Journal of Theoretical and Applied Information Technology
30th April 2026. Vol.104. No.8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CONTRASTIVE SENTENCE EMBEDDINGS–BASED ZERO-SHOT DOCUMENT SUMMARIZATION USING SINGLE-OBJECTIVE GENETIC ALGORITHM

KADRY HAMED[1], MOUSTAFA MAHMOUD AHMED ALI[2], YASER MAHER ABDELMONTELB[3], ABDELMGEID AMIN ALI[1]

[1]Computer Science Department, Faculty of Computers and Information, Minia University, Minia, Egypt
[2]Information Systems Department, Faculty of Computers and Information, Minia University, Minia, Egypt
[3]Information Technology Department, Faculty of Computers and Information, Minia University, Minia, Egypt

E-mail: kadry.hamed@mu.edu.eg, moustafa.m.ali@mu.edu.eg, yaser.wazery@minia.edu.eg, a.ali@mu.edu.eg

ABSTRACT

Amid the rapid developments in the field of natural language processing (NLP), few-shot and zero-shot text summarization have gained prominence as effective methods for generating concise summaries from large texts. These methods employ the capabilities of large language models (LLMs) to perform summarization tasks with minimal or no specific training data. Despite the ability of these approaches to generate open-domain, fluent, and coherent summaries, their extractive summarization performance often remains lagging behind approaches that leverage pre-trained language models (PLMs) with fine-tuning. Thus, we propose a zero-shot extractive summarization framework that leverages the capabilities of PLMs (without fine-tuning) combined with other unsupervised learning strategies. This framework combines the open-domain generation abilities of LLMs with the robust, high-precision performance typically associated with fine-tuned PLMs. Recent advancements in attention mechanisms and evolutionary algorithms have demonstrated significant potential in enhancing the quality of document summarization. This work presents a novel extractive open-domain summarization framework called ETS-BGA that integrates attention-based neural networks with Genetic Algorithms (GAs) to generate high-quality summaries, taking into account three summarization dimensions: (i). Content relevance, which is obtained using a multi-head attention mechanism applied to SimCSE (Simple Contrastive Learning of Sentence Embeddings) without fine-tuning, (ii). Content coverage, which is fulfilled using a hybrid technique that incorporates both Named Entity Recognition (NER) and TF-IDF-based keywords extraction, and (iii). Redundancy minimization, which is obtained by computing cosine similarity between sentence embeddings from SimCSE to catch deeper semantic similarity beyond surface-level syntax. This work also presents a comprehensive cross-paradigm comparative analysis covering statistical-based methods, supervised deep learning-based methods, fine-tuned PLMs, and zero/few-shot methods using LLMs. Experiments on CNN/DailyMail and the New York Times benchmark datasets clarify that our proposed framework generates more informative and less redundant summaries, achieving superior ROUGE scores compared to many existing methods. These findings highlight the effectiveness of combining contrastive learning with evolutionary optimization for zero-shot summarization.

Keywords: *Extractive Text Summarization, Sentence Embeddings, Multi-Head Attention, Zero-Shot Learning, Genetic Algorithms*

1. INTRODUCTION

Text Summarization (TS) refers to the task of generating a brief and coherent version of a longer document, while keeping its main ideas [1, 2]. As digital content -such as news articles, scientific literature, blog posts, and social media data-continues to grow rapidly, effective summarization techniques have become crucial for tasks like information retrieval, content analysis, quick decision-making, and knowledge extraction [3].

Journal of Theoretical and Applied Information Technology
30th April 2026. Vol.104. No.8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

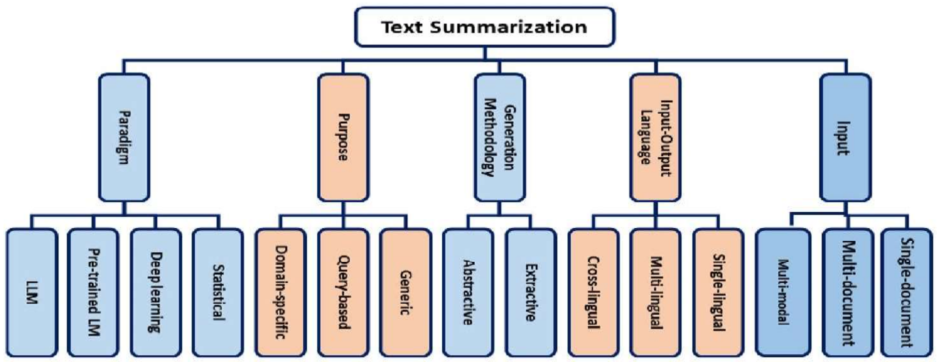

Figure 1: Categorization of Summarization Approaches

As illustrated in Figure 1, TS techniques can be categorized from several perspectives: (i). Based on input type, TS is classified into single-document [4, 5], multi-document [6, 7], and multi-modal summarization [8, 9]. Single-document TS generates a summary from one document, while multi-document TS generates a summary from multiple documents. Multi-modal TS incorporates various data formats—text, images, video, and audio—to produce context-aware summaries. (ii). Based on purpose, summarization can be generic (open-domain) [10, 11], query-based [12], or domain-specific [13, 14]. Generic TS is open-domain, whereas domain-specific and query-based summarization targets specific fields (e.g., biomedical, legal) or specific user queries. (iii). Based on language pairing, TS falls into single-lingual, multi-lingual, and cross-lingual summarization [15]. Single-lingual TS processes and outputs in the same language. Multi-lingual TS supports multiple languages with the same input and output languages, while cross-lingual TS generates summaries in a different language from the input. (iv) Based on the generation method, TS is broadly divided into extractive and abstractive summarization [16]. Extractive TS captures key units (sentences, phrases, or paragraphs) from the original document without modification. In contrast, abstractive methods generate summaries that paraphrase the original document by extracting key terms and knowledge, achieving more flexibility and coherence. (v) Based on the modeling paradigm, TS has evolved through four main phases: statistical, deep learning, fine-tuned pre-trained language models (PLMs), and large language models (LLMs), as illustrated in Figure 2. Statistical methods include heuristic [17], optimization [18], and graph-based techniques [19, 20], which often using TF-IDF and other hand-crafted features [21]. Deep learning approaches employ supervised models trained on document-summary pairs [22, 23, 24, 25], with word embeddings like Word2Vec and GloVe [26, 27]. Fine-tuned PLMs, such as BERT [28] and RoBERTa [29], improved summarization using the “pre-train then fine-tune” paradigm [30, 31]. More recently, LLMs such as GPT-3 [32] have revolutionized TS by enabling zero-shot and few-shot capabilities [33], offering efficient summarization with no or minimal task-specific training.

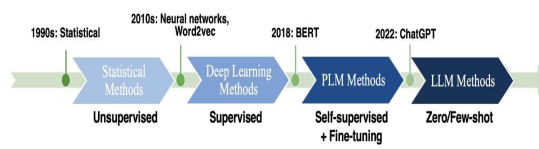

Figure 2: The Evolution of the Four Major Paradigms in TS Research [34]

Despite significant progress in text summarization, most state-of-the-art methods rely heavily on supervised learning and large annotated datasets, which limits their applicability in low-resource or domain-specific scenarios. Consequently, zero-shot summarization remains an open and challenging problem, requiring methods that can generalize without task-specific training.

To address these challenges -building on recent advancements in deep learning and natural language processing (NLP), including sentence embeddings, attention mechanisms, and large-scale PLMs such as BERT and its successors- this work introduces a zero-shot, open-domain, single-document extractive summarization framework that integrates contrastive sentence embeddings with a genetic optimization strategy. First, sentences are encoded using a pre-trained contrastive learning model (SimCSE) to capture deep semantic relationships. Next, a multi-criteria fitness function is designed to evaluate candidate summaries based on relevance, redundancy reduction, and content coverage. A Genetic Algorithm (GA) is then employed as the optimization engine to iteratively evolve candidate summaries towards an optimal solution that is finally chosen to generate the summary. The GA tries to balance multiple conflicting trade-offs: maximizing relevance, minimizing redundancy, and ensuring content coverage, respecting a predefined word limit. The optimization strategy uses tailored genetic operators (selection, crossover, and mutation) designed to maintain solution validity and enhance population diversity, enabling effective exploration of the solution space. This design enables effective summarization without requiring task-specific training, making the approach suitable for diverse and low-resource domains. The key contributions of this work are as follows:

- We present a novel, general-purpose text summarization framework designed for domain adaptability. This zero-shot approach removes the reliance on labeled summarization datasets, making it particularly effective in low-resource or domain-specific settings where annotated data is scarce or unavailable.
- Our framework employs SimCSE-based sentence embeddings to generate semantically rich and context-aware sentence representations, enhancing the quality of extracted summaries.
- A single-objective genetic algorithm is employed to iteratively optimize summary generation based on a unified fitness score that combines relevance, coverage, and redundancy control. This allows for effective sentence selection while strictly adhering to a predefined word count constraint.
- Different from previous work restricted to a single paradigm, we perform an extensive cross-paradigm comparison analysis, including statistical unsupervised methods, supervised deep learning approaches, fine-tuned PLMs, and zero/few-shot LLM summarization techniques.

Experimental results across diverse benchmark datasets reveal that the proposed framework consistently generates high-quality, informative summaries, outperforming standard baselines in terms of informativeness and content coverage.

This work focuses on extractive document summarization in a zero-shot setting, leveraging semantic sentence embeddings and evolutionary optimization. It does not address abstractive summarization or supervised fine-tuning techniques, also it supports monolingual summarization only.

The remainder of this paper is structured as follows: Section 2 presents a comprehensive review of related work. Section 3.1 formally defines the text summarization problem with a focus on its optimization formulation. The proposed system architecture is detailed in Section 3.2, followed by a description of the datasets and experimental setup in Section 3.3. Section 4 discusses the results along with an in-depth cross-paradigm comparative analysis. Finally, Section 5 concludes the paper and outlines directions for future research.

2. LITERATURE REVIEW

The field of TS has evolved significantly, starting from early statistical methods to modern zero/few-shot LLM-based methods. This section surveys key enhancements in extractive summarization, emphasizing the roles of sentence embeddings, attention mechanisms, content coverage strategies, and genetic algorithms. We also discuss existing limitations that motivate the proposed framework.

2.1 Extractive Summarization Techniques

Early extractive summarization techniques were often based on statistical and graph-based methods. For instance, TextRank [20] adapts the PageRank algorithm to determine sentence importance based on co-occurrence patterns, while LexRank [19] employs eigenvector centrality within a similarity graph, where nodes represent sentences and edges represent their cosine similarity score, to identify central sentences. Despite the effectiveness of these methods in identifying salient content, they often suffer from redundancy and fail to ensure coverage of key topics and entities due to their reliance on surface-level features.

More recent methods introduced supervised learning paradigms, where models like those proposed by Kupiec et al. [4] and Conroy and O'Leary [35] leverage features such as sentence position and term frequency to train classifiers for sentence selection. However, these methods depend heavily on annotated training data (document-summary pairs), which may be scarce in specialized domains, and thus lack adaptability without retraining. Unsupervised techniques have remained a focal area of research, with advancements in clustering and graph-based summarization [36].

Recently, pre-trained language models (PLMs) and large language models have transformed summarization tasks by capturing deeper semantic relationships and contextual relevance [16, 37, 38, 39]. Despite these improvements, challenges persist in balancing informativeness, non-redundancy, and comprehensive content coverage.

2.2 Sentence Embeddings and Attention Mechanisms in Summarization

The advent of sentence embeddings has significantly influenced extractive summarization. Earlier models, such as Word2Vec [26] and GloVe [27], offered dense vectors representing words, which can be aggregated into sentence representations. In contrast, recently contextual embeddings from models like BERT [28], RoBERTa [40], and SimCSE [41] provide semantically rich representations that better capture sentence-level meaning. SimCSE, utilizing contrastive learning, excels in modeling semantic similarity, making it effective for both redundancy reduction and identifying important sentences. An et al. [42] introduced CoLo, a contrastive learning-based re-ranking framework that improves one-stage extractive summarization using generated summary-level objectives.

Attention mechanisms, first introduced in neural machine translation [43], have since become fundamental in NLP. In particular, multi-head attention [44], a core component of the Transformer architecture, enables the model to attend to multiple aspects of input sequences simultaneously. This mechanism is particularly useful in extractive summarization for modeling inter-sentence dependencies and assessing contextual relevance. While baseline models like Lead-1 and Random [25] emphasize positional bias, more advanced attention-based techniques weigh sentence relevance in the context of the entire document [45, 46].

Despite these advances, approaches relying solely on sentence embeddings for importance estimation often include redundant content and may overlook critical information. Explicit mechanisms for redundancy avoidance and content coverage are still lacking in many embedding-based models.

2.3 Content Coverage in Summarization

Capturing the most essential information from the source text is a core goal of summarization. Named Entity Recognition (NER) has been widely used to ensure the inclusion of important entities in summaries [47]. Similarly, keyword extraction techniques such as TF-IDF [48] are employed to identify key terms to be included in the generated summaries. More recent models explicitly encourage coverage of key concepts and entities. For instance, [49, 50] proposed mechanisms to encourage coverage of key entities in abstractive summarization.

However, strategies that rely solely on NER or TF-IDF may fail to capture implicit or infrequent yet crucial content. Furthermore, incorporating content coverage as an optimization goal often introduces complexity into reward functions or summary constraints.

The approach proposed in this paper addresses these challenges by combining semantic sentence embeddings (via SimCSE) for assessing relevance and redundancy, multi-head attention for dynamic importance scoring, and a dual coverage strategy utilizing both NER and TF-IDF. This holistic framework is integrated into a genetic algorithm to guide the optimization process, ensuring that generated summaries are not only informative and concise but also semantically diverse and representative of the document's core content, while adhering to length constraints.

3. METHODOLOGY

This study adopts a computational experimental research design, where the proposed summarization framework is evaluated using benchmark datasets and standard evaluation metrics.

This study hypothesizes that integrating contrastive sentence embeddings with a genetic optimization framework, augmented by explicit sentence importance, coverage, and anti-redundancy scoring components, can significantly enhance sentence selection in zero-shot extractive summarization.

3.1 Problem Definition

The problem of single-document extractive summarization is defined as a constrained single-objective optimization problem. Given a document D consisting of N sentences: $\{s_1, s_2, s_3, \dots, s_{N}\}$. the goal is to generate an optimal summary S, which is a subset of sentences from document D, such that: $S = \{s_i \mid s_i \in D,\ i \in I\}$, where I represents the index set of the selected sentences that lead to maximize the overall fitness score. The generated summary must adhere to an upper and lower word count limit defined by $words_{max}$ and $words_{min}$ respectively. In the absence of $words_{min}$, the redundancy minimization objective may leads the summary to consist of very few sentences -possibly only one. The objective function incorporates three key summarization aspects: content relevance and coverage, both of which are to be maximized, and redundancy, which is to be minimized. For uniformity, the redundancy function is reformulated as a maximization criterion. These three aspects collectively define the objective function (OF_{score}) as a maximization problem:

$$\begin{aligned} &\text{Maximize} && OF_{\text{score}} \\ &\text{subject to} && \text{words}_{\min} \le \sum_{s_i \in S} \text{len}(s_i) \le \text{words}_{\max} \end{aligned} \tag{1}$$

language understanding; however, they are not trained explicitly to generate meaningful sentence-level embeddings required for tasks like clustering or semantic similarity measurement. To overcome this limitation, we utilize SimCSE sentence embeddings without any additional fine-tuning. SimCSE represents a state-of-the-art approach that employs contrastive learning to produce semantically informative sentence representations. It is trained to get semantically similar sentences close together in the embedding space, while keeping dissimilar ones far apart. Given an input document D consisting of a sequence of sentences: D = {s_1, s_2, s_3,.., s_N}, each sentence s_i is encoded into a fixed-length dense vector representation $e_i \in R^d$, where d refers to the embedding dimension (e.g., d = 384 for the MiniLM model). The sentence embeddings are computed as follows:

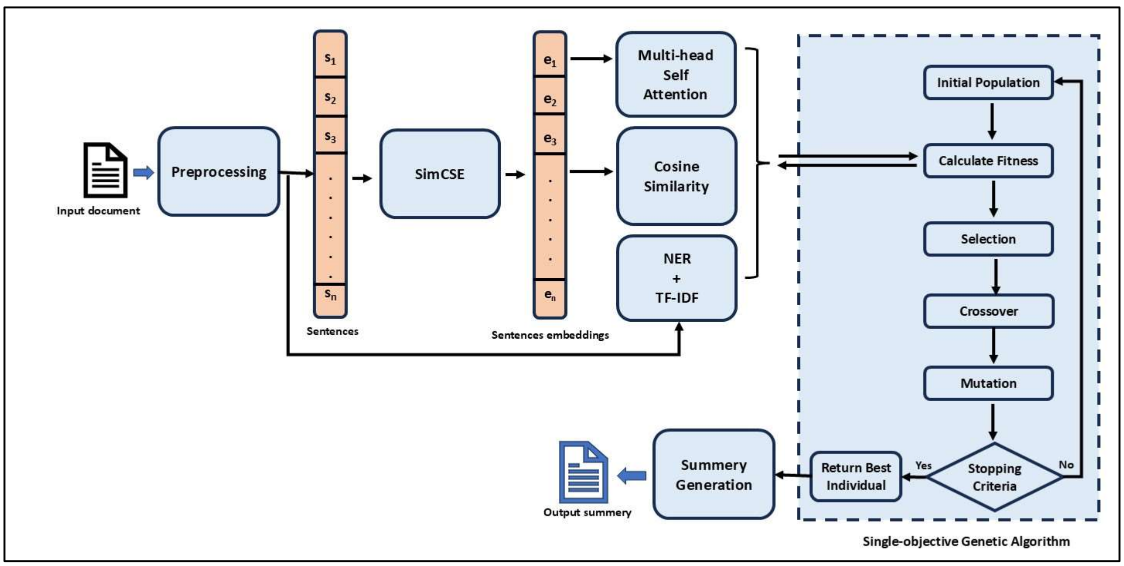

Figure 3: Proposed ETS-BGA System Architecture

$$e_i = \text{SimCSE}(s_i) \tag{2}$$

SimCSE is built upon transformer-based encoders, such as BERT or RoBERTa, and employs a contrastive learning objective:

$$L_{\text{contrastive}} = -\log \frac{\exp(\cos(e_i, e_j^+)/\tau)}{\sum_{k=1}^{2N} \exp(\cos(e_i, e_k)/\tau)} \tag{3}$$

where: $cos(.,.)$ denotes cosine similarity, τ is a temperature hyperparameter, e_j^+ is a positive sample

3.2 Proposed Method

This section discusses the proposed binary genetic optimization-based extractive text summarization approach, which we named ETS-BGA. The proposed architecture of the ETS-BGA is shown in Figure 3.

3.2.1 Pre-processing

Through the pre-processing step, the input document is just tokenized into sentences. Stopwords and punctuation are retained, as their removal can negatively influence sentence structure and the quality of contextual embeddings.

3.2.2 Encoding sentences with SimCSE

Recently, pretrained transformer-based models such as BERT have notably enhanced contextual (usually a dropout-augmented duplicate of s_i), and e_k are negative samples drawn from other sentences in the mini-batch. Once each sentence is encoded into the corresponding embedding, the resulting vector representations are stored as a matrix $E \in R^{N*d}$, where d is the embedding dimension. These embeddings are then used to calculate sentence relevance and inter-sentence redundancy (via cosine similarity).

3.2.3 Population initialization

In the proposed framework, each chromosome (summary) is encoded as a binary vector whose length equals the number of sentences in the document. A bit value of 1 denotes the inclusion of the corresponding sentence in the summary, while 0 indicates exclusion. For example, one such chromosome can be represented as [1, 0, 1, 0, 0, 1, 0, 1, 1, 0], indicating 1^{st},3^{rd}, 6^{th} 8^{th}, and 9^{th} sentences from the input document, with ten sentences present in the generated summary. Only chromosomes satisfying the predefined summary length constraint ($words_{min} \leq textsummarylength \leq words_{max}$) are considered as feasible solutions and retained in the initial population. Constraint checks are applied during initialization and in every generation. Infeasible chromosomes are assigned a zero fitness score, minimizing their likelihood of selection in the next genetic operations.

3.2.4 Objective function

A comprehensive literature review highlights three primary criteria necessary for extractive summarization: content coverage, sentence importance (relevance), and anti-redundancy (redundancy reduction). To enhance content coverage, the proposed method integrates both Named Entity Recognition (NER) and keyword extraction, ensuring that key entities and key topical terms from the original text are included in the final summary. Sentence relevance is evaluated using a Multi-Head Attention mechanism, which captures contextual dependencies and relative importance across document sentences. Redundancy is mitigated using a penalty function based on pairwise cosine similarity between sentence embeddings. The detailed definition of the objective function is presented below:

Summary Coverage Score (SC_{score}): To ensure that generated summaries are both concise and representative of the source document's main ideas, a content coverage objective is incorporated. This objective evaluates the extent to which selected sentences capture the key semantic content of the full document through a hybrid approach combining Named Entity Recognition (NER) and keyword extraction using TF-IDF. Named entities -such as locations, people, organizations, and other specific terms- often carry the main information in the document. A transformer-based NER model is applied to extract entities from both the source document and the candidate summary. Entity coverage is computed as the proportion of entities in the document that also appear in the summary:

$$\text{EntityCoverage} = \frac{|E_S \cap E_D|}{|E_D| + \epsilon} \tag{4}$$

where E_D is the set of named entities in the original document, E_S is the set in the summary, and *epsilon* is a small constant to avoid division by zero when E_D is empty. Higher values indicate better content coverage.

Since named entities alone may not reflect the broader thematic content, TF-IDF is employed to extract the most significant topical terms from the document. Keyword coverage measures the overlap between these top-ranked keywords present in the original document and those present in the summary:

$$\text{KeywordCoverage} = \frac{|K_S \cap K_D|}{|K_D| + \epsilon} \tag{5}$$

where K_D refers to the top-k Keywords from the document and K_S those appearing in the summary.

The total summary coverage score, $SC_{score} \in (0,1)$, is computed as the average of the entity coverage and keyword coverage:

$$SC_{score} = \frac{\text{EntityCoverage} + \text{KeywordCoverage}}{2} \tag{6}$$

This formulation ensures that the final summary retains both critical factual information and the main thematic content of the original document.

Sentence Importance Score (SI_{score}): Once fixed-length sentence embeddings are obtained via SimCSE, the next step is to assess each sentence's contribution to the overall document meaning. To fulfill this, we adopt a *multi-head self-attention* mechanism, which facilitates contextualized

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

interactions among sentences and improves the estimation of their relative importance.

We apply the multi-head attention mechanism proposed by Vaswani et al. [44], in which the self-attention is computed as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{7}$$

Here, Q (query), K (key), and V (value) are linear transformations of the input embeddings. Since all three are derived from the same input E, this constitutes a self-attention setting. This operation enables each sentence to attend to all others, capturing document-level contextual dependencies.

Using h attention heads, each head computes its own attention output, and the results are concatenated and projected through a linear layer:

$$\text{MultiHead}(E) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \tag{8}$$

Where $head_i = Attention(EW_i^Q, EW_i^K, EW_i^V)$, and W_i^Q, W_i^K, W_i^V, and W^O are trainable projection matrices.

The output of the multi-head attention layer is a transformed sequence of vectors $h_1, h_2, \dots, h_N$, with each $h_i \in R^d$ encoding both the intrinsic content of s_i and its contextual importance.

A scalar importance score r_i for each sentence is obtained by applying a feedforward layer followed by a sigmoid activation:

$$r_i = \sigma(\mathbf{w}^\top \mathbf{h}_i + b) \tag{9}$$

where $w \in R^d$ and $b \in R$ are learnable parameters, and $\sigma(.)$ denotes the sigmoid function.

The average importance score of the selected summary sentences is then computed. To normalize this value within the range [0,1], we apply the sigmoid transformation:

$$F_{\text{new}} = \frac{1}{1 + e^{-F_{\text{current}}}} \tag{10}$$

where $F_{current}$ denotes the raw SI_{score} before normalization.

Summary Anti-Redundancy Score (SR_{score}): To ensure that the generated summaries are concise and avoid unnecessary repetition, we explicitly incorporate an anti-redundancy objective. Instead of relying solely on syntactic similarity, we measure redundancy at the semantic level by computing the cosine similarity between sentence embeddings obtained via SimCSE. This approach allows the model to detect similarity even in cases of paraphrasing, synonym usage, or grammatical variation.

The similarity between two sentences s_i and s_j is computed as follows:

$$SM(s_i, s_j) = \begin{cases} 1, & \text{if } \cos\_sim(e_i, e_j) \geq 0.7 \text{ and } i \neq j, \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

Here, e_i and e_j represent the SimCSE embeddings of sentences s_i and s_j, respectively. A value of 1 indicates high similarity (redundancy), while 0 indicates dissimilarity.

The overall redundancy score is then calculated by summing all pairwise similarity values within the summary:

$$SR_{score} = 1 - \left(\frac{\sum_{i=1}^{M} \sum_{j=1}^{M} SM(s_i, s_j)}{2 \cdot M \cdot M}\right) \tag{12}$$

where M is the total number of sentences in the summary. Since our objective is to *maximize* anti-redundancy rather than redundancy, we subtract the normalized redundancy score from 1 as shown in Equation (12).

Finally, the overall objective (fitness) score, OF_{score}, combines the three components—content coverage, informativeness, and anti-redundancy—into a single maximization criterion:

$$OF_{score} = SC_{score} + SI_{score} + SR_{score} \tag{13}$$

3.2.5 Selection operator

After the initialization of the population, the fitness of each individual is evaluated at every iteration using the objective function score defined in Equation (13). Once the fitness values are computed, a tournament selection strategy is employed to select parent individuals to generate offspring.

3.2.6 Crossover operator

A two-point crossover strategy is employed to generate offspring from two selected parents. Two crossover positions are randomly determined within the chromosome structure of the parents, and the segment between these points is exchanged. For example, given: Parent1: [1,0,1,0,0,1,0,0,1,0] and Parent2: [0,1,1,0,1,0,1,0,1,1]. If crossover points are at positions 3 and 7, the generated offspring are: Child1: [1,0,1,0,1,0,0,0,1,0] and Child2: [0,1,1,0,0,1,0,0,1,1].

3.2.7 Mutation operator

This process involves randomly flipping a randomly selected subset of bits (genes) in the chromosome from their original value ($0 \rightarrow 1$ or $1 \rightarrow 0$), introducing new variations into the population.

3.2.8 Termination criteria and summary generation

The genetic algorithm iterates until the maximum number of generations (G) is reached. Hence, after the last iteration, the best individual is selected as a summary candidate.

3.3 Experimental Setup

In this section, we evaluate the performance of the proposed ETS-BGA system, presenting the following components: (i) a description of the dataset, (ii) an overview of the ROUGE valuation metrics, (iii) implementation details along with parameter settings, and (iv) experimental results and comparative analysis.

3.3.1 Dataset description

We evaluated the proposed system on two benchmark summarization datasets: the CNN/DailyMail (CNN/DM; [51]) dataset and the New York Times (NYT; [52]) dataset. The CNN/DM dataset comprises approximately 90,000 articles from the Cable News Network (CNN) and 197,000 articles from the Daily Mail, each paired with a corresponding gold-standard reference summary. The NYT dataset contains about 1.8 million articles published by the New York Times between January 1, 1987, and June 19, 2007. Among these, roughly 650,000 articles have been manually summarized by library scientists, making the dataset suitable for extractive text summarization research.

3.3.2 ROUGE evaluation metrics

The quality of the generated summaries was estimated using the ROUGE metric suite [53]. We report full-length F1-scores for ROUGE-1, ROUGE-2, and ROUGE-L on both the CNN/DM and NYT datasets. The F1-score is defined as the harmonic mean of precision and recall, providing a balanced measure between the two. All ROUGE scores were computed using the rouge-score Python library (version 0.1.2).

3.3.3 Implementation details

All experiments were implemented in Python, utilizing the Hugging Face Transformers library and the DEAP evolutionary algorithm framework. To ensure reproducibility, random seeds for NumPy, PyTorch, and Python's random module were set to the fixed value 42, enabling deterministic score computation.

We employ the supervised SimCSE model based on Bert-base-uncased, trained on the NLI corpora (SNLI and MNLI), to generate high-quality semantic sentence embeddings extracting the [CLS] token representation from the final transformer layer. Importantly, no summarization-specific fine-tuning is applied, ensuring that the entire framework operates in a strictly zero-shot manner with respect to the summarization task.

Sentence importance was estimated through a custom multi-head attention module with 8 heads and an embedding dimension of 768. The attention outputs were passed through a linear projection layer followed by a sigmoid activation, producing normalized relevance scores within the range [0,1].

Content coverage was measured using two complementary strategies. First, entity extraction was performed via the Hugging Face NER pipeline with a generic pretrained bert-base-cased model, invoked through the pipeline("ner", grouped_entities=True) interface. Second, keyword coverage was computed using Scikit-learn's TfidfVectorizer with English stop words removed, selecting the top 10 keywords with the highest TF-IDF scores as representative document keywords for coverage evaluation.

Redundancy between candidate sentences was assessed via cosine similarity on SimCSE embeddings, discarding sentence pairs whose similarity exceeded a threshold of 0.7.

The genetic algorithm was configured with a population size of 30, 20 generations, crossover probability of 0.7, mutation probability of 0.3, tournament selection size of 3, and redundancy threshold of 0.7. The summary length constraint was set to 40–100 words for CNN/DM and 40–200 words for NYT.

All hyperparameters were empirically tuned through preliminary experiments to ensure stable convergence and balanced optimization across relevance, coverage, and redundancy minimization. As shown in Figure 4, the fitness score stabilizes after a round of 15 generations, emphasizing that increasing the population size or the number of generations provides slight improvements while significantly increasing computational overhead.

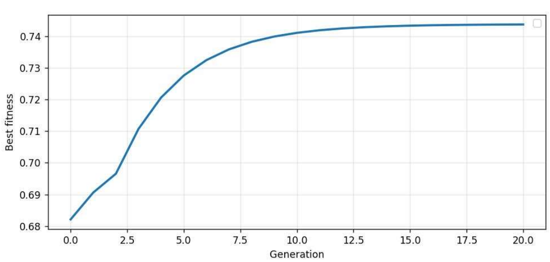

Figure 4: Convergence of Best Fitness per Generation.

4. RESULTS AND DISCUSSION

Since the proposed **ETS-BGA** framework operates entirely in a *zero-shot* setting—without any fine-tuning or task-specific supervision—all experiments are conducted exclusively on the *test splits* of the CNN/DM and NYT datasets. This evaluation protocol ensures a fair assessment of the model's true generalization capability and its robustness when processing previously unseen documents. For each document in the test split, the system is executed once to generate an extractive summary, after which the F1-based ROUGE scores are computed individually. The final results reported in Table 1 correspond to the average ROUGE scores across all test documents. Examples of the highest- and lowest-quality summaries for CNN/DM documents, along with their corresponding ROUGE scores, coverage score ($SC_{textscore}$), relevance score ($SI_{textscore}$), and anti-redundancy score ($SR_{textscore}$), are shown in Figure 5.

Table 1: Experimental results on CNN/DM and NYT test sets using ROUGE F1. Results are taken from the original paper if available

Paradigm	Method	Category	CNN/DM			NYT		
			R-1	R-2	R-L	R-1	R-2	R-L
Statistical (unsupervised)	LEAD-3	Extractive	40.50	17.70	36.70	35.50	17.20	32.00
	TEXTRANK (tf-idf)	Extractive	33.20	11.80	29.60	33.20	13.10	29.00
	TEXTRANK (skip-thought)	Extractive	31.40	10.20	28.20	30.10	9.60	26.10
	PACSUM [54]	Extractive	40.70	17.80	36.90	41.40	21.70	37.50
Deep learning (Supervised)	REFRESH [24]	Extractive	41.30	18.40	37.50	41.30	22.00	**37.80**
	PTR-GEN [25]	Abstractive	39.50	17.30	36.40	42.70	22.10	38.00
	SUMMARUNNER [23]	Extractive	39.60	16.20	35.30	–	–	–
PLM + finetune	BertSumExt [16]	Extractive	43.25	20.24	**39.63**	–	–	–
	BertSumAbs [16]	Abstractive	41.72	19.39	38.76	–	–	–
LLM (Zero/few-shot)	SummIt (ChatGPT) [37]	Abstractive	37.29	13.60	26.87	–	–	–
	Element (GPT3) [38]	Abstractive	37.75	15.20	34.25	–	–	–
	HADAS [39]	Abstractive	–	–	20.12	–	–	–
	ETS-BGA (ours)	Extractive	**43.38**	**22.53**	31.93	**45.26**	**23.04**	28.77

Journal of Theoretical and Applied Information Technology
30th April 2026. Vol.104. No.8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	Golden summary	System generated summary	Evaluation scores	
Examples of the highest quality obtained summary	Australian Fashion Report revealed the Australian-sold brands and companies that ignore the exploitation of their overseas workers. Lowes, Industrie, Best & Less and the Just Group - which includes Just Jeans, Portmans and Dotti - were some of the worst performers. Etiko, Audrey Blue, Cotton On, H&M and Zara had some of the best scores. 75 per cent of companies don't know the source of all their fabrics and inputs.	Lowes, Industrie, Best & Less and the Just Group - which includes Just Jeans, Portmans and Dotti - were identified as some of the worst performing companies by The 2015 Australian Fashion Report. Amongst the best performers were Etiko, Audrey Blue, Cotton On, H&M and Zara. 'Furthermore, 91 per cent of companies still don't know where all their cotton comes from and 75 per cent don't know the source of all their fabrics and inputs.	R-1	75.86
			R-2	61.54
			R-L	62.07
			SC	0.62
			SI	0.61
			Sr	1.0
	This is the moment that a crew of firefighters struggled to haul a giant pig out of a garden swimming pool. The prize porker, known as Pigwig, had fallen into the pool in an upmarket neighbourhood in Ringwood, Hampshire. A team from Dorset Fire and Rescue struggled to haul the huge black pig out of swimming pool water. 'One crew of firefighters from Ferndown and a specialist animal rescue unit from Poole were mobilised to this incident.	The prize porker, known as Pigwig, had fallen into the pool in an upmarket neighbourhood in Ringwood, Hampshire. His owners had been taking him for a walk around the garden when the animal plunged into the water and was unable to get out. Firefighters were also called out to a horse which had fallen into a swimming pool in Heyshott, West Sussex. 'The crew rescued the pig from the swimming pool using specialist animal rescue slide boards, strops and lines to haul the pig from the swimming pool.'	R-1	59.39
			R-2	28.22
			R-L	37.58
			SC	0.58
			SI	0.62
			Sr	1.0
Examples of the lowest quality obtained summary	Arsenal take on Stoke City in Barclays Under 21 Premier League clash. Jack Wilshere and club captain Mikel Arteta have been out since November. Abou Diaby has been ravaged by injuries during nine-year spell at club. Arteta, Wilshere and Diaby are all close to first-team returns. Young winger Serge Gnabry also in the side on return from injury. READ: Arsenal's Alex Oxlade-Chamberlain, Calum Chambers, Jack Wilshere and Danny Welbeck keep their agents close. CLICK HERE for all the latest Arsenal news.	Steve Gatting's young Gunners take on Stoke City at the Emirates Stadium in a U21 Premier League clash. Teenage winger Serge Gnabry is also in the starting line-up, alongside a host of promising youngsters including Dan Crowley. Wilshere watched Arsenal beat Liverpool 4-1 in the Premier League from the stands on Saturday afternoon . Iliev, Maitland-Niles, Connor, Bielik, Ormonde-Ottewill, Diaby, Arteta, Gnabry, Wilshere, Crowley, Iwobi .	R-1	32.47
			R-2	11.84
			R-L	20.78
			SC	0.59
			SI	0.61
			Sr	0.94
	AQAP says a "crusader airstrike" killed Ibrahim al-Rubaish. Al-Rubaish was once detained by the United States in Guantanamo..	Ibrahim al-Rubaish died Monday night in what AQAP's media wing, Al-Malahem Media, called a "crusader airstrike." Al-Rubaish was once held by the U.S. government at its detention facility in Guantanamo Bay, Cuba. He was eventually released as part of Saudi Arabia's program for rehabilitating jihadist terrorists, a program that U.S. Sen. Jeff Sessions, R-Alabama, characterized as "a failure." Yemen, however, has been in disarray since Houthi rebels began asserting themselves last year.	R-1	30
			R-2	18
			R-L	24
			SC	0.62
			SI	0.37
			Sr	1.0

Figure 5: Qualitative Examples of Generated Summaries and Evaluation Scores.

Table 1 also provides a comparative analysis against various recent summarization methods across different paradigms. The first block lists unsupervised approaches. LEAD-3 simply selects the first three sentences as the summary. TEXTRANK (tf-idf and skip-thought variants) [20] models a document as a sentence graph, ranking sentences by node scores and selecting the top k. PACSUM [54] uses BERT-based sentence embeddings in a graph framework, ranking sentences by centrality with position-sensitive edge weighting. Among these, PACSUM performs strongly (R-1 = 40.70 on CNN/DM, 41.40 on NYT), and LEAD-3 remains a surprisingly competitive baseline (R-1 = 40.50 on CNN/DM).

The second block covers recent supervised deep learning approaches. REFRESH [24] is an extractive system trained end-to-end with reinforcement learning to optimize the ROUGE metric directly. PTR-GEN [25] is an abstractive sequence-to-sequence model with copy and coverage mechanism. SUMMARUNNER [23] is an RNN-based extractive model. These methods outperform unsupervised approaches, especially in ROUGE-2 and ROUGE-L. PTR-GEN achieves the highest R-L score on NYT (38.0) and maintains competitive performance in terms of R-1 score (R-1 = 42.70), while REFRESH and SUMMARUNNER deliver robust extractive performance (REFRESH R-1 = 41.30 on both datasets).

The third block presents models based on pre-trained language models (PLMs) with finetune, such as BertSumExt and BertSumAbs [16]. They significantly outperform earlier supervised methods. BertSumExt achieves an R-1 of 43.25 and R-2 of 20.24 on CNN/DM, demonstrating the benefits of transfer learning. BertSumAbs, its abstractive counterpart, follows closely with robust results across all ROUGE metrics.

The fourth block lists recent advancements in large language models (LLMs), including SummIt (ChatGPT) [37], Element (GPT-3) [38], and HADAS [39]. While these methods perform well without task-specific fine-tuning, their ROUGE scores remain below those of fine-tuned PLM-based systems. For example, SummIt achieves an R-1 of 37.29 on CNN/DM, highlighting the potential of zero/few-shot summarization but also revealing room for improvement in matching specialized extractive systems.

The proposed ETS-BGA method outperforms all existing extractive systems on CNN/DM, achieving an R-1 of 43.38 and R-2 of 22.53, and achieving R-1 of 45.26 and R-2 of 23.04 on NYT. On CNN/DM, ETS-BGA outperforms BertSumExt by a clear margin in R-2 (22.53 vs. 20.24) and matches the best abstractive models. These results confirm the effectiveness of integrating SimCSE-based embeddings with a genetic algorithm to balance informativeness, coverage, and redundancy.

Overall, while deep learning and PLM-based fine-tuned models dominate in absolute performance, graph-based unsupervised methods remain strong baselines, and LLM-based zero/few-shot methods offer promising flexibility for rapid deployment. The proposed ETS-BGA system achieves state-of-the-art extractive performance, demonstrating the strength of embedding-enhanced evolutionary optimization, however the high computational cost it incur due to the evolutionary optimization process, which may affect scalability for large documents. Figure 6 presents qualitative comparisons of gold summaries with outputs from ETS-BGA, REFRESH [24], and PACSUM [55] on CNN/DM.

	Document 1	Document 2
Golden	Filmmaker Michael Konig from Cologne, Germany has created an amazing video showing solar activity. It was made by stitching together footage from Nasa's Solar Dynamics Observatory over five years. The footage includes loops of 'coronal rain' showering the surface of the sun. Transits of the moon, Venus and Earth are also seen - and a comet breeze through the outer solar atmosphere.	Queen Victoria's holiday residence was Osborne House on the Isle of Wight. But her journeys there involved train and ferry ride and then another train ride to a station more than two miles from the property. In 1875, a station was built at Whippingham just to serve Royal residence. Building is now a five-bedroom home, currently on the market for 625,000.
REFRESH	An incredible video has stitched together STAS footage from a Nasa spacecraft, revealed the beauty of the solar surface as it bursts with energy. Taken over five years, the footage includes plasma raining down on the sun, an extreme solar eruption and even a comet breezing through the sun's atmosphere. The movie, called Sun, was created by filmmaker Michael K onig from Cologne, Germany. It uses footage recorded by Nasa's Solar Dynamics Observatory (SDO) spacecraft between 2011 and 2015.	'It is impossible to imagine a prettier spot,' Queen Victoria said of her holiday residence, Osborne House on the Isle of Wight. It was at the beach there that she would take to the sea in her wooden bathing machine, and that all of her children learned to swim. She would travel to Portsmouth by train and then by ferry to Ryde. So, in 1875, a station was built at Whippingham, the closest point on the line to Osborne House – just to serve the Royal residence.
PACSUM	You might have seen fantastic images of the sun before, or even clips showing its activity - but you've never seen anything like this. An incredible video has stitched together footage from a Nasa spacecraft, revealed the beauty of the solar surface as it bursts with energy. Taken over five years, the footage includes plasma raining down on the sun, an extreme solar eruption and even a comet breezing through the sun's atmosphere.	'It is impossible to imagine a prettier spot,' Queen Victoria said of her holiday residence, Osborne House on the Isle of Wight. It was at the beach there that she would take to the sea in her wooden bathing machine, and that all of her children learned to swim. But, however pleasing her days at Osborne House, Victoria's journeys there were never easy.
ETS-BGA (ours)	The movie, called Sun, was created by filmmaker Michael Konig from Cologne, Germany. Transits of the moon, Venus and Earth across the sun, as observed from the SDO, are also seen, while towards the end of the video Comet Lovejoy can be seen passing the sun in December 2011. Coronal rain is formed when hot plasma in the corona - the sun's atmosphere - condenses and falls back to the surface .	Pretty spot: Queen Victoria spent her holidays in Osborne House on the Isle of Wight. She would travel to Portsmouth by train and then by ferry to Ryde. Victoria died in 1901 and the station went into public use in 1903. The part of the house now used as a sitting room used to be the station's waiting room, while the main bedroom used to be the points room, where the signalman would pull levers to control trains on the track. Location Whippingham, Isle of Wight.

Figure 6: Examples Of Golden Summaries And System Output Of REFRESH, PACSUM, And ETS_BGA On CNN/DM Dataset.

4.1 Sentence Position Distribution

We further investigate the distribution of sentences selected by various summarization techniques within documents. Specifically, we compare **ETS-BGA**, **LEAD-3**, **PACSUM**, and **ORACLE** based on the positions of their selected sentences, considering only the first 12 sentences of each document in the CNN/DM dataset. The **ORACLE** model serves as an upper performance bound for extractive summarization, as its summaries are generated by selecting subsets of sentences that achieve the highest ROUGE scores [23].

As shown in Figure 7, ORACLE selects sentences distributed relatively evenly across the document, while LEAD-3 consistently selects the opening three sentences. In comparison, PACSUM exhibits a sentence position distribution pattern closely resembling that of LEAD-3. On the other hand, ETS-BGA produces a distribution more aligned with ORACLE, suggesting that our model is less dependent on sentence positions than PACSUM.

To quantitatively assess this observation, we compute the Kullback–Leibler (KL) divergence between the sentence position distributions of an unsupervised model and ORACLE, denoted as $KL(.|ORC)$. The results indicate that $KL(PACSUM|ORC) = 0.614$ is considerably higher than $KL(ETS-BGA|ORC) = 0.11$, confirming that ETS-BGA achieves a sentence selection distribution more closely correlated with ORACLE's.

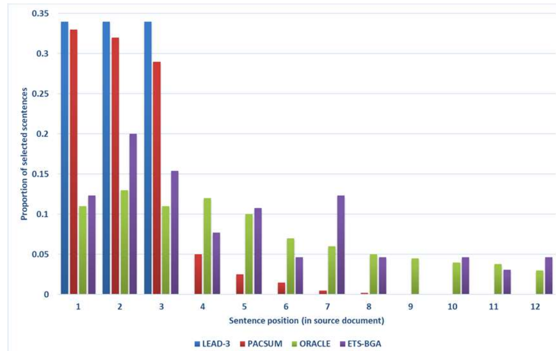

Figure 7: Proportion Of Extracted Sentences By Different Unsupervised Models Against Their Positions.

5. CONCLUSION

In this work, we introduced a zero-shot, single-objective genetic algorithm–based extractive summarization framework that integrates contrastive learning–derived sentence embeddings, multi-head attention–based relevance estimation, redundancy control via cosine similarity between sentence embeddings, and content coverage through named entities and key concepts. The framework jointly addresses relevance, redundancy minimization, and semantic coverage, enabling the generation of concise and informative summaries without further training or fine-tuning.

A key strength of our method is its zero-shot design -once sentence embeddings and relevance scores are computed, the genetic algorithm operates over a fixed scoring landscape, avoiding repeated model updates or incremental learning. This makes the approach computationally efficient and well-suited for real-time summarization of unseen documents. Incorporating named entity recognition and TF-IDF-based keyword extraction into the coverage mechanism further ensures that the summaries retain essential factual and topical content from the source text.

After an extensive cross-paradigm comparative analysis encompassing statistical unsupervised, deep learning supervised, PLM finetuned, and zero/few-shot LLM summarization methods, experiments on the CNN/DailyMail dataset clarify that the proposed method surpasses several strong baselines in ROUGE evaluation metrics. Future directions include extending the framework to multi-document and multilingual summarization, as well as integrating abstractive post-processing techniques to enhance coherence and readability.

REFERENCES

[1] R. M. Alguliyev, R. M. Aliguliyev, N. R. Isazade, A. Abdi, and N. Idris, Cosum: Text summarization based on clustering and optimization," *Expert Systems*, vol. 36, no. 1, p. e12340, 2019.

[2] D. R. Radev, M. T. Joseph, B. Gibson, and P. Muthukrishnan,A bibliometric and network analysis of the field of computational linguistics,'' *Journal of the Association for Information Science and Technology*, vol. 67, no. 3, pp. 683–706, 2016.

[3] M. S. Binwahlan, N. Salim, and L. Suanmali, Swarm based text summarization," in *2009 International Association of Computer Science and Information Technology-Spring Conference*. IEEE, 2009, pp. 145--150.

[4] J. Kupiec, J. Pedersen, and F. Chen,A trainable document summarizer,'' in *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, 1995, pp. 68–73.

[5] A. K. Yadav, Ranvijay, R. S. Yadav, and A. K. Maurya, Graph-based extractive text summarization based on single document," *Multimedia Tools and Applications*, vol. 83, no. 7, pp. 18,987--19,013, 2024.

[6] I. Mani and E. Bloedorn,Multi-document summarization by graph search and matching,'' *arXiv preprint cmp-lg/9712004*, 1997.

[7] 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.

[8] C. Jiang, R. Xie, W. Ye, J. Sun, and S. Zhang,Exploiting pseudo image captions for multimodal summarization,'' *arXiv preprint arXiv:2305.05496*, 2023.

[9] J. Lin, H. Hua, M. Chen, Y. Li, J. Hsiao, C. Ho, and J. Luo, Videoxum: Cross-modal visual and textural summarization of videos," *IEEE Transactions on Multimedia*, vol. 26, pp. 5548--5560, 2023.

[10] Y. Gong and X. Liu,Generic text summarization using relevance measure and latent semantic analysis,'' in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, 2001, pp. 19–25.

[11] J. He, W. Kry{'s}ci{'n}ski, B. McCann, N. Rajani, and C. Xiong, Ctrlsum: Towards generic controllable text summarization," *arXiv preprint arXiv:2012.04281*, 2020.

[12] F. Bayatmakou, A. Mohebi, and A. Ahmadi,An interactive query-based approach for summarizing scientific documents,'' *Information Discovery and Delivery*, vol. 50, no. 2, pp. 176–191, 2022.

[13] L. H. Reeve, H. Han, and A. D. Brooks, The use of domain-specific concepts in biomedical text summarization," *Information Processing & Management*, vol. 43, no. 6, pp. 1765--1776, 2007.

[14] A. Afzal, J. Vladika, D. Braun, and F. Matthes,Challenges in domain-specific abstractive summarization and how to overcome them,'' *arXiv preprint arXiv:2307.00963*, 2023.

[15] S. Akter, A. S. Asa, M. P. Uddin, M. D. Hossain, S. K. Roy, and M. I. Afjal, An extractive text summarization technique for bengali document (s) using k-means clustering algorithm," in *2017 ieee international conference on imaging, vision & pattern recognition (icivpr)*.hskip 1em plus 0.5em minus 0.4emrelax IEEE, 2017, pp. 1--6.

[16] Y. Liu and M. Lapata,Text summarization with pretrained encoders,'' *arXiv preprint arXiv:1908.08345*, 2019.

[17] J. Carbonell and J. Goldstein, The use of mmr, diversity-based reranking for reordering documents and producing summaries," in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 1998, pp. 335--336.

[18] H. Lin and J. Bilmes,A class of submodular functions for document summarization,'' in *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, 2011, pp. 510–520.

[19] G. Erkan and D. R. Radev, Lexrank: Graph-based lexical centrality as salience in text summarization," *Journal of artificial intelligence research*, vol. 22, pp. 457--479, 2004.

[20] R. Mihalcea and P. Tarau,Textrank: Bringing order into text,'' in *Proceedings of the 2004 conference on empirical methods in natural language processing*, 2004, pp. 404–411.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[21] H. P. Edmundson, New methods in automatic extracting," *Journal of the ACM (JACM)*, vol. 16, no. 2, pp. 264--285, 1969.
[22] J. Cheng and M. Lapata,Neural summarization by extracting sentences and words,'' *arXiv preprint arXiv:1603.07252*, 2016.
[23] R. Nallapati, F. Zhai, and B. Zhou, Summarunner: A recurrent neural network based sequence model for extractive summarization of documents," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, 2017.
[24] S. Narayan, S. B. Cohen, and M. Lapata,Ranking sentences for extractive summarization with reinforcement learning,'' *arXiv preprint arXiv:1802.08636*, 2018.
[25] A. See, P. J. Liu, and C. D. Manning, Get to the point: Summarization with pointer-generator networks," *arXiv preprint arXiv:1704.04368*, 2017.
[26] T. Mikolov, K. Chen, G. Corrado, and J. Dean,Efficient estimation of word representations in vector space,'' *arXiv preprint arXiv:1301.3781*, 2013.
[27] J. Pennington, R. Socher, and C. D. Manning, Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532--1543.
[28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova,Bert: Pre-training of deep bidirectional transformers for language understanding,'' in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171–4186.
[29] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
[30] M. Zhong, P. Liu, Y. Chen, D. Wang, X. Qiu, and X. Huang,Extractive summarization as text matching,'' *arXiv preprint arXiv:2004.08795*, 2020.
[31] Y. Liu, P. Liu, D. Radev, and G. Neubig, Brio: Bringing order to abstractive summarization," *arXiv preprint arXiv:2203.16804*, 2022.
[32] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*,Language models are few-shot learners,'' *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
[33] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, Chain-of-thought prompting elicits reasoning in large language models," *Advances in neural information processing systems*, vol. 35, pp. 24,824--24,837, 2022.
[34] H. Zhang, P. S. Yu, and J. Zhang,A systematic survey of text summarization: From statistical methods to large language models,'' *ACM Computing Surveys*, 2024.
[35] J. M. Conroy and D. P. O'leary, Text summarization via hidden markov models," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, 2001, pp. 406--407.
[36] E. Filatova and V. Hatzivassiloglou,Event-based extractive summarization,'' in *Text summarization branches out*, 2004, pp. 104–111.
[37] H. Zhang, X. Liu, and J. Zhang, Summit: Iterative text summarization via chatgpt," *arXiv preprint arXiv:2305.14835*, 2023.
[38] Y. Wang, Z. Zhang, and R. Wang,Element-aware summarization with large language models: Expert-aligned evaluation and chain-of-thought method,'' *arXiv preprint arXiv:2305.13412*, 2023.
[39] Y. Xia, X. Liu, T. Yu, S. Kim, R. A. Rossi, A. Rao, T. Mai, and S. Li, Hallucination diversity-aware active learning for text summarization," *arXiv preprint arXiv:2404.01588*, 2024.
[40] X. Liu, P. He, W. Chen, and J. Gao,Multi-task deep neural networks for natural language understanding,'' *arXiv preprint arXiv:1901.11504*, 2019.
[41] T. Gao, X. Yao, and D. Chen, Simcse: Simple contrastive learning of sentence embeddings," *arXiv preprint arXiv:2104.08821*, 2021.
[42] C. An, M. Zhong, Z. Wu, Q. Zhu, X. Huang, and X. Qiu,Colo: A contrastive learning based re-ranking framework for one-stage summarization,'' *arXiv preprint arXiv:2209.14569*, 2022.
[43] D. Bahdanau, K. Cho, and Y. Bengio, Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
[44] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, {L}. Kaiser, and I. Polosukhin,Attention is all you need,'' *Advances in neural information processing systems*, vol. 30, 2017.
[45] H. Li, J. Zhu, T. Liu, J. Zhang, C. Zong *et al.*, Multi-modal sentence summarization with

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

modality attention and image filtering." in *IJCAI*, 2018, pp. 4152--4158.
[46] S. Xu, X. Zhang, Y. Wu, F. Wei, and M. Zhou,Unsupervised extractive summarization by pre-training hierarchical transformers,'' *arXiv preprint arXiv:2010.08242*, 2020.
[47] A. Nenkova, K. McKeown *et al.*, Automatic summarization," *Foundations and Trends{®} in Information Retrieval*, vol. 5, no. 2--3, pp. 103--233, 2011.
[48] G. G. Chowdhury, *Introduction to modern information retrieval*.hskip 1em plus 0.5em minus 0.4emrelax Facet publishing, 2010.
[49] Z. Cao, F. Wei, W. Li, and S. Li,Faithful to the original: Fact aware neural abstractive summarization,'' in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.
[50] W. Xu, C. Li, M. Lee, and C. Zhang, Multi-task learning for abstractive text summarization with key information guide network," *EURASIP Journal on Advances in Signal Processing*, vol. 2020, pp. 1--11, 2020.
[51] K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom,Teaching machines to read and comprehend,'' *Advances in neural information processing systems*, vol. 28, 2015.
[52] E. Sandhaus, The new york times annotated corpus," *Linguistic Data Consortium, Philadelphia*, vol. 6, no. 12, p. e26752, 2008.
[53] C.-Y. Lin,Rouge: A package for automatic evaluation of summaries,'' in *Text summarization branches out*, 2004, pp. 74–81.
[54] H. Zheng and M. Lapata, Sentence centrality revisited for unsupervised summarization," *arXiv preprint arXiv:1906.03508*, 2019. [55] L. Page, S. Brin, R. Motwani, and T. Winograd,The pagerank citation ranking: Bringing order to the web.'' Stanford infolab, Tech. Rep., 1999.