IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15713

A Scalable Extractive Text Summarisation Model for Hausa Language Documents Using K-Nearest Neighbours and Sparse Matrix Representation

Musa Tanimu Karatu[1], Abdulsamad Muazu[2], Ibrahim Saidu[3]

Department of Computer Science, Federal University, Birnin Kebbi, Nigeria[1,2,3]

Abstract: Extractive text summarisation has become an essential Natural Language Processing (NLP) task for managing the rapid growth of digital textual information. However, conventional graph-based summarisation methods often suffer from high computational complexity and memory consumption due to exhaustive pairwise sentence similarity computations, limiting their applicability to large-scale datasets and low-resource languages such as Hausa. This study proposes a scalable extractive text summarisation model for Hausa language documents using K-Nearest Neighbours (KNN) and Sparse Matrix Representation to improve computational efficiency while preserving summary quality. The proposed approach preprocesses Hausa news articles through tokenisation, stop-word removal, and TF-IDF vectorisation, after which KNN identifies the nearest neighbouring sentences to construct a sparse similarity matrix. This reduces computational complexity from **$O(N^2)$** to **$O(N \log N)$** by eliminating unnecessary similarity computations while preserving the most informative sentence relationships for graph-based ranking. The proposed model was evaluated on a corpus of Hausa news articles and compared with a Graph Convolutional Network–Recurrent Neural Network (GCN–RNN) model using ROUGE-1, ROUGE-2, ROUGE-L, execution time, memory consumption, compression ratio, and computational complexity. Experimental results demonstrate that the proposed model completed summarisation in **32 s** while consuming only **4.79 MB** of memory, compared with **104 s** and **112.96 MB** for the GCN–RNN model, representing reductions of **69.23%** in execution time and **95.76%** in memory usage. It also achieved ROUGE-1, ROUGE-2, and ROUGE-L F1-scores of **85.36%**, **75.64%**, and **76.77%**, respectively, outperforming the baseline while maintaining superior computational efficiency. These findings demonstrate that the proposed KNN with Sparse Matrix Representation provides an efficient, scalable, and practical framework for extractive text summarisation of Hausa language documents and offers a promising solution for other low-resource languages.

Keywords: Extractive Text Summarisation, Hausa Language, Natural Language Processing, K-Nearest Neighbours, Sparse Matrix Representation, TF-IDF, ROUGE Evaluation, Scalable Text Summarisation, Computational Efficiency.

I. INTRODUCTION

Automatic Text Summarization (ATS) aims to condense large volumes of text into concise and informative summaries while preserving the essential information [1]. ATS is broadly categorized into extractive and abstractive summarization. Extractive summarization generates summaries by selecting the most informative sentences from the original document, whereas abstractive summarization produces new sentences that capture the semantic meaning of the source text through natural language generation techniques[2, 3]. These approaches can be applied to both single-document and multi-document summarization tasks, depending on the application domain [4]. A comprehensive classification of ATS methods based on purpose, input, output, and methodology is presented by Jalil, et al. [5] below:

Despite significant advances in ATS, developing models that generalize across different domains remains a major challenge, as summarizing news articles differs considerably from summarizing biomedical, financial, or legal documents [6] Among the existing approaches, graph-based extractive summarization has attracted considerable attention because it models relationships among sentences and ranks them according to their importance.

Several graph-based approaches have demonstrated promising performance. Yongkiatpanich and Wichadakul [6] integrated ontology with the PageRank algorithm to improve sentence ranking and achieved higher ROUGE scores than conventional summarization methods. Similarly, Azadani, et al. [7] proposed a graph-based biomedical summarization framework using itemset mining and sentence clustering, outperforming LexRank [8], TextRank [9], Biochain [10] , and GraphSum [11]. Modified TextRank [12], combined weighting and ranking strategies [13], Sentence-BERT-based graph summarization [14], and supervised graph learning methods [15] have further improved sentence selection accuracy and summary quality.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15713

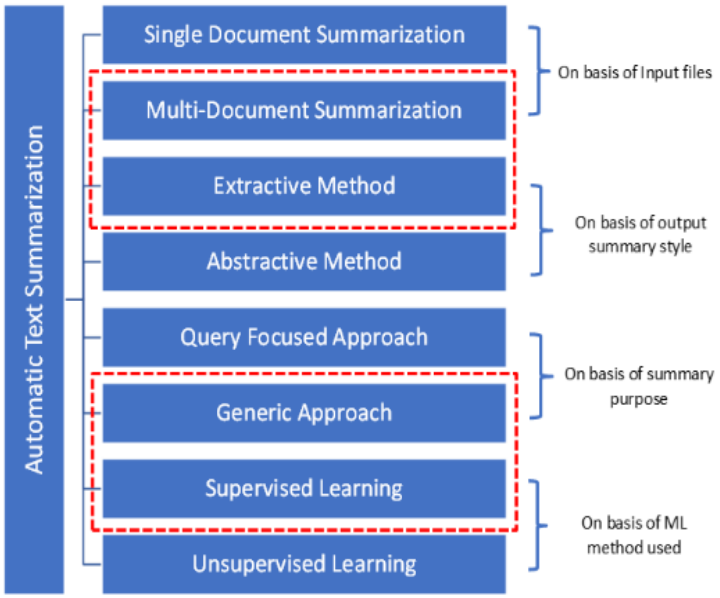

Fig1. Hierarchical representation of categories of ATS (Jalil, et al. [5].

Other studies have incorporated static word embeddings [16], weighted iterative ranking [17], heterogeneous graph attention networks with BERT [18] , graph convolutional networks for multi-document summarization [19], multimodal graph ranking [20], and graph-based approaches for large-scale datasets [21], all demonstrating improved summarization performance.

Recent advances have focused on enhancing graph-based summarization through contextual embeddings and graph neural networks. Domain-specific graph ranking [22] syntactic graph representation learning [23], and multi-granularity heterogeneous graphs [24] have achieved competitive performance on benchmark datasets. Graph-based extractive summarization has also been successfully applied to several languages, including Arabic [25], Urdu Language [26], Marathi Language [27], English language [28]. For Hausa Language [29], proposed a modified PageRank algorithm that ranks sentences based on normalized common bigrams, demonstrating the applicability of graph-based methods to low-resource languages.

Despite these advancements, graph-based extractive text summarization still faces challenges related to scalability, computational complexity, and efficient sentence ranking for large datasets. Kadriua and Obradovic [30] addressed some of these limitations by investigating alternative graph algorithms for more efficient sentence selection. More recently, Anas, et al. [31] proposed a hybrid Graph Convolutional Network (GCN) and Recurrent Neural Network (RNN) framework that improved the coherence and accuracy of Hausa text summaries. Nevertheless, graph-based methods remain computationally expensive when applied to large document collections, often resulting in suboptimal scalability. Therefore, this study proposes a scalable extractive text summarization model for Hausa news articles using a K-Nearest Neighbours (KNN) algorithm with Sparse Matrix Representation for efficient sentence similarity modelling and ranking. The proposed model aims to improve computational efficiency while generating informative and coherent summaries for large-scale Hausa news datasets.

II. METHODOLOGIES

A. Proposed Framework

This study proposes a scalable extractive text summarisation model for Hausa language documents based on K-Nearest Neighbours (KNN) and Sparse Matrix Representation. Unlike conventional graph-based approaches that compute pairwise similarities between all sentences, the proposed model identifies only the nearest neighbouring sentences and stores these relationships in a sparse similarity matrix. This reduces computational complexity and memory consumption while preserving the semantic relationships required for sentence ranking.

The proposed framework comprises five stages: text preprocessing, sentence vector representation, KNN-based sparse matrix construction, sentence ranking, and summary generation, as illustrated in Figure 2. The sparse matrix is

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15713

subsequently used to construct the sentence graph from which the most informative sentences are selected to generate the final extractive summary.

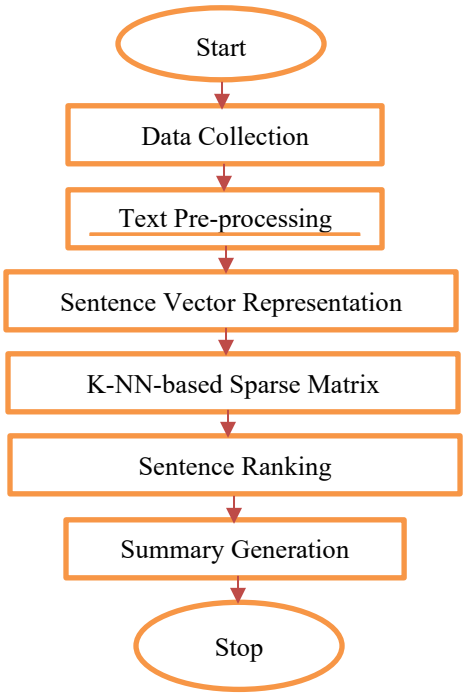

Fig 2: Scalable extractive text summarisation model for Hausa language

B. Dataset and Pre-processing

The experiments were conducted on an updated Hausa news dataset comprising 10,000 articles collected from BBC Hausa, Leadership Hausa, Aminiya, VOA Hausa, RFI Hausa, DW Hausa, and DCL Hausa. The dataset spans multiple domains, including politics, sports, religion, technology, and culture, with each article accompanied by two human-generated reference summaries.

The collected documents were preprocessed through sentence segmentation, tokenization, lowercasing, stop-word removal, and elimination of punctuation and irrelevant symbols to produce normalized sentence representations suitable for feature extraction.

C. Sentence Vector Representation

Each preprocessed sentence is represented using the Term Frequency–Inverse Document Frequency (TF-IDF) weighting scheme. TF-IDF assigns higher weights to terms that occur frequently within a sentence but infrequently across other sentences in the document, thereby emphasizing informative words while reducing the influence of common terms.

The term frequency of term T in sentence S is computed as

$$TF(t,s) = \frac{f(t,s)}{\sum_k f(t,s)}$$

Where:
$TF(t,s)$ Is the term frequency of term t in the sentence S;
$f(t,s)$ Is the number of times term t appears in Sentence S;
$\sum_k f(t,s)$Represent the total number of terms in sentence S;
The second component, **Inverse Document Frequency (IDF)**, measures the uniqueness of a term across all sentences in the document and is computed as:

$$IDF(t) = log\ (\frac{N}{df(t)})$$

Where

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15713

$IDF(t)$= is the inverse document frequency of term t
N= is the total number of sentences in the document
$df(t)$= is the number of sentences containing term t
The TF-IDF weight of a term is then obtained by multiplying its TF and IDF values

$$TF - IDF\ (t,s) = TF(t,s) \times IDF(t)$$

The resulting TF-IDF vectors provide numerical representations of sentence content and serve as the basis for similarity computation.

D. Sparse Matrix Representation

The proposed model employs Sparse Matrix Representation to reduce the computational cost of sentence similarity computation. After preprocessing, sentence similarity is computed using cosine similarity, and the K-Nearest Neighbours (KNN) algorithm identifies the (k) most similar neighbouring sentences for each sentence. Instead of storing similarities between all sentence pairs, only the nearest-neighbour similarities are retained, while the remaining entries are assigned zero values. This produces a sparse similarity matrix that significantly reduces memory usage and computational complexity.

The cosine similarity between two sentence vectors $S(X_i, X_j)$. is computed as

$$S_{ij} = \frac{X_i \,.X_j}{\|X_i\| \, \|X_j\|}$$

Where:

- $X_i \,.X_j$ is the dot product of feature vectors
- $\|X_i\| \; and \; \|X_j\|$ are the Euclidean distance norms (L2 norm) of the vectors

The cosine similarity scores are initially computed to determine sentence neighbourhoods. The KNN algorithm then retains only the k nearest neighbours for each sentence, producing a sparse similarity matrix with substantially lower memory requirements than a dense similarity matrix.

E. Proposed KNN with Sparse Matrix Model

The sparse similarity matrix is used to construct a weighted sentence graph, where each sentence represents a node and each non-zero similarity value represents an edge between neighbouring sentences. Sentence importance is determined based on the weighted connections within the graph, with sentences having stronger relationships assigned higher importance scores.

The importance score of sentence (S_i) is computed as

$$\text{Score}\ (S_i) = \sum_{j=1}^{k} A_i j$$

Where:

- Score (S_i) = importance score of sentence (S_i)
- $A_i j$ = similarity value between sentence $(S_i)$and its j^{th} nearest neighbouring sentence
- k = number of nearest neighbours considered
- $\sum$ = summation of all similarity values from the k nearest neighbours

Sentences are ranked according to their scores, and the top-ranked sentences (20% of the original document) are selected and reordered to generate the final summary.

F. Performance Evaluation

The proposed model is evaluated using ROUGE-1, ROUGE-2, and ROUGE-L by comparing the generated summaries with human-written reference summaries. In addition, the computational efficiency of the proposed KNN-based approach is assessed against conventional graph-based methods to demonstrate its scalability for large Hausa news datasets. However, these evaluation metrics collectively assess both the summarisation quality and computational efficiency of the proposed KNN with Sparse Matrix model.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15713

III. DISCUSSION OF THE EXPERIMENTAL RESULTS

The experimental evaluation compared the Graph Convolutional Network-Recurrent Neural Network (GCN-RNN) ensemble with the proposed cosine similarity against the K-Nearest Neighbours (KNN) with Sparse Matrix approach using four complementary performance dimensions: summary quality, compression capability, computational efficiency, and algorithmic complexity. The results demonstrate that the two approaches optimise different aspects of extractive text summarisation, revealing a clear trade-off between summary conciseness and computational performance. The Table 1 Below show the summary quality between the proposed Approach with GCN-RNN,

TABLE 1: SUMMARY QUALITY

	Metrics	GCN-RNN with Cosine Similarity	KNN with Sparse Metrix	Better Performer
Rouge-1	F1-score	52.94	85.36	KNN
Rouge-2	F1-score	50.24	75.64	KNN
Rouge-L	F1-score	50.00	76.77	KNN
Compression Ratio (%)		41.34	77.06	GCN-RNN (More Concise)
Execution Time (s)		104.00	32.00	KNN
Memory Usage (MB)		112.96	04.79	KNN
Computational Complexity		$O(N^2)$	O(N log N)	KNN

The ROUGE-1, ROUGE-2 and ROUGE-L evaluation of F1-score provides insight into the quality of the generated summaries. The GCN-RNN model achieved F1-scores of 52.94%, 50.24%, and 50.00%, respectively. The KNN with Sparse Matrix approach achieved higher F1-scores of 85.36%, 75.64%, and 76.77%, respectively. Likewise, the KNN model achieved substantially higher recall values across all the ROUGE metrics, indicating that it preserved a greater proportion of the information contained within the reference summaries. This behaviour is consistent with its higher compression ratio, as retaining more of the original document naturally increases the likelihood of matching the reference summaries.

In terms of summary compression, the GCN-RNN ensemble achieved an average compression ratio of 41.34%, indicating that the generated summaries retained approximately 41% of the original document while removing nearly 59% of the source content as shown in Fig 3. In contrast, the KNN with Sparse Matrix approach achieved an average compression ratio of 77.06%, retaining more than three-quarters of the original document. These findings show that the GCN-RNN ensemble generates considerably shorter summaries by selecting fewer sentences from the source document. Such behaviour suggests that the graph-based neural architecture applies a more aggressive sentence selection strategy, which may be advantageous in applications where concise summaries are preferred.

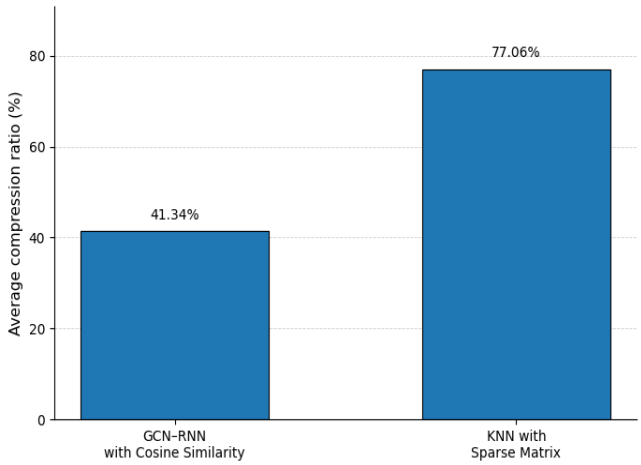

Fig 3: Average Compression Ratio by Summarisation Approach

The evaluation of Precision and Recall for ROUGE-1, ROUGE-2, and ROUGE-L provides additional insight into the quality of the generated summaries. Although, the GCN-RNN model produced considerably shorter summaries, its

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15713

precision values remained consistently high (90.00%, 87.01%, and 85.00% for ROUGE-1, ROUGE-2, and ROUGE-L, respectively), indicating that the information retained in the summaries was generally relevant. Similarly, the KNN model also achieved high precision values (92.50%, 88.58%, and 83.10%), demonstrating that both approaches were effective at selecting relevant sentences despite employing different extraction strategies as shown in the Table 2. Consequently, the principal distinction between the two methods lies not in precision but in information coverage, as reflected by the substantially higher recall achieved by the KNN approach.

TABLE 2: PRECISION AND RECALL EVALUATION

	Metrics	GCN & RNN with Cosine Similarity	KNN with Sparse Metrix
Rouge-1	Precision	90.00	92.50
	Recall	35.50	79.25
Rouge-2	Precision	87.01	88.58
	Recall	34.32	66.00
Rouge-L	Precision	85.00	83.10
	Recall	33.42	71.34

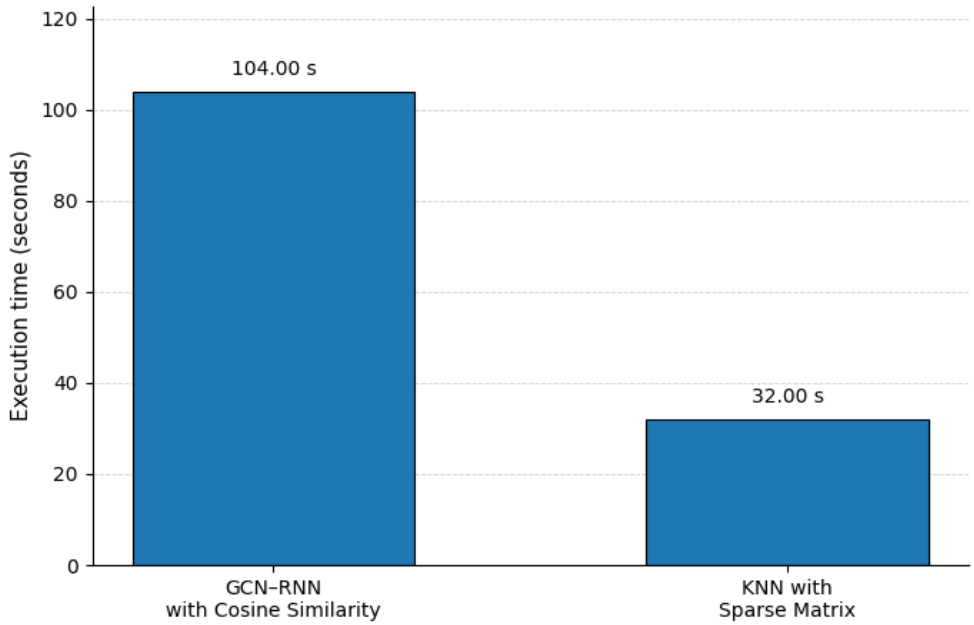

Fig 4: Execution Time by Summarisation Approach

The computational evaluation shown in Fig 4 above, further highlights the contrasting characteristics of the two models. The proposed KNN with Sparse Matrix approach completed the same task in 32 seconds using only 4.79 MB of memory, whereas the GCN-RNN ensemble required 104 seconds to complete the summarization process and consumed 112.96 MB of memory. These results indicate that the KNN approach required substantially fewer computational resources, making it more suitable for large-scale or real-time summarisation applications as shown in Fig 5 below.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15713

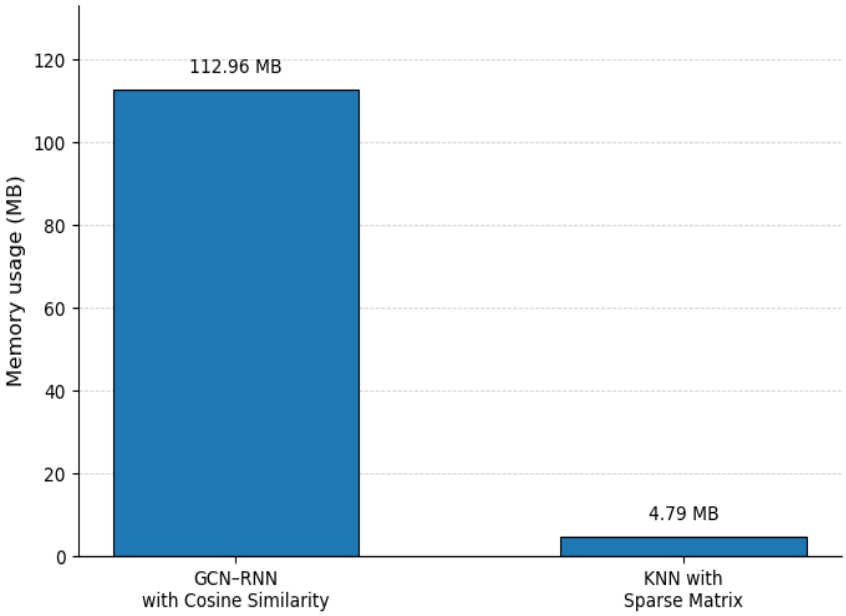

Fig 5: Memory Usage by Summarisation Approach

The observed differences are further supported by the theoretical computational complexities of the two approaches as shown in Fig 6 below. The proposed GCN-RNN ensemble relies on pairwise sentence similarity calculations and graph construction with an approximate computational complexity of $O(N^2)$, resulting in increased processing time and memory consumption as the number of sentences increases. In contrast, the KNN with Sparse Matrix approach reduces the search space by considering only the nearest neighbouring sentences, yielding an approximate complexity of O(N log N). This lower computational complexity contributes directly to its faster execution time and reduced memory requirements.

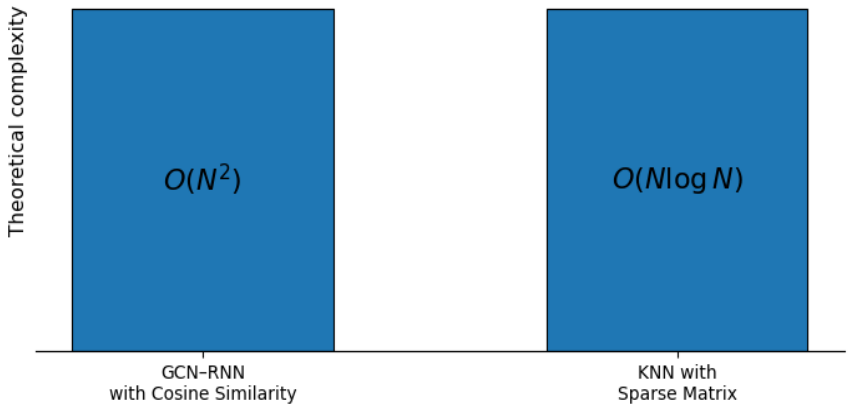

Fig 6: Computational Complexity of the Summarisation Approaches

The results indicate that neither approach is universally superior; rather, each optimises a different objective. The GCN-RNN ensemble demonstrates a stronger ability to generate concise summaries through aggressive content reduction while maintaining high precision. Conversely, the KNN with Sparse Matrix approach achieves higher information retention, superior ROUGE performance, significantly lower execution time, lower memory consumption, and improved computational scalability. Therefore, the choice between the two approaches depends on the intended application. Where concise summaries with substantial redundancy reduction are required, the GCN-RNN ensemble represents a suitable solution. However, where summary completeness, computational efficiency, and scalability are the primary objectives, the KNN with Sparse Matrix approach provides a more appropriate alternative.

DOI: 10.17148/IJARCCE.2026.15713

The central conclusion is that, the experimental results demonstrate a trade-off between summary conciseness and information retention. The proposed GCN-RNN ensemble generates substantially shorter summaries while maintaining high precision, whereas the KNN with Sparse Matrix approach achieves greater information coverage and significantly better computational efficiency. Consequently, the preferred method depends on whether the application prioritises brevity or comprehensive information preservation.

A. Statistical Significance Analysis

To determine whether the observed differences in summarisation performance between the proposed K-Nearest Neighbours (KNN) with Sparse Matrix model and the GCN-RNN model were statistically significant, the Wilcoxon signed-rank test was performed on the paired ROUGE F1-scores. The Wilcoxon signed-rank test was selected because it is a non-parametric statistical test suitable for comparing two related samples without requiring the assumption of normally distributed data.

The results of the statistical analysis indicate that the proposed KNN with Sparse Matrix approach significantly outperformed the GCN-RNN model across the evaluated ROUGE metrics. For ROUGE-1 F1, the Wilcoxon signed-rank test produced a test statistic of $W = 0.000$ with a p-value < 0.001, leading to the rejection of the null hypothesis. Similarly, for ROUGE-2 F1, the test also produced $W = 0.000$ with a p-value < 0.001, indicating a statistically significant difference between the two summarisation approaches. These findings demonstrate that the improvements achieved by the proposed KNN with Sparse Matrix model are unlikely to have occurred by chance and therefore represent genuine performance gains over the GCN-RNN model.

The statistical significance of the ROUGE-1 and ROUGE-2 results provides strong evidence that the proposed optimisation strategy not only improves computational efficiency but also produces higher-quality summaries. When considered alongside the superior ROUGE F1-scores, lower execution time, reduced memory consumption, and lower computational complexity ($O(N \log N)$), the statistical analysis further validates the effectiveness of the proposed KNN with Sparse Matrix approach for scalable extractive text summarisation of Hausa language documents.

IV. CONCLUSION AND FUTURE WORK

This study presented a scalable extractive text summarisation model for Hausa language documents based on K-Nearest Neighbours (KNN) and Sparse Matrix Representation. By retaining only the nearest-neighbour sentence relationships, the proposed approach significantly reduced the computational complexity of graph construction from **$O(N^2)$** to **$O(N \log N)$**, resulting in lower execution time and memory consumption while maintaining high summarisation quality. Experimental evaluation demonstrated that the proposed model outperformed the GCN–RNN baseline in ROUGE-1, ROUGE-2, and ROUGE-L scores, while completing summarisation more efficiently using substantially fewer computational resources. These findings confirm that integrating KNN with Sparse Matrix Representation provides an effective and scalable framework for extractive text summarisation, particularly for low-resource languages such as Hausa.

Despite these promising results, the study is limited to extractive summarisation using a Hausa news corpus. The performance of the proposed model may vary across other document domains and languages with different linguistic characteristics. Furthermore, the use of TF-IDF for sentence representation may not fully capture contextual and semantic information in complex documents.

Future research will focus on extending the proposed framework to abstractive text summarisation by incorporating contextual language models and transformer-based embeddings. The model will also be evaluated on larger multilingual and cross-domain datasets to assess its generalizability.

REFERENCES

[1] M. Bidoki, M. R. Moosavi, and M. Fakhrahmad, "A semantic approach to extractive multi-document summarization: Applying sentence expansion for tuning of conceptual densities," *Information Processing and Management,* vol. 57, p. 102341, 2020.

[2] A. A. Bichi, R. Samsudin, R. Hassan, L. R. A. Hasan, and A. A. Rogo, "Graph-based extractive text summarization method for Hausa text," *PLOS ONE | https://doi.org/10.1371/journal.pone.0285376* vol. 5, pp. 1-15, 2023.

[3] N. Moratanch and S. Chitrakala, "A survey on abstractive text summarization," in *2016 International Conference on Circuit, power and computing technologies (ICCPCT)*, 2016, pp. 1-7.

DOI: 10.17148/IJARCCE.2026.15713

[4] Ramesh and B. Rajan, "Extractive Text Summarization Using Graph Based Ranking Algorithm And Mean Shift Clustering," presented at the In Proceeding of International Conference on Recent Trends in Computing, Communication & Networking Technology(ICRTCCNT) Chennai, Tamil Nadu, 2019.

[5] Z. Jalil, M. Nasir, M. Alazab, J. Nasir, T. Amjad, and A. Alqammaz, "Grapharizer: A Graph-Based Technique for Extractive Multi-Document Summarization," *Electronics (MDPI)2023, 12, 1895. https://doi.org/10.3390/electronics12081895,* vol. 12, pp. 1-26, 2023.

[6] C. Yongkiatpanich and D. Wichadakul, "Extractive Text Summarization Using Ontology and Graph-Based Method," presented at the 2019 IEEE 4th International Conference on Computer and Communication Systems, 2019.

[7] M. N. Azadani, N. Ghadiri, and E. Davoodijam, "Graph-based biomedical text summarization: An itemset mining and sentence clustering approach," I. U. o. T. Department of Electrical and Computer Engineering, Isfahan, Ed., ed, 2017, pp. 1-55.

[8] R. Lawrence, H. Hyoil, and B. Ari-D, "BioChain: lexical chaining methods for biomedical text summarization," in *Proceedings of the 2006 ACM symposium on Applied computing*, 2006, pp. 180-184.

[9] M. Rada and T. Paul, "Textrank: Bringing order into text," in *Proceedings of the 2004 conference on empirical methods in natural language processing*, 2004, pp. 404-411.

[10] L. Reeve, H. Han, and A. D. Brooks, "BioChain: lexical chaining methods for biomedical text summarization," in *Proceedings of the 2006 ACM symposium on Applied computing*, 2006, pp. 180-184.

[11] B. Elena, C. Luca, M. Naeem, and F. Alessandro, "GraphSum: Discovering correlations among multiple terms for graph-based summarization," *Information Sciences,* vol. 249, pp. 96-109, 2013.

[12] C. Mallick, A. K. Das, M. Dutta, A. D. Kumar, and A. Sarkar, "Graph-Based Text Summarization Using Modified TextRank," in *In soft Computing in Data Analytics: Proceeding of International Conference on SCDA 2018*, 2019, pp. 137-146.

[13] A. Alzuhair and M. Al-Dhelaan, "An Approach for Combining Multiple Weighting Schemes and Ranking Methods in Graph-Based Multi-Document Summarization," *IEEE Access,* vol. 7, pp. 120375-120386, 2019.

[14] T. Gokhan, P. Smith, and M. Lee, "GUSUM: Graph-Based Unsupervised Summarization using Sentence Features Scoring and Sentence-BERT," in *Proceedings of TextGraphs-16: Graph-based Methods for Natural Language Processing*, 2022, pp. 44–53.

[15] X. Mao, H. Yang, S. Huang, Y. Liu, and L. Rongsheng, "Extractive summarization using supervised and unsupervised learning," *Expert Systems With Applications ELSEVIER,* vol. 133, pp. 173–181, 2019.

[16] U. Barman, V. Barman, M. Rahman, and N. K. Choudhury, "Graph Based Extractive News Articles Summarization Approach leveraging Static Word Embeddings," presented at the 2021 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2021.

[17] S. Chitrakala, N. Moratanch, B. Ramya, C. G. R. Raaj, and B. Divya, "Concept-Based Extractive Text Summarization Using Graph Modelling and Weighted Iterative Ranking," presented at the International Conference on Emerging Research in Computing, Information, Communication and Applications ERCICA 2016, 2017.

[18] M. Umair, I. Alam, A. Khan, I. Khan, N. Ullah, and M. Y. Momand, "Retracted: N-GPETS: Neural Attention Graph-Based Pretrained Statistical Model for Extractive Text Summarization," *Computational Intelligence and Neuroscience, Hindawi,* vol. 2022, pp. 1-15, 2022.

[19] M. Yasunaga, R. Zhang, K. Meelu, A. Pareek, K. Srinivasan, and D. Radev, "Graph-based Neural Multi-Document Summarization," *arXiv:1706.06681v3,* pp. 1-11, 2017.

[20] J. Zhu, L. Xiang, Y. Zhou, J. Zhang, and C. Zong, "Graph-based Multimodal Ranking Models for Multimodal Summarization," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.,* vol. 20, pp. 60-81, 2021.

[21] M. Moradi, M. Dashtib, and M. Samwald, "Summarization of biomedical articles using domain-specific word embeddings and graph ranking," *Journal of Biomedical Informatics, Elsevier* vol. 107, p. 103452, 2020.

[22] T. Vo, "An approach of syntactical text graph representation learning for extractive summarization," *International Journal of Intelligent Robotics and Applications,* vol. 7, pp. 190–204, 2023.

[23] H. Zhao, W. Zhang, M. Huang, S. Feng, and Y. Wu, "A Multi-Granularity Heterogeneous Graph for Extractive Text Summarization," *Electronics* vol. 12, p. 2184, 2023.

[24] Z. Zhang, H. Elfardy, M. Dreyer, K. Small, H. Ji, and M. Bansal, "Enhancing Multi-Document Summarization with Cross-Document Graph-based Information Extraction," in *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 2023, pp. 1696–1707.

[25] E. Reda, B. Gamal, and K. A. El, "Graph-Based Extractive Arabic Text Summarization Using Multiple Morphological Analyzers.," *Journal of Information Science & Engineering,* vol. 36, p. 347, 2020.

[26] A. Nawaz, M. Bakhtyar, J. Baber, I. Ullah, W. Noor, and A. Basi, "Extractive Text Summarization Models for Urdu Language," *Information Processing and Management,* vol. 57, pp. 1-12, 2020.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15713

[27] V. V. Sarwadnya and S. S. Sonawane, "Marathi Extractive Text Summarizer using Graph Based Mode," presented at the Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.

[28] B. Jing, Y. Zeyu, Y. Tao, F. Wei, and T. Hanghang, "Multiplex graph neural network for extractive text summarization," *arXiv preprint arXiv:2108.12870,* 2021.

[29] A. A. Bich, R. Samsudin, R. Hassan, L. R. A. Hasan, and A. A. Rogo, "Graph-based extractive text summarization method for Hausa text," *PLOS ONE https://doi.org/10.1371/journal.pone.0285376,* vol. 18, pp. 1-15, 2023.

[30] K. Kadriua and M. Obradovic, "Extractive approach for text summarization using graphs," *arXiv:2106.10955,* pp. 1-5, 2021.

[31] S. Anas, M. T. Karatu, R. Abubakar, and S. Abdullahi, "A graph convolutional network and recurrent neural network ensemble for extractive text summarisation in the hausa language," *Journal of Engineering Science,* vol. 32, pp. 32 - 46, 2025.

