IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 13 May 2024, accepted 14 June 2024, date of publication 27 June 2024, date of current version 9 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3420163

RESEARCH ARTICLE

Abstractive Summarization Model for Summarizing Scientific Article

MEHTAP ULKER AND A. BEDRI OZER
Computer Engineering Department, Firat University, 23190 Elâziğ, Türkiye
Corresponding author: Mehtap Ulker (m.ulker@firat.edu.tr)

ABSTRACT Researchers consistently publish articles to contribute to science. However, it has become difficult to understand the terms employed in the document along with the way the semantic content is associated with other terms because of the rapid growth in the publication of scientific journals. Therefore, the generation of summaries based on scientific terms is more difficult with longer articles. The preparation of a summary with semantic relations between terms is addressed with graph-based techniques. However, graph-based methods are inadequately focused on generating summaries of scientific articles. To address this problem, a novel graph-based abstractive summarization (GBAS) model based on SciBERT and the graph transformer network (GTN) is proposed in this paper. The scientific content is encoded with SciBERT, terminology-related word extracts from the article with the Scientific Information Extractor (SciIE) system, and long documents are encoded and summarized with GTN. The proposed model is compared with baseline models. Experimental results show that the proposed model outperforms baseline methods in summarizing long scientific articles with ROUGE-L scores of 34.96.

INDEX TERMS Text summarization, abstractive method, SciBERT, SciIE, graph transformer.

I. INTRODUCTION

Many researchers from different research areas publish scientific papers to contribute to the scientific community. The massive flow of scientific articles makes accessing the required information more challenging. Researchers need to get a handle on salient information without reading the entire document. Therefore, the tendency toward automatic text summarization has started to grow. Automatic text summarization (ATS) presents the content of the document as quickly and concisely as possible while preserving the original document's integrity. ATS can be performed with an extractive or abstractive approach, depending on the selection, combination, or paraphrasing of salient sentences in the source document [1]. Extractive summarization generates a summary from sentences that best express the main idea in the source document [2], whereas abstractive summarization generates rewritten summaries with sentences or words that are different from the source document [3]. In this respect, the abstractive approach generates summaries that are similar to human-written summaries.

The rapid development of deep learning techniques has resulted in great advances in abstractive summarization methods [4], [5]. However, it is still a challenge to intelligently summarize scientific articles via traditional neural summarization methods. The main reasons are: (I) These methods have been trained for task-specific applications on general domain datasets, which consist of news articles, blog posts, tweets, etc. (II) The data presented in scientific articles is quite complex because it requires expertise to identify new technologies and their connections [6]. Therefore, information extraction (IE) systems may be preferred in practice to automatically determine these relations. However, the annotations of the relations extracted by these methods are insufficient for a scientific summary of the text [7], [8]. (III) In current studies [4], existing approaches become insufficient in capturing the semantic contexts of phrases inside a document as document lengths increase. The summarization of scientific long documents is beyond the capacity of attention-based models and Long-Short Term

The associate editor coordinating the review of this manuscript and approving it for publication was Binit Lukose.

Memory (LSTM) In this paper, considering all the above problems, a novel scientific text summarization model based on SciBERT and the graph transformer network (GTN) has been proposed that generates abstracts from the text of scientific articles. First, entity, co-reference, and relation annotations were extracted from the source document with the Scientific Information Extractor (SciIE) to handle the hierarchical structure of the document [7]. Second, a knowledge graph has been constructed with this information. The output of the last hidden state of SciBERT was used to encode the text. Finally, GTN [8] was performed to summarize text from the knowledge graph. The main contributions of this paper are as follows:

- A novel model has been proposed for the summarization of scientific articles consisting of SciBERT trained on a large corpus of scientific text and a graph transformer that benefits from the relational structures of the knowledge graph without linearization or hierarchical constraints.
- A novel graph-based model is presented that summarizes long documents.
- The effectiveness of summarization models in summarizing short and long documents was compared with the proposed model.
- The proposed model can summarize scientific articles containing current topics related to computer science such as "fingerprint", "image processing", "natural language processing", "cyber security" and "machine learning".

The remainder of the article is organized as follows: Section II presents previous studies. The proposed model is presented in Section III. Experimental studies and discussion are included in Section IV. The conclusion is given in the last section.

II. RELATED WORK

A. TEXT SUMMARIZATION APPROACHES IN SCIENTIFIC ARTICLES

Text summarization approaches are performed by directly extracting salient sentences from the source documents or rewriting these sentences with words that differ from the source documents [9], [10]. Previous researchers have focused on extractive methods to summarize scientific articles because abstractive methods are more difficult and complex because they require advanced Natural Language Processing (NLP) techniques. These studies are summarized in Table 1. In this study [3], they proposed a generic summarizer model that is language-independent and based on a quantum-inspired approach for the extraction of important sentences. In this study [11], they proposed a graph-based framework that can also be applied to scientific articles without any domain or language constraints. This model utilizes the advantages of graph-based, statistical-based, semantic-based, and centrality-based methods. In this study [12], they proposed a regression-based model to highlight salient sentences in scientific articles. They experimented on three different scientific datasets (CSPubSum, AlPubSum, and BioPubSum) to demonstrate the effectiveness of their method. In this study [13], they constructed a large-scale manually annotated dataset (SciSummNet) for summarizing scientific articles. In addition, a hybrid summarization model was proposed. The effectiveness of their corpus on this model and the data-driven neural models was evaluated. In this study [14], they presented a new model for summarizing scientific articles by inspiring SummPipon [15]. In this study [16], they constructed a novel corpus (SciTLDR) consisting of scientific papers related to the computer science domain. In addition, a novel model (CATTS) was proposed to evaluate their corpus. The proposed model is suitable for both extractive and abstractive methods.

Abstractive methods are closer to reality in terms of generating a summary by thinking human-like. Recently, researchers have focused on it for generating scientific summaries [9], [10]. In this study [5], they presented a graph network-based model based on a sentence-level denoiser and an auto-regressive generator. To demonstrate the effectiveness of their model, PubMed and CORD-19 datasets containing scientific articles in the biomedical domain were used. In this study [17], they proposed a sequence-to-sequence-based model with three encoders and one decoder. In addition, they proposed novel evaluation metrics, namely ROUGE1-NOORDER, ROUGE1-STEM, and ROUGE1-CONTEXT. In this study [4], they presented a SciBERT-based summarization model to summarize scientific articles related to COVID-19. This model consists of a graph attention network and a pretraining language model (SciBERT). To evaluate the proposed model, the CORD-19 (COVID-19 Open Research Dataset) consisting of scientific articles was used. In this study [18], they presented a novel model consisting of timescale adaptation over the pointer-generator-coverage network. It has been mentioned that this model is successful in summarizing long articles. Bao et al. [19] developed GEMINI model based on an abstractive model that a rewriter and a fuser to emulate the sentence rewriting and fusion approaches. The model switches between rewriting and fusing modes according to the characteristics of each training sentence. Cao et al. mentioned that the noise in the training dataset will cause problems in summary generations. They [20] proposed a rejection learning-based model for abstractive summarization. Li et al. [21]. present a HGSum model that includes a heterogeneous graph to represent different semantic units of documents (e.g., words and sentences). Thus, they generated a graph-based summary with the most salient information for summarization. Rehman et al. [22] proposed a model with a GRU-based encoder-decoder for abstractive text summarization.

B. PRE-TRAINED LANGUAGE MODELS FOR THE ATS

Pre-trained language models (PTLMs) have made significant progress in abstractive methods and many NLP tasks.

TABLE 1. The summary of the scientific text summarization literature.

Ref.	Target	Methodology	Dataset	Evaluation Metric	Performance
[3]	Extractive	Modified quantum-inspired genetic algorithm	DUC 2005; DUC 2007	Rouge	DUC 2005: R1:47.67, R2:12.87, RSU4: 18.85 DUC 2007: R1:41.06, R2:08.98, RSU4: 14.72
[11]	Extractive	Graph structure, statistical-based, semantic-based, and centrality-based methods	DUC 2001 DUC 2002	Rouge	DUC 2001: R1:51.37, R2:27.16, RL:47.36, RSU4: 25.65 DUC 2002: R1:53.37, R2:28.58, RL:49.79, RSU4: 27.65
[12]	Extractive	Word2vec, Decision Tree, Random Forest, Multi-Layer Perceptron, Gradient Boosting	CSPubSum AlPubSum BioPubSumm	Rouge	CSPubSumm: RL:31.6 AlPubSum: RL:28.00 BioPubSumm: RL:28.90
[13]	Extractive	The graph-based multi-document summarization, Reference paper, Graph Convolutional Network	CL-SciSumm SciSummNet	Rouge, Human Evaluation	SciSummNet: R2:29.30 R3:24.65 RSU4:18.56 CL-SciSumm:R2:18.46, R3 12.77 RSU4:12.21
[14]	Extractive Abstractive	SciBERT, PageRank	LongSumm	Rouge, Human Evaluation	R1:40.90, R2:9.52, RL:15.47
[16]	Extractive Abstractive	BART, Shuffled Data	SciTLDR	Rouge, Human Evaluation	R1:43.8, R2:21.3, RL:35.5 R1:31.7, R2:11.1 RL: 25.0
[5]	Abstractive	Graph network, auto-regressive generator, knowledge domain	CORD-19 PubMed	Rouge	CORD-19: R1:33.68, R2:22.56, RL:32.84 PubMed: R1:33.03 R2:13.51 RL:29.30
[17]	Abstractive	Seq2seq model, AraVec based word2vec	CNN/Daily Mail, Specific dataset	Rouge, Human Evaluation	R1:38.6, R1-Noorder:46.5, R1-stem: 52.6, R1-context:58.1
[4]	Abstractive	Graph Attention Networks, BioBERT, SciBERT	CORD-19	Rouge, Human Evaluation	R1:44.56, R2:18.89, RL: 36.53
[18]	Abstractive	Multiple timescales gated recurrent unit (MTGRU), The pointer-generator-coverage network	CNN/Daily Mail, Specific dataset	Rouge	CNN/Daily Mail: R1:40.94, R2:18.14, RL:38.57 Specific dataset: R1:56.91, R2:37.48, RL:54.02

The main aim of this study is to analyze the relations at the sentence/token level with large-scale corpora. Most researchers have proposed many language models to enhance specific NLP tasks. These are performed through two basic strategies: feature-based and fine-tuning. The feature-based approach requires task-specific architectures with pre-trained representations as additional features. In the fine-tuning approach, a classification layer must be added to the pre-trained model. Fine-tuning approaches are widely preferred to enhance the quality of the generated summaries in ATS [23], [24].

PTLMs are preferred in general-domain text summarization tasks and have achieved successful results. The BERT [24] model was constructed to pre-train deep bidirectional representations of unlabeled text. This model has only an encoder. Therefore, it is stated that it cannot be suitable for abstractive approaches. To address this problem, many researchers have proposed novel models based on BERT. BERTSUMABS [25] is a model that consists of an encoder and a decoder. GSUM [26], based on a neural encoder-decoder, is a model that takes several types of external guidance as input text. The Text-To-Text Transfer Transformer model (T5) [27], based on encoder-decoder architecture, aims to generate a novel text from the text it receives as input. Refactor [28] is a model consisting of a two-stage training process to identify candidate summaries from both document sentences and different base model outputs. To obtain semantic results from scientific articles, the researchers have focused on the SciBERT model, which has the same architecture and configuration as BERT. SciBERT is a model with a maximum sequence length of 512 tokens pre-trained with large-scale scientific papers (1.14M papers) collected from semantic scholars related to different disciplines [4], [6].

C. GRAPH TRANSFORMER NETWORK (GTN)

GTN is a model designed to learn node representation and identify the relations between disconnected nodes in graph structures [29]. In most studies related to text summarization [4], [8], [30], [31], [32], [33], Graph Attention Transformer (GATs) are widely used among graph neural network-based approaches. GATs [34] are models with an attention-based architecture constructed to operate data in a graph structure. The main aim is to find the representations of each node in the graph by adopting the attention mechanism. With GATs, the hierarchical structure of the document can be handled as a whole. In addition, meaningful relationships can be revealed between sentences, tokens, or entities through the preservation of the global context [43].

According to [4], GATs are successful in representing word co-occurrence graphs because they utilize a masked self-attention mechanism to capture dependency between neighbors and prevent information flow between disconnected nodes. According to [30], GATs can extract the hierarchical structure of a document simultaneously as tokens, sentences, paragraphs, and documents. In addition, it provides consistency with the multi-head attention module in the BERT model. According to [31], the advantage of GAT is that it can enhance the impact of the most salient parts of the source document. According to [32], GATs can effectively capture the content in the source document through propagating contextual information [33] proposed

TABLE 2. The properties of the datasets.

Dataset	Doc. Size	Target Avg. Word	Source Avg. Word	Avg. Entities	Avg. Relation
SciSummNet	710	132.90	338.18	13.35	6.36
SciTLDR	1548	166.63	3036.71	15.29	10.20
ArxivComp	16.807	158.24	549.06	16.14	5.92

mix-order graph attention networks for handling indirectly connected nodes inspired by the traditional model of GATs. According to [8], the use of self-attention in GATs constrains the vertex updates of information from adjacent nodes, despite eliminating the deficiencies of previous methods based on graph convolutions. Therefore, a graph transformer encoder built on the GATs architecture was proposed. It provides a more global contextualization of each vertex with a transformer-style architecture. However, GATs [33] have not achieved as many efficient results as GTN. In this paper, a graph-based abstractive summarization (GBAS) approach consisting of three stages was proposed, inspired by the SciBERT and graph transformer, to generate a summary from the text.

III. PROPOSED MODEL

The framework of the proposed model is illustrated in Figure 1. The proposed model consists of five stages: dataset preparation, text encoder, graph construction, graph encoder, and summary decoder. First, the dataset is prepared with the text section and abstract of the scientific articles and the features extracted through the SciIE system. Second, word embedding is obtained with the output in the last layer of SciBERT from the text. Then, a knowledge graph is constructed with features. The knowledge graph is encoded with a graph transformer. In the last stage, a summary is obtained from the knowledge graph. These are explained in detail below.

A. DATASET PREPARING

In the current research, abstractive summaries are generated from the title or full text of scientific articles [8], [35]. The title is inadequate to sufficiently summarize the substance of the document. Therefore, it is foreseen within the scope of this study that the summaries generated from the text can improve the performance of scientific text summarization. In this study, three datasets of different lengths were prepared. First, the scientific articles up to April 2023 from the arXiv website was crawled containing current topics related to computer science such as "fingerprint", "image processing", "natural language processing", "cyber security" and "machine learning" for the ArxivComp dataset. The dataset properties are given in Table 2. Considering the following factors, the SciIE system has been preferred for extracting salient information from scientific articles:

- Traditional Information Extraction (IE) systems [36], [37] in the scientific domain are designed to obtain these within sentences. However, SciIE makes it possible to extract information by taking it into account across sentences [7].
- The SciIE system is designed to identify six entity types (task, method, metric, material, other-scientific-term, and generic), seven relationship types (compare, part-of, conjunction, evaluate-for, feature-of, used-for, and hyponym-of), and co-reference annotations are used to obtain entity types and relations annotations.

B. THE GRAPH-BASED ABSTRACTIVE SUMMARIZATION MODEL: (GBAS)

Text Encoder: SciBERT was used as the text encoder. The main aim of the pre-trained model is to analyze the relations at the sentence/token level with large-scale corpora. SciBERT has a multi-transformer architecture [24]. To obtain the corresponding sequences given in the samples as word embedding, we used the output of the last hidden state in SciBERT. Thus, word embedding was obtained for the text of each article as follows:

$$S = \left(\underbrace{W_{11}, W_{12}, \ldots, W_{1K}}_{word_1}, \ldots, \underbrace{W_{n1}, W_{n2}, \ldots, W_{nm}}_{word_n} \right) \quad (1)$$

where S represents source word sequences. The sub-word of each word is shown with w_{nm}, in which n and m indicate the order of words and sub-word order, respectively. In response to S, the target word sequence is obtained with $T = \{t_1, t_2, \ldots, t_p\}$.

Graph Construction: The graph transformer operates the graph as an input. To construct the graph, the entities, their relations, and co-reference annotations for each text are established as outlined in the dataset preparation. Then, it benefits from the graph preparation process of the GraphWriter model [8] based on [38]. Differently, the graph is constructed by considering the text and abstract sections together. According to the document structure given as an example in Table 3, the stage of graph construction is as follows:

- The "abstract-relations" and "document-relations" arrays are rebuilt according to the "relations types", "abstract-entities", and "document-entities" indexes.
- The "abstract-relations" array is converted to "0 1 1", according to the example of "brain tumor segmentation — CONJUNCTION — treatment outcome evaluation", so that the index of brain tumor segmentation is "0", the index of CONJUNCTION is "1" and the index of treatment outcome evaluation is "1".
- The "abstract-relations" array is converted to "2 0 3", according to the example of "Deep learning techniques – USED-FOR – brain tumor segmenta-tion-method", so that the index of deep learning techniques is "2", the index of USED-FOR is "0" and the index of brain tumor segmentation method is "3".
- Entity names that could not be extracted because of spelling errors were excluded from the novel array.

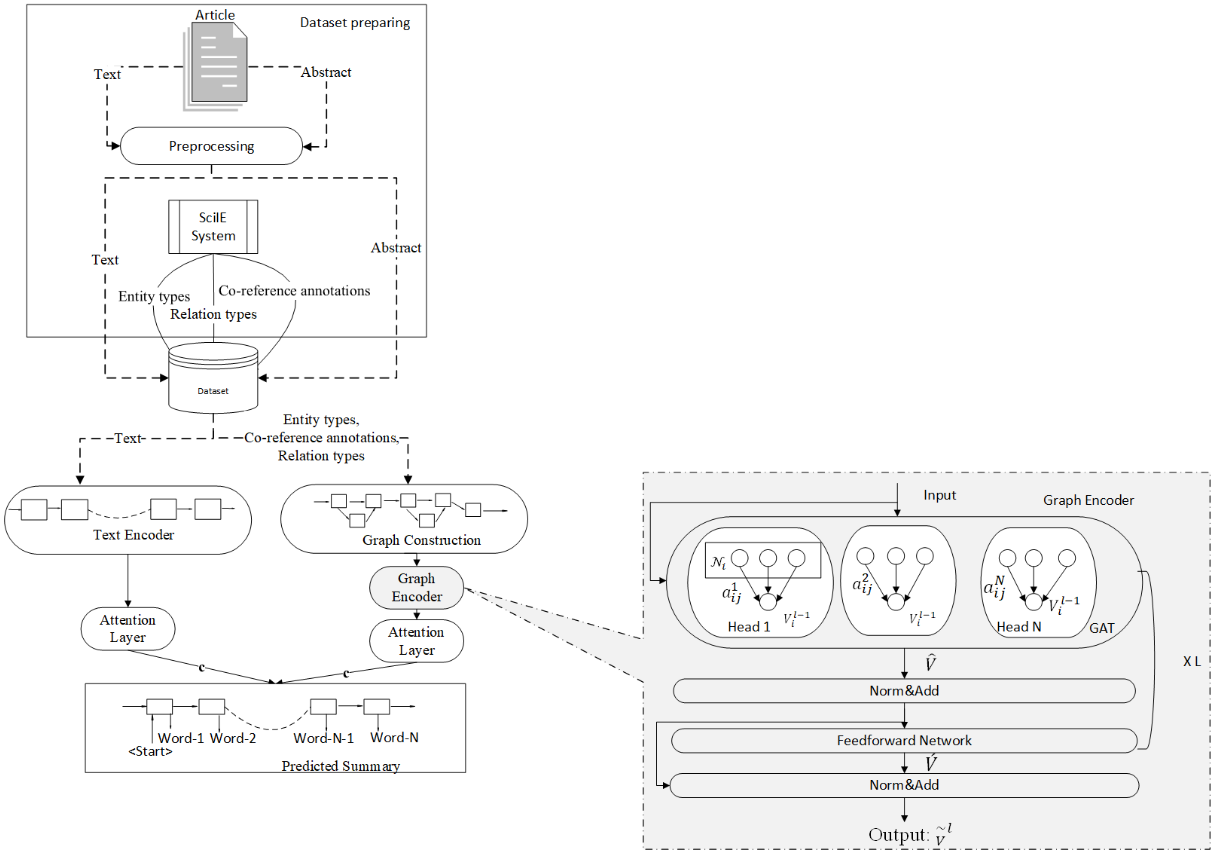


FIGURE 1. The overview of the proposed model.

Accordingly, the "abstract-relations" array is [0 1 1; 2 0 3; 4 1 5; 9 0 8; 9 1 10; 10 0 8; 12 0 8; 10 0 8; 13 0 81; 9 0 17; 21 0 17; 23 0 22; 28 0 22; 28 1 29; 29 0 22].

- The same transformation was performed in the text.
- The array of "document-relations" is [1 1 2; 19 0 18; 15 0 21; 25 0 26; 25 1 28; 28 0 26; 25 0 30; 31 0 5; 34 0 5; 34 0 36; 37 6 36; 39 6 36; 40 0 41; 49 1 50; 54 0 53; 55 6 54; 60 0 52; 63 0 66; 66 0 67; 34 0 75].
- As a result, a comprehensive graph was constructed by combining the new transformation array of "abstract-relations" and "document relations".

where the list of entities and relations retrieved from the abstract section with the SciIE system is referred to by the terms abstract-entities and abstract-relations. The list of entities and relations that the SciIE system collected from the text is referred to as the document-entities and document-relations. The procedures mentioned above are performed with Equations 2, 3, and 4. Figure 2 shows the graph constructed for the text and abstract sections. G_1 and G_2 denote global nodes. G_1 and G_2 represent global nodes and contain entity name lists in Table 3. To maintain the flow of information in the graph, the global node is connected to all nodes. This node is used as the start of the decoder. Nodes consist of entity names. Each labeled edge is replaced by two nodes. One of them represents the forward direction of the relations (Rel.), and the other represents the reverse direction of the relations. A novel node is connected to nodes consisting of entities (Ent.), preserving the directions of previous edges.

$$G = (G_1 \cup G_2) \tag{2}$$

$$All - Ent. = Abstract - Ent. \cup Document - Ent. \tag{3}$$

$$All - Rel. = Abstract - Rel. \cup Document - Rel. \tag{4}$$

Within the scope of the study, entity names (abstract-entities, document entities), relations (abstract-relations, document-relations), and global nodes (G_1 and G_2) in Equations 2, 3, and 4 are combined to build the comprehensive graph in Figure 3.

Graph Encoder: To encode the graph structure, a graph transformer architecture is based on GATs. This model uses the N-head self-attentional system, as shown in Figure 4. In this model, each vertex is contextualized by attending to another connected vertex in the graph. Equations (5) and (6) are utilized in the computation of the N-independent attentions.

$$\widehat{v_i} = V_i + \|_{n=1}^{N} \sum_{j \epsilon n_i} a_{ij}^n W_V^n V_j \tag{5}$$

$$a_{ij}^n = a^n(V_i, V_j) \tag{6}$$

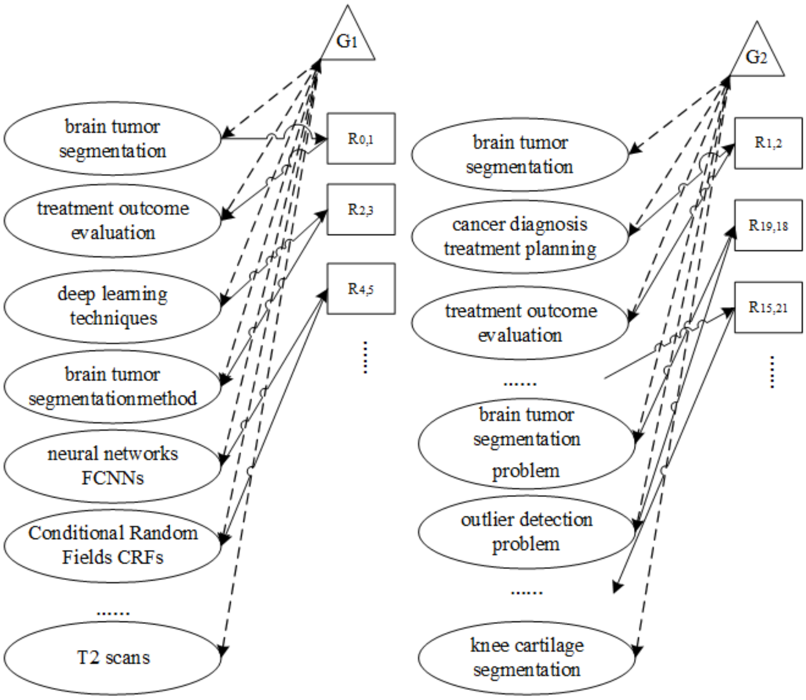


FIGURE 2. The example of relation graphs a) The abstract graph b) The text graph.

where, N_i indicates each vertex representation with the participation of the other vertices in G, α^n, and W_V^n are the attention mechanisms parameterized per head, and $\|$ indicates the concatenation of the N-head attention, N_i represents the neighborhood of V_i. For each head, independent transformations are learned with α respectively. Equation (7) is used for attention functions.

$$a(q_i, k_i) = \frac{exp((W_k k_j)^T w_Q q_i)}{\sum_{z \epsilon N_i} exp((W_k k_j)^T w_Q q_i)} \tag{7}$$

where, $W_K W_Q \epsilon \mathbb{R}^{dxd}$ learnable parameters of attention mechanism, N_i represents the neighborhood of v_i. In this model, equations (8) and (9) are applied L times for each block: Where $FNN(x)$ donates a two-layer feedforward network. As a result, each vertex encoding indicates with $V^L = \left[V_i^L\right]$ which consists of relation, entities, and global vertex in G.

$$\widehat{v}_i = LayerNorm(\widehat{v}_i + LayerNorm(\widehat{v}_i) \tag{8}$$
$$v_i' = FFN(LayerNorm(\widehat{v}_i)) \tag{9}$$

Summary Decoder: In this stage, the content vectors (c) obtained from the graph and text sequences are calculated by adding a decoder hidden state h_t at each t timestep. While vertex embedding V^L is used for graph sequences, and T is used for the text sequence. This is given in equation (10). In the last stage, it is given as input to RNN with h_t by concatenating the context vectors namely c_t from both graphs and the text.

$$c_t = \begin{cases} h_t + \|_{n=1}^{N} \sum_{j \epsilon V} a_j^n W_G^n V_j^L, & graphsequence \\ \quad a_j = a(h_t, V_j^L) & \\ h_t + \|_{n=1}^{N} \sum_{j \epsilon V} a_j^n W_G^n V_j^L, & textsequence \\ \quad a_j = a(h_t, T_j) & \end{cases} \tag{10}$$

where c_t represents the context vectors, h_t refers to the hidden state at each t timestep, V^L represents vertex embedding in G, T represents the text sequence embedding, and W_V^n are the attention mechanisms parameterized per head. To calculate the probability of copying from the input, it is applied in equation (11) [44]. Taking into account this equality, the probability of the final next token is given by equation (12).

$$p = softmax(W_{copy}[h_t, \|c_t] + b_{copy}) \tag{11}$$
$$p * a^{copy} + (1-p) * a^{vocab} \tag{12}$$

where p refers to the probability of copying from the input. a^{copy} and a^{vocab} refer to the probability distribution over entities and input tokens, and remaining probability respectively.W_{copy} and b_{copy} are learnable parameters. In a^{copy}, $a([h_t, \|c_t], x_i)$ is performed for $x_i \epsilon V \| T$.

IV. EXPERIMENTS

To ensure a fair comparison of the proposed model's performance, GBAS was performed using the default hyperparameter settings for the three datasets. SciBERT was used [6] by following the default hyperparameter settings.

TABLE 3. The example of source document.

Type	Example
"relations types:"	["USED-FOR", "CONJUNCTION", "FEATURE OF", "PART OF", "COMPARE", "EVALUATE FOR", "HYPONYM-OF"]
"abstract-entities":	"entities": ["brain tumor segmentation", "treatment outcome evaluation", "deep learning techniques", "brain tumor segmentationmethod", "neural networks FCNNs", "Conditional Random Fields CRFs", "unified framework", "appearance and spatial consistency",..., "T2 scans"]
"abstract-relations":	"relations": ["brain tumor segmentation – CONJUNCTION – treatment outcome evaluation", "deep learning techniques – USED-FOR – brain tumor segmentationmethod", "neural networks FCNNs – CONJUNCTION – Conditional Random Fields CRFs", "2D image patches – USED-FOR – deeplearning based segmentation model", "2D image patches – CONJUNCTION – image slices",..., "T2 scans – USED-FOR – those"]
"document-entities":	["brain tumor segmentation", "cancer diagnosis treatment planning", "treatment outcome evaluation", "manual segmentation of brain tumors", "semi automatic or automatic brain tumor segmentation methods", "brain tumor segmentation studies", "gliomas", "magnetic resonance imaging mri", "segmentation of gliomas", "appearance", "gliosis", "mri data gliomas", "fuzzy boundaries",...,"knee cartilage segmentation"]
"document-relations":	["cancer diagnosis treatment planning – CONJUNCTION – treatment outcome evaluation", "manual segmentation – USED-FOR – manual segmentation of brain tumors", "mri data – USED-FOR – segmentation of gliomas", "discriminative model based methods – CONJUNCTION – discriminative model", "probabilistic image atlases – USED-FOR – brain tumor segmentation problem", "outlier detection problem – USED-FOR – brain tumor segmentation problem", "discriminative model based methods – USED-FOR – tumor segmentation problem",..., "back propagation algorithms – CONJUNCTION – mrfs crfs"]

A graph transformer [8] based on the default parameter setting that activation function, dropout, attention head, graph layer (L), and feedforward network in block layer size were PReLU, 3, 6, 4, 2000, respectively, was used to encode graphs. To decode, a beam search was used with size 4. At the last stage, the words mentioned as $< unk >$ were removed from the generated summaries.

A. EVALUATION METRIC

To evaluate the similarity of the reference summary (human-created) and the generated summary by the GBAS model, the Rouge metric (Recall Oriented Understudy of Gisting Evaluation) was used. ROUGE calculates n gram-based recall to determine whether the abstract written by the author overlaps with the summarization generated by the model [39]. The most commonly preferred metrics are ROUGE-1, ROUGE-2, and ROUGE-L, which overlap the uni-gram, bi-gram, and the longest common sequences (LCS) in the word level, respectively. The calculation of the ROUGE-N is shown in the equation (13):

$$ROUGE - N = \frac{\sum_{S \epsilon H} \sum_{gram_n \epsilon S} Count_{match}^{gram_n}}{\sum_{S \epsilon H} \sum_{gram_n \epsilon S} Count(gram_n)} \quad (13)$$

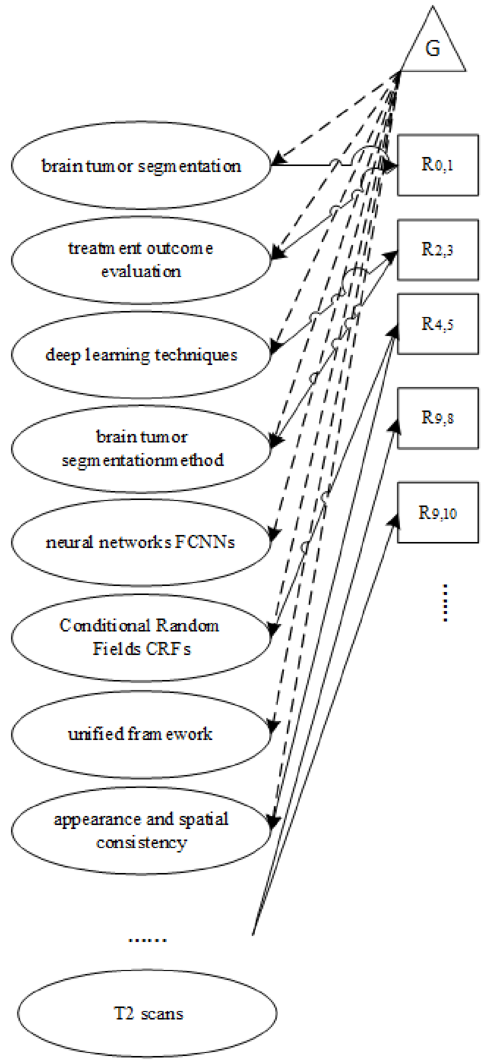


FIGURE 3. The graph representation.

where N, $Count(gram_n)$, $Count_{match}^{gram_n}$ are represented as the length of the N-gram, the count of n-grams in the references, and the maximum number of matching words in the candidate summary, H refers to the sentences in the reference summary.

B. BASELINE METHODS

To evaluate the performance of the proposed model GBAS, experiments were conducted in a baseline approach for both abstractive and extractive methods. These methods are summarized as follows:

- TextRank [40] is a graph-based method applied to extractive methods.
- LexRank [41] is a graph-based method that calculates based on cosine similarity. Based on the eigenvector centrality, the graph's representation of the sentences is shown.
- LSA [42] is an algebraic method for extractive methods. This model is performed in three stages: input matrix

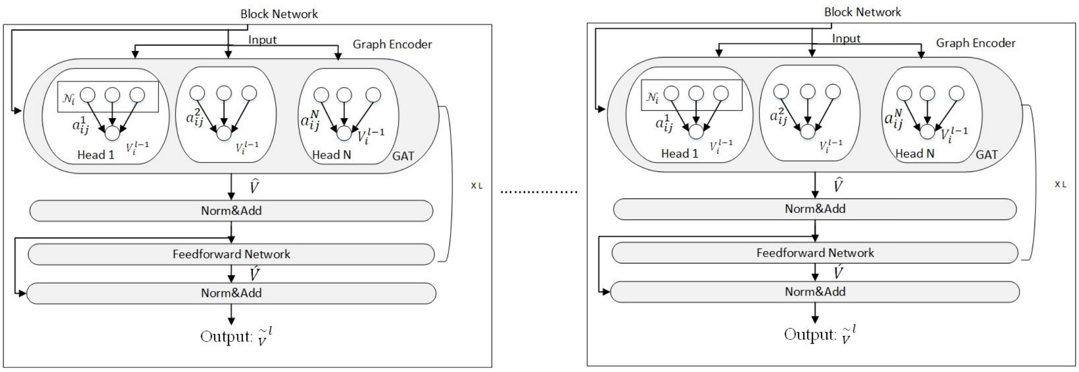


FIGURE 4. Graph encoder model [8].

creation, singular value decomposition (SVD), and sentence selection.

- A fine-tuned Text-to-Text Transfer Transformer (T5) model is used for text summarization trained with 4515 English news articles.[1]
- Billsum [45] is a T5-based fine-tuned model on the Billsum dataset which consists of US Congressional bills.
- BART [43], which uses the standard sequence-to-sequence Transformer architecture, is effective when the model is fine-tuned for the abstractive summarization task. The BART fine-tuned model is trained to summarize scientific articles.[2]
- GraphWriter [8] is a graph-based tool that generates a summary from the title for abstractive methods.

C. RESULT AND DISCUSSION

The experiments were conducted with the above baseline methods on the SciTLDR and SciSummNet datasets and ArxivComp. The datasets (ArxivComp, SciSummNet, and SciTLDR) have average word lengths of 549.06, 338.18, and 3036.71, respectively. The results of the experiment are shown in Tables 4, 5, and 6.

Graph-based methods are highly successful among abstractive and extractive methods. Because Seq2Seq models are not very good at transferring information in long-term sequences, abstractive methods concentrate on attention-based methods. Attention-based models fix this issue, but when taking the document's hierarchical structure into account, they may result in a loss of semantic integrity. Therefore, graph-based methods are superior to other methods because they ensure the integrity of the document. Compared to the proposed method, graf-based methods performed better on a dataset of average word lengths 338.18, with a difference of 1.42 in Rouge-L.

According to the results shown in Table 4, the proposed model outperforms the baseline methods on the ArxivComp dataset. The GraphWriter model achieves the closest result to the proposed model among the abstractive methods. This model generates the summary from the title, whereas the proposed model generates the summary from the text. This is how this model differs from the proposed model. Token embedding for the proposed model used SciBERT, which was trained using scientific articles. The text and abstract sections were combined to create the proposed model's graph. The main difference between these approaches and others (BART, T5-based, and Billsum) is the restriction on token sequence lengths. The text section is also of variable length for each article. Therefore, the summaries produced by models do not correspond with the author's summary for articles of different lengths.

TABLE 4. The evaluation results on the ArxivComp dataset.

Model	Rouge-1	Rouge-2	Rouge-L
TextRank	28.52	9.20	25.67
LexRank	36.63	10.94	33.18
LSA	30.18	8.02	27.90
BART (fine-tuned)	24.39	9.52	23.17
T5-model	27.00	13.53	21.79
Billsum	30.96	22.87	28.86
GraphWriter	43.63	18.63	36.31
GBAS (proposed)	45.05	19.35	37.10

The auto-regressive decoder in the BART model enables it to achieve the greatest summarization performance when comparing the results of different models. As can be seen in Table 6, graph-based methods are more successful than baseline methods in summarizing long documents because they preserve the integrity of the document.

The performance of the model for both Billsum and T5-model approaches dramatically declines as the document length increases. However with the SciSummnet dataset—whose average word length is shorter—better performance was obtained in both models. These results indicate that T5-based approaches are insufficient for summarizing long scientific articles.

Examining the results of the extraction methods (TextRank, LexRank and LSA) reveals that these methods are not

[1]https://huggingface.co/mrm8488/t5-base-finetuned-summarize-news
[2]https://huggingface.co/sana-ngu/bart-base-finetuned-summarize-scientific-articles

TABLE 5. The evaluation results on the SciSummNet dataset.

Model	Rouge-1	Rouge-2	Rouge-L
TextRank	33.77	18.39	31.99
LexRank	34.22	18.18	32.39
LSA	34.25	18.13	32.30
BART (fine-tuned)	42.37	18.49	41.37
T5-model	41.67	27.67	36.41
Billsum	43.27	35.79	39.65
GraphWriter	43.64	15.11	36.94
GBAS (proposed)	45.05	15.11	38.36

TABLE 6. The evaluation results on SciTLDR dataset.

Model	Rouge-1	Rouge-2	Rouge-L
TextRank	16.11	4.34	13.67
LexRank	16.68	4.50	14.07
LSA	18.22	4.93	15.26
BART (fine-tuned)	25.24	4.41	25.25
T5-model	14.85	8.48	7.51
Billsum	16.11	10.11	9.15
GraphWriter	42.19	15.46	34.62
GBAS (proposed)	42.93	14.90	34.96

TABLE 7. Human evaluation results.

	Con.	In.	Coh.	R.	G.
Mean/Var.	3.11(0.57)	4.06 (0.61)	4.03(0.56)	3.92(0.51)	3.15(0.56)
Fleiss kappa	0.671	0.611	0.534	0.603	0.613

as successful as the abstraction methods. It is seen that the lowest results are obtained in the SciTLDR dataset. The main reason for this is that as the length of the document increases, the number of sentences containing more general information also increases. When sentence selections are performed with these algorithms, the overlapping rate of the summary is decreased.

As can be seen in Tables 4 and 6 the proposed method outperforms baseline methods on long documents. As can be seen in Table 5, it is a comparable method for documents with a shorter average word length. An advantage of the proposed method is that it handles the document hierarchically for long documents.

D. HUMAN EVALUATION

The Rouge metric compares the generated summary based on the overlapping of the n-grams with the ground-truth summary. However, it is not sufficient to prove the quality of the generated summaries. To overcome this problem, the generated summaries were also evaluated with human judgment. The evaluation criteria are as follows: 1) Conciseness(Con.) is whether you avoid redundant information; 2) Informativeness(I) is whether it contains salient information; 3) Coherence(Coh.) is whether the content of the generated summary is appropriate for the ground-truth summary; 4) Readability(R) means that the generated summary is easy to understand and fluent; and 5) Grammatically(G), the question is whether the sentences are appropriate to the grammar rules.

TABLE 8. The samples of generated summaries.

Summary Type	Summary
Gen:	"Aiming at the problems of color distortion, blur, and excessive noise in underwater images, an underwater image enhancement algorithm-based texture reconstruction is proposed. First, the color equalization of the degraded image is realized by the automatic color enhancement algorithm. Second, the relative total variation is introduced to decompose the structure layer and texture layer. Then, the best background light point is selected based on brightness gradient discrimination, and the transmittance of the backscatter component is obtained by the red-dark channel prior, which is inserted into the imaging model to remove the fogging phenomenon in the structure layer and enhance effective details in the texture layer by a multi-scale detail enhancement algorithm and binary mask. Finally, the structure layer and texture layer are reconstructed to get the image. The experimental results show that the hue saturation and clarity of the underwater image can be improved and have good performance in different underwater environments."
Gen:	" We describe the use of a degraded aerial image for a reflectance component. The term directional sensitivity and the adaptive variant of total generalized variation in uneven illumination are obtained from the reflectance image. The illumination of the reflectance image is determined by the adaptive variant of total generalized variation based on the adaptive variant of total generalized variation. The key difference of this paper is then applied to a reflectance component to solve the derived function in an objective function. The reflectance image of reflectance image can only be reconstructed with an adaptive variant of total generalized variation. The result is tested in terms of three stages of results on the state of the art in the art restoration methods and discusses the development of an objective function for the derived function. The nonlocal retinex model is compared to the state of the art.".
Gen:	"In this paper, we propose a way of global restoration task which has many useful applications eg denoising we attempt to accomplish such synthesis given a source image and target text description of using only a single image one main focus of this paper we consider the problem of denoising our goal is to estimate where denoising where an image to the denoised image from we employ an end to leading denoising methods that segments the sos procedure for the problem of the degraded input image in the sos procedure to capture the denoised image from in the space of the sos algorithm the proposed the training set the test set which could be used to improve the existing image denoising methods extensive experiments show the effectiveness of denoising and can be conducted on the segmentation of denoising and in vivo showing images"

Five expert volunteers rated the summaries from 1 (worst) to 5 (best) for each criterion. Fleiss's Kappa analysis was performed to determine whether the evaluator scores were compatible with each other.

From the results in Table 7, it is seen that the evaluations of the volunteers mostly agree with each other in that the result of each criterion is greater than 0.5. According to the results, the generated summary is generally informative, fluent, and overlaps with a ground-truth summary. However, it has been observed that grammatical problems remain. For instance, some words repeat more than once in the generated summary.

In addition, it adversely affects conciseness and readability because $< unk >$ and some entities extracted as "generic" by the SciIE system are not learning. It caused inconsistencies in the meaning of sentences because these words were removed from the generated summaries at the last stage. Because the proposed model is pretrained with topics related to computer science, it will produce successful summaries on these topics. In Table 8, sample abstracts generated for scientific articles related to image degradation are given.

V. CONCLUSION

This paper addresses the issue of a lack of semantic consistency between words in scientific articles which makes it challenging to summarize the content. This study proposes a graph-based model for summarizing long scientific articles as a solution to this problem because it preserves the hierarchical integrity of a graph-based model. The graph has been created using the terms from the scientific publication that the SciIE system is capable of extracting from text and abstract. The performance of the proposed model was evaluated with baseline methods on three datasets with different average word lengths. The proposed model is outperformed in summarizing long documents. The results of the human evaluation show that the proposed model generally generated an informative, fluent, and overlapping ground-truth summary. The proposed model is trained with current topics related to computer science. The proposed model provides superior summaries of topics in this field. The limitation of the model is that it is insufficient to summarize documents containing mathematical expressions, figures, and tables. These sections were eliminated from the study during the preprocessing phase. In light of these, it may be of great value to condense the articles for future study. Articles in science articles are published frequently. The definitions of terminology used in scientific documents may differ. The dataset updated regularly will be more useful for research in the future.

REFERENCES

[1] A. Alomari, N. Idris, A. Q. M. Sabri, and I. Alsmadi, "Deep reinforcement and transfer learning for abstractive text summarization: A review," *Comput. Speech Lang.*, vol. 71, Jan. 2022, Art. no. 101276.

[2] M. Zhong, P. Liu, D. Wang, X. Qiu, and X. Huang, "Searching for effective neural extractive summarization: What works and what's next," 2019, *arXiv:1907.03491*.

[3] M. Mojrian and S. A. Mirroshandel, "A novel extractive multi-document text summarization system using quantum-inspired genetic algorithm: MTSQIGA," *Expert Syst. Appl.*, vol. 171, Jun. 2021, Art. no. 114555, doi: 10.1016/j.eswa.2020.114555.

[4] X. Cai, S. Liu, L. Yang, Y. Lu, J. Zhao, D. Shen, and T. Liu, "COVIDSum: A linguistically enriched SciBERT-based summarization model for COVID-19 scientific papers," *J. Biomed. Informat.*, vol. 127, Mar. 2022, Art. no. 103999, doi: 10.1016/j.jbi.2022.103999.

[5] Y. Du, Y. Zhao, J. Yan, and Q. Li, "UGDAS: Unsupervised graph-network based denoiser for abstractive summarization in biomedical domain," *Methods*, vol. 203, pp. 160–166, Jul. 2022, doi: 10.1016/j.ymeth.2022.03.012.

[6] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A pretrained language model for scientific text," 2019, *arXiv:1903.10676*.

[7] Y. Luan, L. He, M. Ostendorf, and H. Hajishirzi, "Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction," 2018, *arXiv:1808.09602*.

[8] R. Koncel-Kedziorski, D. Bekal, Y. Luan, M. Lapata, and H. Hajishirzi, "Text generation from knowledge graphs with graph transformers," 2019, *arXiv:1904.02342*.

[9] A. Fan, D. Grangier, and M. Auli, "Controllable abstractive summarization," 2017, *arXiv:1711.05217*.

[10] Z. Liang, J. Du, and C. Li, "Abstractive social media text summarization using selective reinforced Seq2Seq attention model," *Neurocomputing*, vol. 410, pp. 432–440, Oct. 2020, doi: 10.1016/j.neucom.2020.04.137.

[11] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "EdgeSumm: Graph-based framework for automatic text summarization," *Inf. Process. Manage.*, vol. 57, no. 6, Nov. 2020, Art. no. 102264, doi: 10.1016/j.ipm.2020.102264.

[12] L. Cagliero and M. La Quatra, "Extracting highlights of scientific articles: A supervised summarization approach," *Expert Syst. Appl.*, vol. 160, Dec. 2020, Art. no. 113659, doi: 10.1016/j.eswa.2020.113659.

[13] M. Yasunaga, J. Kasai, R. Zhang, A. R. Fabbri, I. Li, D. Friedman, and D. R. Radev, "ScisummNet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks," 2019, *arXiv:1909.01716*.

[14] J. Ju, M. Liu, L. Gao, and S. Pan, "SciSummPip: An unsupervised scientific paper summarization pipeline," 2020, *arXiv:2010.09190*.

[15] Q. Zhou, F. Wei, and M. Zhou, "At which level should we extract? An empirical analysis on extractive document summarization," 2020, *arXiv:2004.02664*.

[16] I. Cachola, K. Lo, A. Cohan, and D. S. Weld, "TLDR: Extreme summarization of scientific documents," 2020, *arXiv:2004.15011*.

[17] D. Suleiman and A. Awajan, "Multilayer encoder and single-layer decoder for abstractive Arabic text summarization," *Knowl.-Based Syst.*, vol. 237, Feb. 2022, Art. no. 107791, doi: 10.1016/j.knosys.2021.107791.

[18] D. S. Moirangthem and M. Lee, "Abstractive summarization of long texts by representing multiple compositionalities with temporal hierarchical pointer generator network," *Neural Netw.*, vol. 124, pp. 1–11, Apr. 2020, doi: 10.1016/j.neunet.2019.12.022.

[19] G. Bao, Z. Ou, and Y. Zhang, "GEMINI: Controlling the sentence-level summary style in abstractive text summarization," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2023, pp. 831–842.

[20] M. Cao, Y. Dong, J. He, and J. C. K. Cheung, "Learning with rejection for abstractive text summarization," 2023, *arXiv:2302.08531*.

[21] M. Li, J. Qi, and J. H. Lau, "Compressed heterogeneous graph for abstractive multi-document summarization," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 11, pp. 13085–13093.

[22] T. Rehman, S. Das, D. K. Sanyal, and S. Chattopadhyay, "Abstractive text summarization using attentive GRU based encoder–decoder," in *Applications of Artificial Intelligence and Machine Learning*. Singapore: Springer, 2022, pp. 687–695.

[23] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever. (2018). *Improving Language Understanding by Generative Pre-training*. [Online]. Available: https://www.mikecaptain.com/resources/pdf/GPT-1.pdf

[24] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[25] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," 2019, *arXiv:1908.08345*.

[26] Z.-Y. Dou, P. Liu, H. Hayashi, Z. Jiang, and G. Neubig, "GSum: A general framework for guided neural abstractive summarization," 2020, *arXiv:2010.08014*.

[27] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," 2019, *arXiv:1910.10683*.

[28] Y. Liu, Z.-Y. Dou, and P. Liu, "RefSum: Refactoring neural summarization," 2021, *arXiv:2104.07210*.

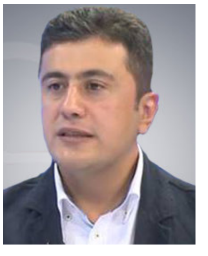
[29] S. Yun, M. Jeong, R. Kim, J. Kang, and H. J. Kim, "Graph transformer networks," 2019, *arXiv:1911.06455*.

[30] B. Zheng, H. Wen, Y. Liang, N. Duan, W. Che, D. Jiang, M. Zhou, and T. Liu, "Document modeling with graph attention networks for multi-grained machine reading comprehension," 2020, *arXiv:2005.05806*.

[31] Y. Zhou, J. Shen, X. Zhang, W. Yang, T. Han, and T. Chen, "Automatic source code summarization with graph attention networks," *J. Syst. Softw.*, vol. 188, Jun. 2022, Art. no. 111257.

[32] H. Chen, P. Hong, W. Han, N. Majumder, and S. Poria, ''Dialogue relation extraction with document-level heterogeneous graph attention networks,'' 2020, *arXiv:2009.05092*.
[33] Y. Zhao, L. Chen, Z. Chen, R. Cao, S. Zhu, and K. Yu, ''Line graph enhanced AMR-to-text generation with mix-order graph attention networks,'' in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, p. 73241.
[34] P. Velič ković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, ''Graph attention networks,'' 2017, *arXiv:1710.10903*.
[35] F. Xie, J. Chen, and K. Chen, ''Extractive text-image summarization with relation-enhanced graph attention network,'' *J. Intell. Inf. Syst.*, vol. 61, no. 2, pp. 325–341, Oct. 2023, doi: 10.1007/s10844-022-00757-x.
[36] Y. Kumar, K. Kaur, and S. Kaur, ''Study of automatic text summarization approaches in different languages,'' *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 5897–5929, Dec. 2021, doi: 10.1007/s10462-021-09964-4.
[37] I. Augenstein, M. Das, S. Riedel, L. Vikraman, and A. McCallum, ''SemEval 2017 task 10: ScienceIE–extracting keyphrases and relations from scientific publications,'' 2017, *arXiv:1704.02853*.
[38] D. Beck, G. Haffari, and T. Cohn, ''Graph-to-sequence learning using gated graph neural networks,'' 2018, *arXiv:1806.09835*.
[39] C. Y. Lin, ''Rouge: A package for automatic evaluation of summaries,'' in *Text Summarization Branches Out*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014, pp. 74–81.
[40] R. Mihalcea and P. Tarau, ''TextRank: Bringing order into text,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Jul. 2004, pp. 404–411.
[41] G. Erkan and D. R. Radev, ''LexRank: Graph-based lexical centrality as salience in text summarization,'' *J. Artif. Intell. Res.*, vol. 22, pp. 457–479, Dec. 2004.
[42] Y. Gong and X. Liu, ''Generic text summarization using relevance measure and latent semantic analysis,'' in *Proc. 24th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Sep. 2001, pp. 679–688.
[43] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, ''BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension,'' 2019, *arXiv:1910.13461*.
[44] A. See, P. J. Liu, and C. D. Manning, ''Get to the point: Summarization with pointer-generator networks,'' 2017, *arXiv:1704.04368*.
[45] T. Rehman, S. Das, D. K. Sanyal, and S. Chattopadhyay, ''An analysis of abstractive text summarization using pre-trained models,'' in *Proc. Int. Conf. Comput. Intell., Data Sci. Cloud Comput.* Singapore: Springer, 2021, pp. 253–264.

MEHTAP ULKER received the B.S. degree in computer engineering from Firat University, and the M.S. degree in computer engineering from Gazi University, Ankara. She is currently pursuing the Ph.D. degree in computer engineering with Firat University. She is also a Research Assistant with the Computer Engineering Department, Firat University. Her research interests include NLP, text summarization, and biometrics.



A. BEDRI OZER is currently a Professor of computer engineering with Firat University. He has authored a number of studies on information security. He has consulted many M.S. and Ph.D. students.