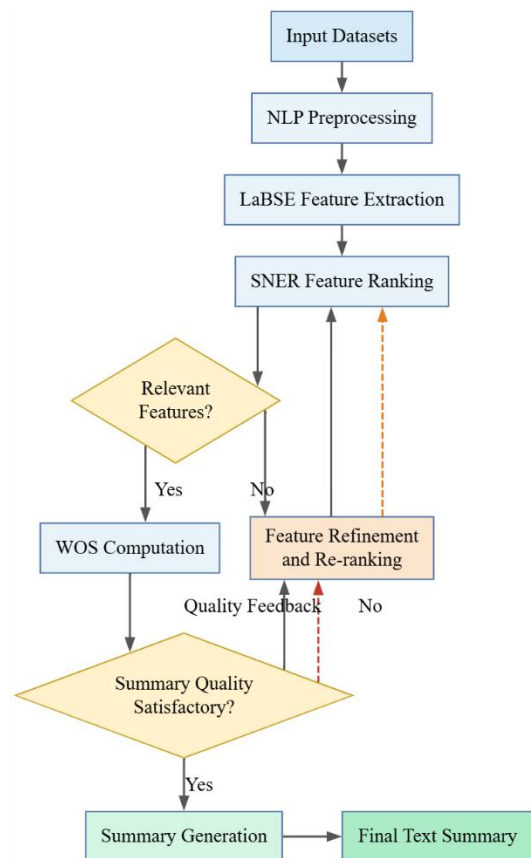


Fig. 1. Proposed System Architecture.

A. LaBSE Feature Extraction

The The LaBSE Feature Extraction in Figure 2 offers several benefits to more advanced applications of text summarization. It generates long context vectors that can represent the semantic context better than just the words in the text and can provide a reasonable approximation of relations in complex text. It can be multilingual and language-agnostic and can be applied to many kinds of document collections, from different domains.

Step 1: Domain Specific Document Acquisition: Acquire textual documents for specific domains, such as research papers, social media posts, legal documents, business reports, customer reviews or news articles. The ability to do this allows the summarization framework to process the

heterogeneous information in the text and retain the semantic consistency of the various types of documents.

Step 2: Document Standardization: Once they have been acquired, the textual documents are standardized by performing some pre-processing operations, including case normalisation, punctuation refinement, whitespace correction and the elimination of irrelevant symbols. This is where the textual content is structured in a uniform structure and semantic representation is performed.

Step 3: Sentence Boundary Identification: Sentence segmentation techniques are used to divide the standardized document into meaningful sentences. Each sentence becomes a semantic unit of its own, which contributes to the meaning of the document.

Step 4: Token Representation Generation: Contextual Tokens represent the linguistic structure and semantic relationships within a given sentence. LaBSE does not just look at the words themselves but considers the relations between words that are close by and their context surrounding those words.

Step 5: Contextual Semantic Encoding: In this step, LaBSE converts the tokenized sentences into high-dimensional representations of context. The encoder is based on deep bidirectional language modelling which allows it to learn the context of the previous and next words.

Step 6: Cross Sentence Relationship Modeling: Sentence embeddings generated in Step 5 are used to find semantic relationship between various sentences in the document. Sentences that are semantically similar to each other (same thing, same event, same entity, same idea or opinion) are grouped together.

Step 7: Information Density Assessment: The semantic representations are further analyzed to assess the information density of each sentence. More relevant sentences tend to have semantic relevance when they include facts, important subjects, significant events, or key context.

Step 8: Redundancy Detection and Filtering: There is repetitive information in many documents, which are presented in different word patterns. Redundant sentences are detected and removed based on sentences similarity measures using LaBSE embeddings.

Step 9: Semantic Feature Consolidation: The enhanced sentence-level embeddings are combined into a complete semantic feature space for the entire document. This single representation retains context, topic distributions, entity interactions and semantic relationships. The resulting feature set is thus a sound basis for further feature ranking and feature selection mechanisms.

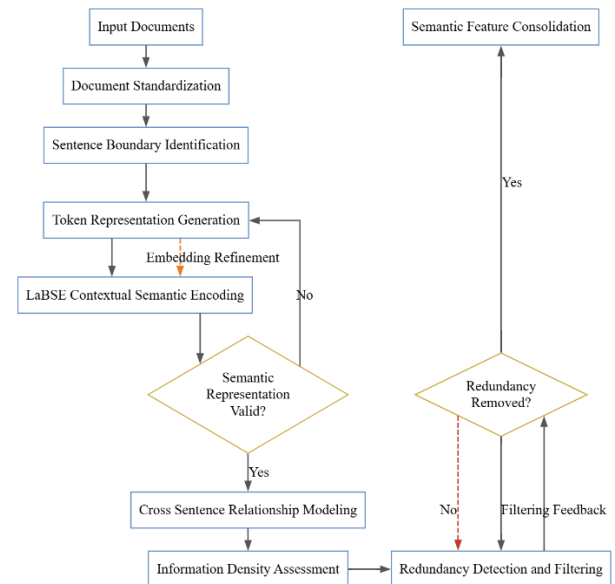


Fig. 2. Proposed LaBSE Feature Extraction Architecture.

B. SNER with WOS Classifier

The SNER with WOS Classifier as illustrated in Figure 3 framework has proven to be advantageous in application-specific summarization tasks such as news analytics, healthcare documentation, legal intelligence, business reporting, analysis of academic literature, and customer opinion mining. Sentiment data and named entity significance work together to effectively highlight domain-relevant content without missing out on the important information in the context.

Step 1: Application Specific Semantic Feature Acquisition: It involves obtaining a set of semantic feature vectors from textual documents from a particular application domain, e.g., news articles, healthcare reports, legal documents, business intelligence records, scientific publications, and customer feedback repositories.

Step 2: Sentiment Context Identification: The semantic features are analysed to identify the sentiment characteristics in individual sentences. Patterns of positive, negative, neutral and mixed sentiments are identified to capture the emotional meaning of the textual information.

Step 3: NER Processing: The second step is to perform NER and recognize important entities like persons, organizations, locations, products, events, diseases, institutions and such domain-specific terms.

Step 4: Sentiment Entity Association Analysis: This step involves identifying the relationships between the entities identified and their sentiments. This is the step to see if certain entities are associated with significant opinions, events, achievements, risks, or critical observations.

Step 5: SNER Feature Importance Evaluation: Sentiment NER Feature Ranking mechanism is used to jointly analyze extracted sentiment and entity attributes. Sentiment intensity, entity prominence, contextual significance and semantic relevance are used to give each sentence an importance estimation.

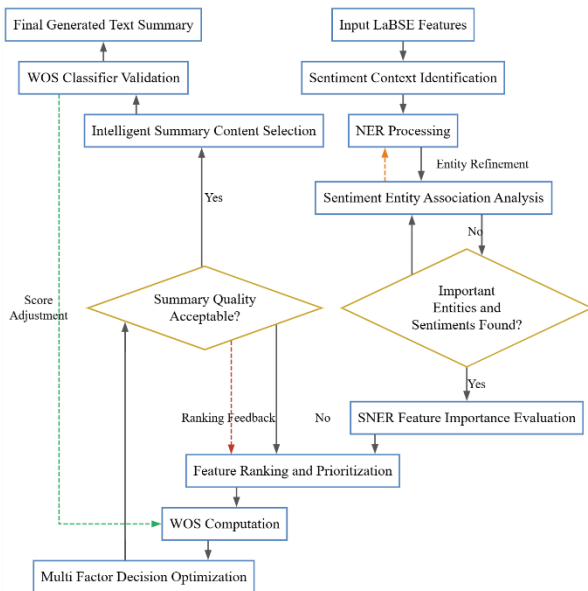


Fig. 3. Proposed SNER with WOS Classifier Architecture.

Step 6: Feature Ranking and Prioritization: The scores of importance generated in the SNER analysis are used to prioritize the potential semantic features available. Informative sentences, with important entities and high contextual relevance, have a higher priority, and less relevant information is considered less important.

Step 7: WOS Computation: The ranked features are passed on to the WOS mechanism. At this stage, it is computing a degree of importance overall, from multiple factors including semantic relevance of the terms, sentiment importance, entity importance, consistency of terms in the context, information density and coverage of the document.

Step 8: Multi Factor Decision Optimization: WOS framework is used to compare alternative textual segments by means of a multi-dimensional decision approach. Sentences are contrasted about their value in advancing the document's meaning.

Step 9: Intelligent Summary Content Selection: The most relevant chunks of text from the optimized WOS are placed in the summary in the frame. Comprising the selected sentences, these are some of the key ideas, entities, feelings, and context relations included in the original text.

Step 10: Final Summary Classification and Generation: During the last step the chosen text is processed with the WOS Classifier to check the quality of the summary, the coherence and the informativeness. A classifier performs the decision, whether the generated summary is relevant and covers the required information. The validated content is then added together to form a short, compact, coherent and context-dependent summary, which can be used for decision-making or knowledge extraction.

IV. RESULTS AND DISCUSSION

In this section, different methods are compared as they have the same data and experimental setup to give a comprehensive evaluation of the various methods. The methods are tested with several test measures to provide a clear understanding of the pros and cons of each method, and how effective they are at completing the task.

A. Datasets

The two benchmark datasets, which are commonly used in summarization research, BBC and XSUM, are used for the proposed text summarization framework evaluation. The BBC data set is composed of 2225 news articles from five different categories: business, entertainment, politics, sport, and technology, covering a variety of themes for assessing document understanding and document condensation. XSUM is a large-scale extreme summarization corpus containing 226,711 BBC news documents with a single-sentence summary written by a human abstractive summarizer, ideal for testing abstractive and extractive summarization. The articles include a wide range of domains, and range from 2010 to 2025, providing a high degree of both linguistic and contextual variation.

B. Hyperparameters

The hyperparameter configuration that was used in the LaBSE feature extraction is shown in TABLE I. The length of the document ranged from 50 to 1000 words, and 512 was found to be a sweet spot that gave a decent amount of context while also being efficient in computation. The same applies to sentence length, which varied from 8 to 128 tokens with 64 tokens being the best semantic representation. The batch size was determined as 32 in a search range of 8-128, and the learning rate of 0.0001 yielded stable convergence, between the minimum learning rate of 0.00001 and the maximum learning rate of 0.01. Best results were obtained after 100 training epochs. The best contextual representation was obtained by a LaBSE embedding dimension of 768; semantic and redundancy thresholds of 0.80 and 0.85, respectively, were shown to be good enough for preserving important information and eliminating repetitive content. In addition, the 0.30 dropout resulted in better model generalization to avoid overfitting the features extracted from the dataset.

TABLE I. HYPERPARAMETERS OF LABSE FEATURE EXTRACTION

Hyperparameter	Minimum Value	Maximum Value	Best Value
Document Lengths	50	1000	512
Sentence Length	8	128	64
Batch Size	8	128	32
Learning Rate	0.00001	0.01	0.0001
Number of Epochs	10	200	100
LaBSE Dimension	128	768	768
Semantic Threshold	0.50	0.95	0.80
Redundancy Threshold	0.50	0.95	0.85
Dropout Rate	0.10	0.60	0.30

The hyperparameters used in the proposed SNER with WOS classifier are summarized in TABLE II. Sentiment weight factor reached its best value at 0.65, while entity importance and context relevance obtained their best values, 0.70 and 0.75 respectively. A weight of 0.60 was an effective way to measure information density. The SNER module was the most discriminative feature ranking when 150 most important features were chosen from 10 to 500 features. Under WOS mechanism, the best results for semantic, sentiment, entity, coverage and coherence weights are 0.80, 0.65, 0.70, 0.75, and 0.85, respectively, which requires a balance of multiple information aspects. Further, concise yet informative summaries were obtained with summary compression ratio set

to 25%, and maximum content relevance and summary quality were obtained with the classification threshold set to 0.80.

TABLE II. HYPERPARAMETERS OF SNER WITH WOS CLASSIFIER.

Hyperparameter	Minimum Value	Maximum Value	Best Value
Sentiment Weight Factor	0.10	1.00	0.65
Entity Importance Weight	0.10	1.00	0.70
Context Relevance Weight	0.10	1.00	0.75
Information Density Weight	0.10	1.00	0.60
SNER Top Ranked Features	10	500	150
WOS Semantic Weight	0.10	1.00	0.80
WOS Sentiment Weight	0.10	1.00	0.65
WOS Entity Weight	0.10	1.00	0.70
WOS Coverage Weight	0.10	1.00	0.75
WOS Coherence Weight	0.10	1.00	0.85
Summary Compression Ratio (%)	10	50	25
Classification Threshold	0.50	0.95	0.80

C. Performance Estimation

The comparative evaluation of the BERT scores of the proposed model is shown on multiple benchmark datasets in TABLE III. The proposed framework got the highest semantic similarity on BBC dataset among the other datasets based on the precision, recall and F1-score of 0.962, 0.951, and 0.956 respectively. Likewise, the XSUM dataset achieved scores of 0.954, 0.946, and 0.950, showing good semantic preservation even for the extreme summarization task. The framework achieved F1-scores of 0.944, 0.938, and 0.943 for CNN-DailyMail, Multi News, and Newsroom datasets, respectively, consistently. The overall average precision, recall and F1 score were 0.951, 0.942 and 0.946 respectively, which proved that the proposed SSGWOS framework could produce summaries close to human written reference summaries in various datasets.

TABLE III. COMPARATIVE EVALUATION OF BERT SCORES FOR PROPOSED MODEL ACROSS DIFFERENT DATABASES

Dataset	Precision	Recall	F1-Score
BBC	0.962	0.951	0.956
XSUM	0.954	0.946	0.950
CNN DailyMail	0.948	0.940	0.944
Multi News	0.942	0.935	0.938
Newsroom	0.947	0.939	0.943
Average	0.951	0.942	0.946

Results of comparing summarization methodologies in terms of accuracy in News, Business, Politics, and Technology are shown in TABLE IV. The traditional graph-based method TextRank [11] and LexRank [13] obtained the accuracy of 82.94% to 86.43%, which shows that they had moderate summarization effectiveness. Transformer-based models like BERTSum [14], PEGASUS [16] and T5-BERT [18] achieved the highest accuracy of up to 94.36% in News category. But the proposed SSGWOS framework consistently outperformed all existing approaches with accuracy of 98.72%, 98.35%, 98.11% and 98.58% in News, Business, Politics and Technology respectively. The findings provide proof of the effectiveness of the application of semantic embeddings, entity ranking based on sentiment and weighted omni scoring to obtain the best summary.

TABLE IV. ACCURACY EVALUATION FOR SUMMARIZATION TECHNIQUES ACROSS DIFFERENT CATEGORIES

Methodology	News	Business	Politics	Technology
TextRank [11]	84.26	83.55	82.94	84.12
LexRank [13]	86.43	85.81	85.34	86.02
BERTSum [14]	91.65	90.92	90.45	91.12
PEGASUS [16]	93.81	93.16	92.95	93.42
T5-BERT [18]	94.36	93.84	93.47	94.05
Proposed SSGWOS	98.72	98.35	98.11	98.58

The comparison of F1-scores of different summarization methods in various content categories is shown in TABLE V. The performance of text classification models was enhanced to 82.31%–85.97% with the F1-scores by TextRank [11] and LexRank [13] and further boosted to 89.88%–93.91% by advanced transformer-based models like BERTSum [14], PEGASUS [16] and T5-BERT [18]. The proposed SSGWOS framework achieved the best F1 in all the categories with 98.44% on the News category, 98.07% on Business, 97.86% on Politics and 98.23% on Technology. The significant advantage compared to previous techniques underscores the potential of the proposed framework to achieve good precision and recall, leading to informative and accurate summarizations.

TABLE V. F1-SCORE EVALUATION FOR SUMMARIZATION TECHNIQUES ACROSS DIFFERENT CATEGORIES

Methodology	News	Business	Politics	Technology
TextRank [11]	83.91	83.07	82.31	83.68
LexRank [13]	85.97	85.23	84.88	85.61
BERTSum [14]	91.08	90.31	89.88	90.76
PEGASUS [16]	93.22	92.74	92.33	92.88
T5-BERT [18]	93.91	93.28	92.95	93.56
Proposed SSGWOS	98.44	98.07	97.86	98.23

TABLE VI measures the readability of the summarization methods based on the standard readability metrics on the BBC, XSUM and CNN DailyMail datasets. The proposed SSGWOS method gave the Flesch Reading Ease value of 78.94 for the BBC dataset which is much higher than TextRank (61.42) and PEGASUS (67.85) methods, demonstrating the ability to make the articles easier to read. Similarly, Flesch Kincaid Grade, Gunning Fog Index and SMOG Index values were lowered to 6.12, 7.45 and 6.91, respectively, from higher complexity scores that were generated by available competing methods. Similar gains were found on the XSUM dataset with a Reading Ease score of 77.63, as well as lower complexity indices of 6.35, 7.68, and 7.05. The proposed approach achieved a score of 78.16 for Reading Ease on CNN DailyMail, and the lowest Flesch Kincaid Grade (6.28), Gunning Fog Index (7.51), and SMOG Index (6.98). The results show that the proposed framework can not only enhance the accuracy of summarization but also generate a summary that is much easier to read and comprehend.

TABLE VI. READABILITY METRICS FOR DIFFERENT DATASETS AND METHODS

Dataset	Method	Flesch Reading Ease	Flesch Kincaid Grade	Gunning Fog Index	SMOG Index
BBC	TextRank	61.42	9.81	11.92	10.44
	PEGASUS	67.85	8.23	10.14	9.07
	SSGWOS	78.94	6.12	7.45	6.91
XSUM	TextRank	59.87	10.05	12.31	10.82
	PEGASUS	65.94	8.58	10.52	9.41
	SSGWOS	77.63	6.35	7.68	7.05

CNN	TextRank	60.71	9.74	11.88	10.36
DailyMail	PEGASUS	66.48	8.41	10.36	9.25
	SSGWOS	78.16	6.28	7.51	6.98

V. CONCLUSION

The experimental results show the high performance of the proposed SSGWOS framework for advanced text summarization on several benchmark sets, including BBC, XSUM, CNN DailyMail, Multi News, and Newsroom datasets, with an average precision of 0.951, recall of 0.942, and F1-score of 0.946. Moreover, SSGWOS outperforms current methods like TextRank [11] with 98.72%, 98.35%, 98.11%, and 98.58% classification accuracy for News, Politics, and Technology categories, respectively, and F1-scores of 98.44%, 98.07%, 97.86%, and 98.23% for the same categories, respectively. Additionally, the proposed framework was found to be superior based on readability analysis, which yielded higher scores in Flesch Reading Ease (78.94, 77.63, 78.16) on BBC, XSUM, and CNN DailyMail datasets, respectively, and low readability complexity indices. The enhancement of these results shows that semantic representation learning, sentiment-aware entity ranking and weighted omni scoring can greatly improve the quality of the summarization, semantic preservation, and readability. This framework can be expanded by integrating large language models, multimodal document summarization (text, images, and speech), cross-lingual summarization for low-resource languages, real-time streaming document summarization, and explainable summarization mechanisms in future studies.

REFERENCES

- [1] Jain, Rajanish Kumar, and Anubha Jain. "Catchphrase Extraction in Natural Language Processing: A Comprehensive Survey of Methods, Challenges." *Mathematics and Logics in Computer Science: Proceedings of ICMLCS 2025* (2026): 332.
- [2] Gunnu, Venkata, Shubham Shah, Anvesh Minukuri, and Jayanth Gopu. "Introduction to LLMs, SLMs, and Modern NLP Challenges." In *Practical Solutions for Modern NLP Challenges: Mastering LLMs and SLMs for Real-World NLP in Cloud and Open-Source*, pp. 1-40. Berkeley, CA: Apress, 2026.
- [3] Shah, Zohaib Ali, Ibrar Ali Shah, Muhammad Usman, and Bilal Khan. "Automatic Text Summarization by Using Deep Learning Models." In *Computational Intelligent Systems: Applications of AI and Machine Learning*, pp. 161-172. Singapore: Springer Nature Singapore, 2026.
- [4] Dhanda¹, Namrata, Vandana Yadav, Ayantika Das, and Aritra Ghosh. "Summarization with Semantic Analysis and Context-Aware Response Suggestions." *Data Science and Applications: Proceedings of ICDSA 2025, Volume 8 8* (2026): 1.
- [5] Ayesha Camran, D., and Afreen Khan. "The Role of LLMs for Optimizing Clinical Insights in Text Summarization." *Intelligent Vision and Computing: Proceedings of ICIVC 2025, Volume 3 3* (2026): 159.
- [6] Sarkar, Sandip, and Mousiki Singha Roy. "Enhancing Text Summarization with Deep Learning: A Seq2Seq Model Approach Using Attention Mechanisms and Stacked-LSTM Networks." *International Journal of Asian Language Processing* (2026).
- [7] Gosai, Krupali, Damodharan Palaniappan, and J. V. R. Ravindra. "Enhancing legal text summarization using fine-tuned and hybrid models on the billsum dataset." In *Artificial Intelligence and Sustainable Innovation*, pp. 634-639. CRC Press, 2026.
- [8] O'Shaughnessy, Douglas. "An Overview of Recent Advances in Natural Language Processing for Information Systems." *Applied Sciences* 16, no. 2 (2026): 1122.
- [9] Luo, Senlin, and Limin Pan. "Big Data and Natural Language Processing (NLP)." *Big Data Analysis: Theory and Technology* (2026): 249-331.
- [10] Amarendra, Ch, P. Nayeem Khan, A. Chandra Sekhar, and P. Praseedhaand K. Srinivas. "AI Powered NLP Chatbot for Efficient Document Query Resolution and Summarization in Healthcare." *Proceedings Copyright* 595: 600.
- [11] Tyagi, Aditi, Himanshu Kumar, Amit Kumar, Vishal Mishra, Ugra Sen, and Ajeet Singh. "Comparative analysis of legal text summarization using transformer-based models." In *Smart Technologies and Intelligent Computing*, pp. 414-421. CRC Press, 2026.
- [12] Mushtaq, Shafiya, and K. Veningston. "Biomedical text summarization with large language models: methodologies, challenges, and future directions." *International Journal of Data Science and Analytics* 22, no. 1 (2026): 29.
- [13] Tang, Yijia, YaoShen Yu, Zhiqiu Huang, Bowei Xia, and Yukun Cao. "P2N2S: Bridging the gap between natural and programming languages for code summarization via large language models." *World Wide Web* 29, no. 1 (2026): 7.
- [14] Patel, Dharil, Shruti Patil, Deepali Vora, and Aniket K. Shahade. "Legal text summarization with optimized hybrid models and fine-tuned LLaMA-2." *Discover Computing* 29, no. 1 (2026): 31.
- [15] Sharma, Neeru, Saravjeet Singh, and Monit Kapoor. "Context Based Review on Auto Text Summarization Techniques Using Natural Language Processing for Performing Evaluation and Comparative Analysis of ATS Techniques." *SN Computer Science* 7, no. 1 (2026): 15.
- [16] Kadhim, Estabraq Abdulreda, Mohammad-Reza Feizi-Derakhshi, and Hadi S. Aghdasi. "Advanced text summarization model incorporating nlp techniques and feature-based scoring." *IEEE Access* 13 (2025): 19302-19319.
- [17] Verma, Babita, Ani Thomas, and Rohit Kumar Verma. "Innovative abstractive Hindi text summarization model incorporating Bi-LSTM classifier, optimizer and generative AI." *Intelligent Decision Technologies* 19, no. 2 (2025): 585-593.
- [18] Abadi, Vahid Nejad Mahmood, and Fahimeh Ghasemian. "Enhancing Persian text summarization through a three-phase fine-tuning and reinforcement learning approach with the mT5 transformer model." *Scientific Reports* 15, no. 1 (2025): 80.
- [19] Alotaibi, Alanoud, and Farrukh Nadeem. "An unsupervised integrated framework for Arabic aspect-based sentiment analysis and abstractive text summarization of traffic services using transformer models." *Smart Cities* 8, no. 2 (2025): 62.
- [20] Santosa, Evan, Alexander Brian Susanto, Henry Wunarsa, and Muhammad Amien Ibrahim. "Exploring Abstractive Indonesian News Text Summarization Using BART Model." *Procedia Computer Science* 269 (2025): 1142-1151.