

Date of publication

Integrating Topic-Aware Heterogeneous Graph Neural Network with Transformer Model for Medical Scientific Document Abstractive Summarization

Ayesha Khaliq¹, Atif Khan², Salman Afsar Awan³, Salman Jan⁴, Muhammad Umair⁵, Megat Farez Azril Zuhairi⁶

¹ Department of Computer Science, University of Agriculture Faisalabad, 03802, Faisalabad, Pakistan

² Department of Computer Science, Islamia College Peshawar, KP, 25120, Peshawar, Pakistan

³ Department of Computer Science, University of Agriculture Faisalabad, 03802, Faisalabad, Pakistan

⁴ Department of Information Technology, Alburaimi University College, 512, Alburaimi, Oman

⁵ Department of Computer Science, City University of Science and Information Technology, 25000, Peshawar, Pakistan

⁶ Malaysian Institute of Information Technology, University Kuala Lumpur, 50250, Kuala Lumpur, Malaysia

Corresponding authors: Megat Farez Azril Zuhairi, Salman Jan (e-mail: megatfarez@unikl.edu.my, salman@buc.edu.om).

ABSTRACT The development of abstractive summarization methods is a crucial task in Natural Language Processing (NLP) that presents challenges, which require the creation of intelligent systems that are capable of extracting the main idea from text effectively and generate coherent summary. Numerous existing abstractive approaches do not take into account the importance of the broader context or fail to capture the global semantics in identifying salient content for summary. Moreover, there is lack of studies that extensively evaluated abstractive summarization models for specific domains, such as medical scientific document summarization. With this motivation behind, this paper developed an integrated framework for abstractive summarization of medical scientific documents that integrates topic-aware Heterogeneous Graph Neural Network with a Transformer model. The suggested framework uses Latent Dirichlet Allocation (LDA) for topic modeling to uncover latent topics and global information, thus preserving document-level attributes important for creation of effective summaries. In addition to topic modeling, the framework utilized a Heterogeneous Graph Neural Network (HGNN), capable of capturing the relationship between sentences through graph-based document representation, and allows for the concurrent updating of both local and global information. Finally, the framework is integrated with a Transformer decoder, which greatly enhances the ability of model to produce accurate and informative abstractive summaries. The performance of propose framework is evaluated on publicly available PubMed dataset related to medical scientific papers. Experimental results illustrate that the suggested framework for abstractive summarization showed superior performance as compared to the state-of-the-art models, achieving high F1-Scores: 46.03 for Rouge-1, 21.42 for Rouge-2, and 39.71 for Rouge-L.

INDEX TERMS BERT, LDA, GAT, TF-IDF, TRANSFORMER, MEDICAL DOCUMENTS, ABSTRACTIVE SUMMARIZATION

I. INTRODUCTION

The primary goal of automatic text summarization is to enable a quick review of documents by generating a coherent and concise summary that retains the original essence of the documents [1]. Text summarization can be classified into two categories: extractive and abstractive. Extractive summarization involves the systematic selection of original sentences and content directly from the source document. Conversely, the abstractive summarization technique utilizes sophisticated procedures to significantly transform and rephrase sentences from the original input

documents [2]. Most of the research in text summarization has focused news reports (brief documents), conversation records (collaborative documents), critiques (informal documents), and social media posts (tweets). Conversely, researchers have devoted less attention to summarization in medical domain, specifically for medical scientific documents concerning various diseases.

Abstractive models align more closely with the human nature of summarization and can produce high-quality summaries [3]. Consequently, the research presented in this

paper utilizes an abstractive technique for the summarization of medical scientific documents.

Modeling the relationships between sentences is a crucial step for identifying the relevant content and main theme of the input text in the process of generating a summary [4]. In recent years, several studies [5, 6], utilized Recurrent Neural Networks (RNNs) to grasp the relationships between sentences. However, recurrent models face challenges in accurately capturing long-range dependencies among sentences and efficiently managing computational resources [7]. Modeling global information, especially topical information, constitutes another critical aspect of summarization. It significantly influences the selection of sentences and provides additional context for understanding documents [3, 8]. However, it has been observed that several previous studies [9, 10] have considered topic models as a standalone source of information, neglecting to investigate a cohesive approach that concurrently improves both the task of text summarization (capturing complex inter-sentence relationships within documents) and the effectiveness of topic models.

In recent years, there have been notable achievements in applying Graph Neural Networks (GNNs) or Graph Attention Networks (GAT) [11] to summarize documents [4, 12-17]. This success arises from their ability to capture complex inter-sentence relationships within documents. Specifically, GNN models demonstrate proficiency in representing complex structural data by encapsulating semantic units (nodes) and their interconnections (edges). Due to these capabilities, there is growing interest in leveraging the GAT framework alongside topic modeling for abstractive summarization tasks. The researchers in [12] proposed a heterogeneous neural graph structure using BERT to obtain contextual sentence representations. They simultaneously trained with latent topics using the Neural Topic Model (NTM) to model global information. The study of [18] proposed an expanded model that employed NTM for abstractive text summarization.

The authors of [15] introduced a graph structure enriched with latent topical information. They employed K-Means and Gaussian Mixture Models (GMM) for topic extraction. The study of [3] proposed an innovative Graph-Based Topic-Aware abstractive text summarization model. Initially, documents were encoded with BERT to generate sentence representations. Concurrently, NTM identified potential topics, while GAT refined these sentences and topic nodes. Finally, the sentence embeddings, enriched with topic information, were fed into a Transformer-driven decoder to generate summaries.

Another study by [19] introduced a heterogeneous graph that incorporated topical information for abstractive summarization of Chinese complaint reports. To achieve this, GAT was formulated using sentence context embeddings and latent topics. Both the document context embeddings and topic data were updated concurrently, thereby addressing the issue of semantic fragmentation [3].

Despite the effectiveness of existing methods, there remains a need for a comprehensive, topic-aware neural graph model that integrates global semantic information and is coupled with an advanced decoder designed specifically for abstractive summarization of medical scientific documents.

This study introduces an innovative Topic-aware Heterogeneous Graph Neural Network (HGNN) model aimed at summarizing medical documents. Instead of focusing solely on creating a sentence-level graph, the model integrates additional semantic components such as words and latent topic nodes by incorporating them as auxiliary elements within the graph. The initial phase involves using a BERT-Encoder (Bidirectional Encoder Representations) [20] to encode the entire document, thereby generating sentence and word nodes. Following this, latent topic nodes are identified using a well-established generative statistical topic modeling technique known as Latent Dirichlet Allocation (LDA) [21].

In the next phase, a heterogeneous document graph is constructed, comprising sentences, words, and global topic nodes linked by edge features quantified as TF-IDF values. The representations of these nodes are then refined and updated using GAT. Ultimately, the sentence representations, enriched with topic information, are fed into a Transformer-based decoder to generate abstractive summaries. Our approach differs from previous works. For instance, the research by [3] relies on the Neural Topic Model (NTM) for topical information, which requires complex training setups [15]. Similarly, our method differs from work conducted by [19], which focused on constructing a heterogeneous graph solely for word and topic nodes, specifically for summarizing Chinese complaint reports.

The key contributions of our model are outlined as follows:

- This study introduces a novel Neural Graph structure enriched with topical information for abstractive summarization of medical scientific documents. This approach integrates the LDA Topic Extractor, BERT-Encoder, Graph Attention Network, and Transformer model into a cohesive framework.
- The framework effectively captures global semantic information, guiding the summary generation process and addressing the challenge of limited semantic context in traditional summarization methods.
- Leveraging data from the publicly available PubMed dataset, our research demonstrates consistent and favorable results compared to state-of-the-art heterogeneous graph structures used for both extractive and abstractive summarization tasks.

The paper is organized as follows: Section II critically reviews existing abstractive summarization techniques, Topic models, and Neural Graph-based models. Section III presents the proposed model's methodology in detail. Section IV illustrates a comparative analysis of our model with state-of-the-art approaches. It also details the experimental setup, overall performance, and ablation

studies of the proposed model. Finally, Section V concludes the paper and outlines potential avenues for future research.

II. RELATED WORK

Our research includes three key areas: 1. A review of fundamental methods used for abstractive summarization. 2. An examination of topic-aware models that combine topic modeling with other techniques to improve summarization. 3. An exploration of how neural graph-based networks are used for both extractive and abstractive summarization tasks.

A. ABSTRACTIVE TEXT SUMMARIZATION

Abstractive text summarization is dominated by the sequence-to-sequence framework [22]. Early versions relied on RNNs [23], specifically Long Short-Term Memory (LSTM) units, to encode the input text into a vector representation and then decode it into a summary. [3]. However, LSTMs struggle to capture long-range dependencies in lengthy sequences, leading to poor summarization and inefficient use of computational resources [7]. Attention mechanisms [24] were introduced to address this issue. However, they can still suffer from issues like out-of-vocabulary (OOV) words and repetitive outputs. The Pointer-Generator Network with Coverage [25], tackles these problems by cleverly controlling the inclusion of words during generation, achieving impressive results.

Conventional RNN architectures, although well-suited for NLP tasks due to their ability to handle variable-length sequences, suffer from several limitations. These limitations include challenges in parallelization during training, increased training complexity, and difficulties in capturing long-range dependencies and hierarchical relationships within the data. To address these shortcomings, Google introduced the Transformer model [7], a state-of-the-art architecture that leverages self-attention mechanisms to enable parallel processing of the model. This approach represents a significant leap forward compared to traditional RNNs.

BERT, a powerful pre-trained language model based on Transformers [20], has significantly improved performance in various summarization tasks [26-32]. Leading models like BART [33] (Bidirectional and Auto-Regressive Transformers), PEGASUS [34] (Pre-training with Extracted Gap-sentences for Abstractive Summarization), and ProphetNet [35], all employ Transformer architecture for abstractive summarization task. The researchers in [36] introduced a groundbreaking approach for summarization and question answering using Large Language Models (LLMs). Their approach ensures LLMs are not overloaded with irrelevant data, saving time and resources.

Additionally, scholars of [37] explored how knowledge graphs can enhance summaries [37]. They integrated knowledge graphs with BART and developed multi-source transformer modules that can process both textual and

graph-based information, leading to more accurate and cohesive summaries.

B. TOPIC-AWARE TEXT SUMMARIZATION

Modeling topical information is another crucial aspect for selecting sentences in text summarization [8]. The authors in [12] used BERT to capture contextual information in sentences and employed the Neural Topic Model (NTM) to model global topics concurrently. The study in [18] proposed an expanded model using NTM for abstractive summarization. The authors of [15] introduced a graph structure enriched with topics extracted through clustering methods like K-means and Gaussian Mixture Models (GMM). An unsupervised extractive summarization approach is explored in [38]. This approach integrated clustering with topic modeling (using LDA) to reduce topic bias and utilizes K-Medoid clustering for generating summaries.

Several studies explored integrating topic modeling with summarization techniques. The work in [39] proposed a novel approach that combined BERT for contextual embeddings, an NTM for topic discovery, and a transformer network to capture long-range dependencies. This facilitated an end-to-end process for topic inference and summarization. The study in [40] leveraged latent topic information derived from topic vectors and sequential networks to improve the quality and accuracy of summaries.

The study of [41] effectively combined a graph contrastive topic model with a pre-trained language model to utilize both broad and detailed contextual information for extractive summarization of long documents. Building upon the concept of topic awareness, the study in [3] proposed a novel Graph-Based Abstractive Summarization model. The document undergoes BERT encoding to generate sentence embeddings, capturing the semantic meaning of each sentence. Simultaneously, a Neural Topic Model (NTM) is used to identify potential topics within the document. A heterogeneous document graph is then constructed. This graph incorporates nodes for both the sentence embeddings and the identified topics. To further refine the relationships within the graph, a modified version of Graph Attention Network (GAT) is employed. Finally, the sentence embeddings, enriched with the topic information from the refined graph, are given to a Transformer decoder to generate the summaries.

C. NEURAL GRAPH-BASED TEXT SUMMARIZATION

Modern text summarization research has shifted from homogeneous graphs with static nodes to heterogeneous networks. These networks allow for the inclusion of diverse node types, which represent wide range of textual elements. In addition, they allow for dynamic updates during the summarization process. The approach of [4] constructed a graph neural network based on word co-occurrences within the document to capture word-level relationships. The study in [42] utilized syntactic graph convolutional networks

(GCNs) to model the non-Euclidean structure of documents. This approach effectively captured long-range dependencies beyond simple word order. In addition, an attention mechanism is integrated to focus on relevant content for summarization. The authors of [43] employed a network with three distinct node types: sentences, Elementary Discourse Units (EDUs), and entities. They utilized RST discourse parsing to grasp the relationships between EDUs, which provides a deeper comprehension of the document's structure.

The research in [44] presented a heterogeneous graph network that incorporated information from both words and sentences. The model focused on redundancy dependencies between sentences, and iteratively refined sentence representations through a redundancy-aware graph. This iterative process aimed to enhance the model's ability to capture the essential meaning of each sentence. Recent research examined novel network structures and attention mechanisms for extractive summarization of long documents [45]. One approach utilized a transformer-based architecture within a heterogeneous network and included distinct node types for tokens, entities, and sentences. The study discovered that a multi-granularity sparse attention mechanism assisted in focusing on important relationships between these diverse nodes during the summarization process.

Another study proposed a novel Multiplex Graph Convolutional Network (M-GCN) architecture [46], which were effective at capturing various kinds of relationships between words and sentences. The model took into account both relationships within sentences (intra-sentential) and relationships between sentences (inter-sentential) to produce better summaries, particularly for long documents. Moreover, the work of [14] improved the current heterogeneous graph approaches by adding passage nodes alongside word and sentence nodes. The inclusion of passage nodes enhanced the network's ability to represent each sentence within the graph structure. This resulted in a more thorough comprehension of the document's structure, which ultimately assisted the creation of more accurate summaries.

Researchers of [47] employed the Text Graph Multi-Headed Attention Network (TGA) to effectively capture sentence representations across various types of text graphs at different levels. This method leveraged multi-headed attention to focus on relevant information within the graph structure. The study in [48] proposed a novel method based on hypergraph transformers. The method iteratively refined and developed strong representations of sentences by strategically incorporating various relationships between them. These relationships include underlying themes, key terms, references to the same entities, and document organization. The detailed methodology of our proposed model is discussed in the following section.

III. METHODOLOGY

This section illustrates the methodology of our proposed Heterogeneous Graph Neural Network framework, enriched

with topical information, for abstractive summarization as shown in Fig. 1. The proposed framework consists of four core trainable modules: a BERT Document Encoder, an LDA Topic Extractor, a Graph Attention Layer, and a Transformer Decoder. Given a corpus of scientific medical document with n documents, $D = [D_1, D_2, \dots, D_n]$, and the corresponding 'M' human generated gold summaries $Y = [Y_1, Y_2, \dots, Y_m]$, our model aims to create abstractive summaries of the documents (D) that capture the essential meaning of the documents using their own phrasing, achieving quality comparable to human-written summaries (Y). The proposed framework utilizes a multi-step process to generate abstractive summaries for scientific medical documents. First, the BERT Document Encoder processes each document to generate contextualized embeddings for each sentence, capturing its semantic meaning. These embeddings become the foundation for sentence and word nodes within the document graph. Next, the LDA model identifies a set of latent topics that represent the underlying thematic structure of each document. These topics are also incorporated as nodes in the document graph. A heterogeneous document graph is then constructed, where nodes represent sentences, words, and topics and the edges capture the relationships between these nodes, and are represented by TF-IDF weights.

The Graph Attention Layer then operates on heterogeneous document graph. It Utilizes TF-IDF weights on the edges to emphasize connections between important words and sentences within the document. It leverages an attention mechanism to selectively focus on crucial relationships within the graph, refining the representation of each node. Finally, the refined document graph is fed into a Transformer decoder. The decoder utilizes these refined node representations to generate an abstractive summary that captures the essential meaning of the document using its own phrasing. The details behind each module are discussed below as follows.

A. DOCUMENT ENCODER

BERT's encoder leverages an attention-based architecture, offering significant advantages compared to standard language models. Notably, it utilizes deep bidirectional training, considering both the left and right context of each word during training. This enhances its ability to understand the sequence effectively. However, the original BERT model aligns output vectors with individual tokens, not entire sentences [20]. For extractive summarization, the focus shifts to manipulating sentence-level representations. While the original BERT uses segmentation embeddings for sentence pairs, our task requires handling multi-sentence inputs [29]. Therefore, this study employs a modified BERT version that utilizes an external [CLS] token at the beginning and a [SEP] token at the end of each sentence. This systematic approach ensures consistent representation of individual sentences, similar to the method used in [12, 17, 29].

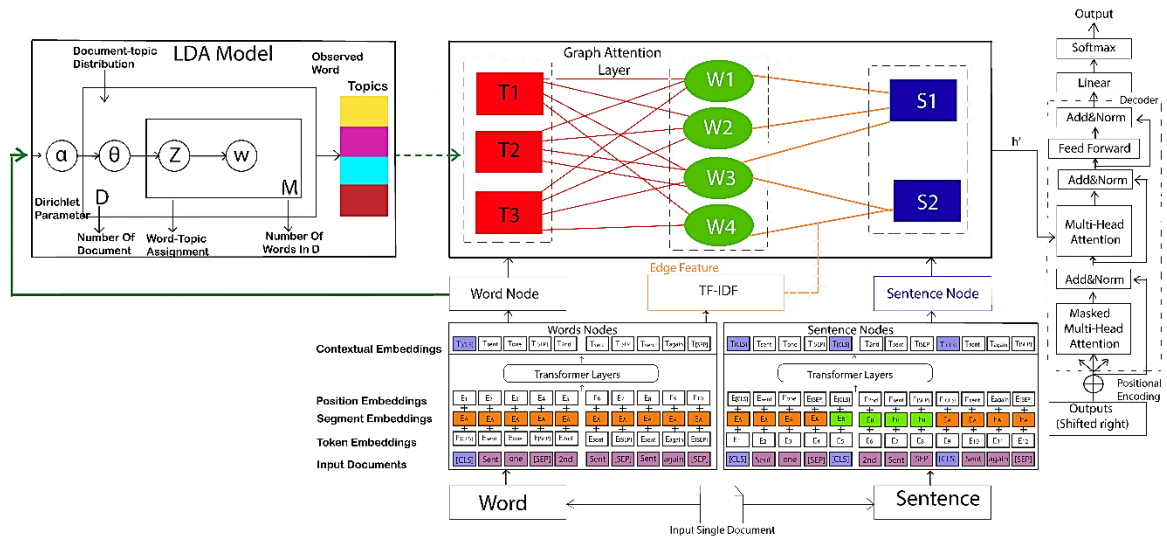


FIGURE 1. Overall framework of the proposed model

B. LDA TOPIC EXTRACTOR

A key aspect of our approach involves enriching the document representation by integrating latent topic information within the graph structure. This approach deviates from methods that rely solely on neural networks or clustering techniques like K-means and GMM [12, 15]. Instead, we leverage the Latent Dirichlet Allocation (LDA) model to extract initial topic features from each document, utilizing pre-trained word embeddings. LDA is a well-established unsupervised learning technique for analyzing document collections. It operates under two core assumptions: (1) Each document can be understood as a mixture of various hidden topics, with each topic having a specific probability of being present, (2) These hidden topics, across all documents, follow a common underlying thematic structure. This structure influences the types of words that are likely to appear within each topic [49]. By incorporating these topic features into the graph structure, we aim to capture a more comprehensive understanding of the document's content beyond just its individual words and their relationships.

While Latent Dirichlet Allocation (LDA) traditionally works with bag-of-words representations, our approach utilizes word embeddings. To bridge this gap, we created a Document-Term Matrix (DTM). This matrix represents documents using their word embeddings, with rows corresponding to documents and columns corresponding to the dimensions of the BERT embeddings. The complete generative process by LDA for latent topics discovery is as follows:

For each topic 't' from 1 to T:

- Sample a word distribution β_t from a Dirichlet distribution: $\beta_t \sim \text{Dirichlet}(\eta)$.

- This step determines how words are distributed within each topic.

For each document 'd' from 1 to D:

- Sample a topic distribution Θ_d from a Dirichlet distribution: $\Theta_d \sim \text{Dirichlet}(\alpha)$.
- This step defines the distribution of topics within each document.

For each word 'n' in document (d):

- Sample a topic assignment $z_{d,n}$ from the topic distribution $\Theta_d : z_{d,n} \sim \text{Multinomial}(\Theta_d)$.
- Sample a word $w_{d,n}$ from the word distribution corresponding to the assigned topic $\beta_{z_{d,n}} : w_{d,n} \sim \text{Multinomial}(\beta_{z_{d,n}})$.
- This step represents how words are generated within documents based on their topic assignments.

The probability of the observed data D is computed for a corpus using the following Equation 1:

$$p(D | \alpha, \beta) = \prod_{d=1}^D \int p(\theta_d | \alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{d,n}} P(z_{d,n} | \theta_d) p(w_{d,n} | z_{d,n}, \beta) \right) d\theta_d \quad (1)$$

Where D represents the total number of documents within the collection, w signifies the size of the vocabulary, which is the total number of unique words encountered, T stands for the number of latent topics identified by LDA, N_d indicates the number of words within a specific document (d). Θ_d represents the topic distribution for document d. This captures the likelihood of each topic existing within

that document, β_t signifies the word distribution for topic t . It reflects the probability of each word appearing within that particular topic, $z_{d,n}$ represents the topic assigned to the n th word in document d , and $w_{d,n}$ denotes the specific word found at position ' n ' within document d . Within this generative model, we estimated three crucial parameters:

Alpha (α), the document-topic distribution, captures the prior probability of each topic being present in a document.

Eta (η), the topic-word distribution, captures the prior probability of each word appearing in a topic.

Topic assignments $z_{d,n}$ represent the most likely topic assignments for each word within each document.

The ultimate goal is to discover the most probable topics and their corresponding word assignments based on the content observed within the documents. Although BERT is very effective at capturing the meaning of individual sentences and words. To further enrich the overall understanding of the document, we integrate topic distribution information alongside the sentence and word nodes within the graph structure. Additionally, we leverage TF-IDF weights on the edges of the graph. These weights emphasize the importance of connections between the various nodes, leading to a more refined grasp of the document's internal relationships.

C. HETEROGENOUS GRAPH LAYER

The vectors of nodes are initialized with embedding features, where $H_s^0 = X_s$ (sentence nodes), $H_w^0 = X_w$ (word nodes), and $H_t^0 = X_t$ (topic nodes), respectively. Sequentially, the node representations are updated with the graph attention network.

1) GRAPH ATTENTION NETWORK

Considering the heterogeneous graph architecture and the initial attributes of each node, we employ GAT to determine the hidden states of these nodes. Specifically, let $h_i \in R^{d_h}$ represent the input hidden representation of the i th node and N_i denotes its neighbors. Equations (2-4) shows the computation of the graph attention layer [4]:

$$z_{ij} = \text{Leaky ReLU}(W_a[W_q h_i; W_k h_j; e_{ij}]) \quad (2)$$

$$\alpha_{ij} = \frac{\exp^{z_{ij}}}{\sum_{l \in N_i} \exp^{z_{il}}} \quad (3)$$

$$u_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} W_v h_j \right) \quad (4)$$

Here, e_{ij} are the scalar edge weights W_a, W_q, W_k, W_v and vector α represent trainable parameters that are optimized throughout the training process. α_{ij} is the attention weight between h_i and h_j . Alternatively, to enhance performance, multi-head attention can be utilized, and it is computed as shown in (5).

$$u_i = \parallel_{k=1}^K \left(\sum_{j \in N_i} \alpha_{ij}^k W^k h_j \right) \quad (5)$$

The symbol \parallel is used to represent the concatenation operator. σ represents the non-linear transformation function, while h_i signifies the hidden state, which encapsulates information gathered from the neighboring nodes.

In addition, to tackle the challenge of gradient vanishing, a residual connection is incorporated. This connection is appended to the original representation, culminating in the derivation of the ultimate hidden state.

$$h'_i = u_i + h_i \quad (6)$$

2) GRAPH UPDATION

After initialization, the sentence nodes are updated by incorporating information from their neighboring word nodes. The topic nodes are also updated through the use of GAT and FFN layers as shown in (7).

$$\begin{aligned} U_{T \leftarrow w}^1 &= \text{GAT}(H_T^0, H_w^0, H_w^0) \\ H_T^1 &= \text{FFN}(U_{T \leftarrow w}^1 + H_T^0) \end{aligned} \quad (7)$$

Here in above equation H_T^0, H_w^0, H_s^0 are the node features of the topic, word, and sentence respectively. During each iteration, the update process [15] includes interactions between word-to-sentence, sentence-to-word, and word-to-topic connections. This can be formulated as shown in (8):

$$\begin{aligned} U_{w \leftarrow s}^{t+1} &= \text{GAT}(H_w^t, H_s^t, H_s^t) \\ U_{w \leftarrow T}^{t+1} &= \text{GAT}(H_w^t, H_T^t, H_T^t) \\ U_{w \leftarrow s, T}^{t+1} &= \sigma(U_{w \leftarrow s}^{t+1} + U_{w \leftarrow T}^{t+1}) \\ H_w^{t+1} &= \text{FFN}(U_{w \leftarrow s, T}^{t+1} + H_w^t) \\ U_{s \leftarrow w}^{t+1} &= \text{GAT}(H_s^t, H_w^{t+1}, H_w^{t+1}) \\ H_s^{t+1} &= \text{FFN}(U_{s \leftarrow w}^{t+1} + H_s^t) \\ U_{T \leftarrow w}^{t+1}, A_{T \leftarrow w}^{t+1} &= \text{GAT}(H_T^t, H_w^{t+1}, H_w^{t+1}) \\ H_T^{t+1} &= \text{FFN}(U_{T \leftarrow w}^{t+1} + H_T^t) \end{aligned} \quad (8)$$

Here in above equation, $A_{T \leftarrow w}$ signifies the attention matrix originating from word nodes towards the topic nodes. Next, a topic representation for the document is created, which is achieved by combining all the topic-related features extracted through the GAT layer. Equation (9) explains this combination process [15]:

$$\alpha_i = \frac{\sum_{n=1}^{N_d} c(w_n) \times A_{i,n}}{\sum_{j=1}^k \sum_{n=1}^{N_d} c(w_n) \times A_{j,n}}$$

$$H_{T_d} = \sum_{i=1}^K \alpha_i \times H_{T_i} \quad (9)$$

In the above equation (9), $A_{i,n}$ represents the information that the word n contributes to the topic i . $c(w_n)$ signifies the frequency of the word in the document, K stands for the number of topics, and α_i denotes the dominance level of topic i about the overall document topic. In addition, the hidden state of each sentence is carefully combined with the aforementioned topic vector. This produces a new representation that captures both the meaning of the sentence and its thematic relevance. This process is formulated as given in (10):

$$H'_{s_i, T_d} = F F N(H_{T_d}) \oplus H_{s_i} \quad (10)$$

D. DECODER

Our framework employed a state-of-the-art Transformer decoder. This decoder has its own attention mechanism, which allow it to focus on important parts of the input. The Transformer is recognized for its ability to analyze connections within the text (self-attention) and between the input text and the desired summary (cross-attention). Our model leveraged this capability to effectively collect salient information from the input text. It selectively focuses on the most informative sections and intelligently extracts and combines necessary details to generate clear and informative summaries.

Our work used a 6-layer Transformer decoder similar to the research in [3]. Each layer utilizes a multi-head attention mechanism and a feed-forward neural network, following standard Transformer practices. Each layer has 768 hidden states, resulting in a total of $768 \times 6 = 4608$ hidden states across all layers. The final output of the heterogeneous graph layer H'_{s_i, T_d} are the core representations that capture the overall meaning and underlying topical information. This information is then fed into the decoder for further processing and summary generation. In our study, the transformer decoder employed the Cross-Entropy Loss function. The loss function serves as primary objective of the model and it measures the difference between the predicted probability distribution of the generated summary tokens and the true distribution of actual summary tokens. By minimizing the Cross-Entropy Loss, the model assigns higher probabilities to the correct tokens in the summary. This, in turn, enhances its capability to generate accurate and coherent summaries.

IV. PERFORMANCE EVALUATION

This section presents a systematic evaluation of our proposed model. We compare its performance against state-of-the-art extractive and abstractive summarization models,

including neural graph-based models that utilize various nodes. We also provide details of the long scientific medical documents dataset used in this research, along with the hyperparameters, implementation settings, and the impact of latent topics (global information) on the system's summaries.

A. DATASET

While news article summarization is well-studied, summarizing significantly longer scientific papers poses a greater challenge due to the difficulty of accurately encoding lengthy text for effective summaries [7, 12]. To address this, we focus on the PubMed dataset of long scientific medical documents, introduced by [50]. The statistics of the PubMed dataset are illustrated in Table I.

TABLE I
EVALUATED DATASET STATISTICS

DATASET	Documents			Avg. Tokens	
	Train	Test	Val	Doc.	Sum
PubMed	119,924	6633	6658	3016	203

B. PERFORMANCE EVALUATION TOOLS

This study utilizes the ROUGE metrics to assess the quality of our machine-generated summaries. These metrics function by comparing machine-generated summaries to those written by human experts. This work used ROUGE-1(R-1), ROUGE-2(R-2), and ROUGE-L(R-3) assess different aspects of similarity. R-1 specifically evaluates the unigram recall, which measures the proportion of individual words (unigrams) that appear in both the machine-generated summary and the corresponding human-written abstract. R-2, on the other hand, focuses on bigram recall, assessing the number of two-word phrases (bigrams) shared between the summaries. Finally, R-L identifies the longest sequence of words that appear in the same order in both summaries. A higher ROUGE score [3] indicates a closer resemblance between the machine-generated and human-written summaries, suggesting a more effective summarization system. ROUGE-N is computed as given in (11).

$$ROUGE-N = \frac{Count_{mach}(ref, pred)}{Count(ref)} \quad (11)$$

C. IMPLEMENTATION SETTINGS

Our end-to-end trainable model leverages a base BERT model with a vocabulary size of 30,000 words to extract semantic features from text. BERT generates 768-dimensional word embeddings for each word. To improve computational efficiency and potentially capture task-relevant information, these embeddings are further processed and reduced to a lower dimension (60) using a dimensionality reduction technique. The model is trained for 50 epochs using the Adam optimizer with a batch size of 32. The node feature size within a single Graph Attention Network (GAT) layer is 60. The Transformer

decoder utilizes 6 multi-head attention heads. Dropout regularization (0.1 probability) is applied across layers. The model is trained on sentences with up to 100 tokens and documents with up to 50 sentences. The learning rate is set to 0.0001, and a maximum gradient normalization threshold of 1 is used.

To address the computational demands of model, we trained the model on Google Colab Pro+ leveraging its powerful GPUs. This provided significant processing power, which was essential for training. The training process itself required 15 GB of GPU memory, 6 GB of RAM, and 15 GB of hard drive space. The model training time was 72 hours.

D. OVERALL PERFORMANCE OF THE MODEL

This section evaluates the proposed model's accuracy and effectiveness in generating abstractive summaries for a medical scientific dataset. We compare our model against recent benchmark models categorized into five approaches:

1) TRADITIONAL EXTRACTIVE SUMMARIZATION

This category includes models like SumBasic [51], LexRank [52], LSA [53], and Oracle [8] that focus on extracting key sentences to create summaries.

2) NEURAL ABSTRACTIVE SUMMARIZATION

This approach utilizes neural networks to generate entirely new summaries that capture the document's essence. Examples include Attn-Seq2Seq [54], Pntr-Gen-Seq2Seq [25], and Discourse Aware network [50].

3) NEURAL EXTRACTIVE SUMMARIZATION

Similar to traditional extractive methods, these models leverage neural networks but still focus on selecting important sentences. Some examples are Cheng&Lapata [55], SummaRuNNer [5], NeuSum [6], and Xiao&Carenini [8].

4) PRE-TRAINED BASED MODELS

Match-Sum [56] falls into this category. It leverages a pre-trained BERT model for summarization, making it a sophisticated BERT-based approach.

5) GRAPH BASED MODELS

This category includes models like Topic-GraphSum [12], SSN-DM [57] HeterGraphLongSum [14] and GTASum [3] that utilize graph structures to represent document information and relationships between sentences or topics, facilitating summarization.

Table II in this study shows the Rouge F-scores for various models. The first segment of Table II demonstrates results for traditional summarization models and Oracle. The second segment presents the results of neural abstractive summarization models. The third segment reveal the results of Match-Sum, a sophisticated BERT-based summarizing model. The fourth segment illustrates the results achieved by recent advanced graph-based models for abstractive summarization. The last segment of Table II presents the results of our proposed model.

Our model significantly outperformed conventional extractive summarization models on all ROUGE metrics, as shown in Table II. This achievement extends to state-of-the-art neural abstractive and extractive models, where our approach also achieved higher ROUGE scores. This improvement highlights the benefit of integrating global semantic information with a dedicated graph layer. This layer facilitates the model's ability to generate summaries by considering the broader context within the document.

Additionally, traditional sequence-to-sequence (seq2seq) models with attention and pointer networks often struggle with lengthy scientific documents. This limitation arises from the challenges encoders face in capturing long-range dependencies within long texts [57]. Our model overcomes this limitation. Furthermore, our model surpasses the advanced BERT-based Match-Sum model, which experiences performance drops on the PubMed scientific dataset. Match-Sum's difficulty lies in grasping semantic and global information, hindering its ability to interpret the meaning of sentences and summaries. In contrast, our model can learn semantic information and leverage latent topics to focus on salient in long documents.

Our model's performance is compared with highly advanced neural graph-based models for both extractive and abstractive summarization, which utilize rich semantic information. Our model achieved near-identical results in R-1, comparable results in R-2, and surpassed Topic-GraphSum in R-L. This demonstrates the effectiveness of our model, which combined a GAT layer with a Transformer model, for abstractive summarization task. Topic-GraphSum, a state-of-the-art model for extractive summarization of long scientific documents, employed an NTM model for topic modeling along with BERT and GAT networks. GraphSum utilizing an NTM model contributes to its higher R-1 score compared to our proposed model. This is because NTM can be jointly optimized with the document encoder and graph networks. However, NTM training configurations are more complex [15], and aligning it with a graph neural network is more challenging compared to the simpler and easier-to-train LDA model. It is important to note that extractive summarization focuses on selecting the most important sentences from a document, which often leads to higher ROUGE scores compared to abstractive methods [50].

Our analysis also compares our model with two other advanced models: HeterGraphLongSum and GTASum. HeterGraphLongSum: This graph-based model focuses on extracting summaries from long scientific papers. It leverages three semantic units – words, sentences, and passage nodes – within its graph structure. Notably, the inclusion of passage nodes as high-level semantic units contributes to HeterGraphLongSum's stronger performance on ROUGE scores compared to our proposed abstractive model. Another reason for this performance difference is that extractive summarization, by design, selects the most important sentences from a document, which often leads to higher ROUGE scores compared to abstractive methods.

GTASum: This Graph-Based Topic-Aware abstractive summarization model is a direct competitor. GTASum employs a combination of techniques, including a BERT encoder, NTM for topic modeling, GAT, and a Transformer decoder. Our model surpassed GTASum in R-1 score, while achieving comparable results in R-2 and R-L scores.

Our model differs from GTASum in its use of semantic units within the graph structure. GTASum relies solely on sentence and topic nodes, whereas our model incorporates three types of nodes: sentences, words, and topics. Additionally, we leverage TF-IDF values of the entire document as edge features within the graph for richer information representation.

TABLE II

ROUGE F1 SCORES/RESULTS ON THE PUBMED DATASET SET. RESULTS WITH * ARE TAKEN FROM [50], RESULTS WITH + ARE TAKEN FROM [8], WHILE REMAINING RESULTS FROM THEIR ORIGINAL PAPERS.

<i>Traditional Extractive Models</i>	<i>R-1</i>	<i>R-2</i>	<i>R-L</i>
SumBasic*	37.15	11.36	33.43
LexRank*	39.19	13.89	34.59
LSA*	33.89	9.93	29.70
Oracle+	55.05	27.48	38.66
<i>Neural Abstractive Models</i>			
Attn-Seq2Seq*	31.55	8.52	25.56
Pntr-Gen-Seq2Seq*	35.86	10.22	25.16
Discourse Aware Network*	38.93	15.37	31.80
<i>Neural Extractive Models</i>			
Cheng&Lapata+	43.89	18.53	30.17
SummaRuNNer+	43.89	18.78	30.36
Xiao&Carenini	44.85	19.7	31.43
<i>Pre-Trained Models</i>			
Match-Sum	41.21	14.91	36.75
<i>Graph Based Models</i>			
Topic-GraphSum	46.13	20.91	33.27
SSN-DM	46.73	21.00	34.10
HeterGraphLongSum	48.75	22.45	43.97
GTASum	44.46	21.32	39.84
<i>Proposed Model</i>			
	46.03	21.42	39.71

Another key distinction lies in the topic modeling technique. We employed the simpler and more efficient LDA model compared to GTASum's NTM model, which requires complex configuration and training.

In summary, our model consistently delivers competitive results compared to both extractive and abstractive summarization models. It strikes a balance between sophisticated architecture, resource efficiency, and ease of implementation, ensuring effective performance across various scenarios. The model may be further improved by incorporating additional high-level semantic units as global information within the graph structure. Additionally, modifications to the encoder might be necessary to address the limitation of sentence token length for handling very long documents.

E. ABLATION WORK

To understand how each component/part affects our model's performance on summaries of scientific papers

from the PubMed dataset, we conducted an ablation study on the PubMed dataset. We evaluated three configurations:

1) FULL MODEL

This includes all components (word/sentence nodes, LDA, GAT).

2) WITHOUT TOPIC NODES (LDA)

This removes topic nodes from the graph, resulting in a structure with only word and sentence nodes.

2) WITHOUT GAT

This removes the Graph Attention Network (GAT) module. The comprehensive topic vector is then merged with sentence representations before feeding them into the decoder.

Table III depicts the performance of these configurations on the PubMed dataset.

Our findings are as follows:

- The full model achieved superior performance as compared to both ablated configurations, which illustrates the importance of each component and their joint effect for optimal model's performance.
- Removing topic nodes drastically reduced performance, which emphasizes the importance of latent topic information for model effectiveness.
- Removing GAT also led to a significant performance drop, which highlights the vital role of inter-sentence relationships in summarizing long documents.

TABLE III

IMPACT OF DIFFERENT COMPONENTS ON OVERALL MODEL

PubMed Dataset	Rouge 1	Rouge 2	Rouge L
Full Model	46.03	21.42	39.71
W/o Topic Nodes	43.12	19.42	38.01
W/o GAT	41.35	17.86	36.61

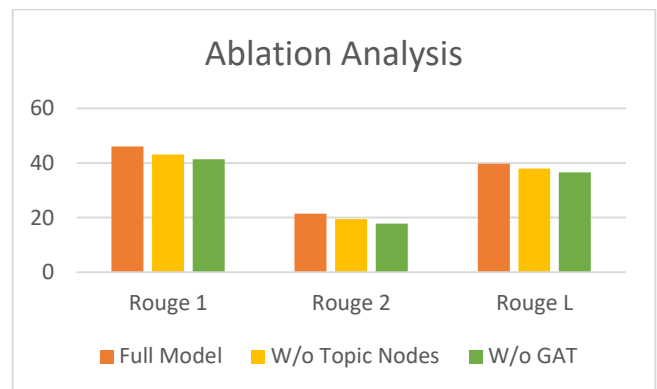


FIGURE 2. ROUGE SCORES OF OUR FULL MODEL AND TWO ABLATED VARIANTS

V. CONCLUSION AND FUTURE DIRECTIONS

This study proposed a novel Topic-aware Graph Neural Abstractive Summarization model particularly developed for lengthy scientific medical texts. Our model surpassed sentence-level neural graphs by considering additional semantic elements like words and latent topic nodes. This supplements the graph structure, leading a deeper understanding of the text. The model utilizes a powerful technique called BERT to encode the entire document. This exhaustive understanding of the text enables the model to better grasp the overall semantics and relationships between concepts.

LDA is also employed to identify hidden topics within the text and recognizes its underlying themes. This thematic comprehension further improved the model's ability to generate summaries that precisely capture the gist of the document.

In addition, a Heterogeneous Graph Neural Network is integrated into the framework to handle the complexity of scientific medical text by capturing meaningful connections between words, sentences, and latent topics within the document. The network can effectively model the complex connections within the text using the diverse node types. Finally, a Transformer decoder is utilized to ensure that the generated summaries are accurate, clear, and closely reflect the original text. Moreover, the decoder produced high-quality summaries using the comprehensive understanding and rich relationships captured in the previous stages.

We evaluated our model against various methods using the publicly available PubMed dataset of medical research papers. The results revealed that our approach outperformed most traditional models and achieved performance closer to the leading methods. For future research, we propose several directions to further improve our model: We will investigate the integration of more complex semantic units into the model to enhance its performance and robustness. We plan to explore advanced topic modeling techniques that are effective even in resource-limited environments, aiming to maintain simplicity without sacrificing performance. Additionally, we will examine the potential of advanced, sophisticated decoder components that could more effectively synergize with other neural components.

ACKNOWLEDGMENT

This research is funded by the Malaysian Institute of Information Technology, Universiti Kuala Lumpur, Malaysia.

CONFLICT OF INTEREST

The Author declares no conflict of interest.

REFERENCES

- [1] A. P. Widyassari, S. Rustad, G. F. Shidik, E. Noersongko, A. Syukur, and A. Affandy, "Review of automatic text summarization techniques & methods," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 4, pp. 1029-1046, 2022.
- [2] J. Pilault, R. Li, S. Subramanian, and C. Pal, "On extractive and abstractive neural document summarization with transformer language models," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 9308-9319.
- [3] M. Jiang, Y. Zou, J. Xu, and M. Zhang, "GATSum: Graph-based topic-aware abstract text summarization," *Information Technology and Control*, vol. 51, no. 2, pp. 345-355, 2022.
- [4] D. Wang, P. Liu, Y. Zheng, X. Qiu, and X. Huang, "Heterogeneous graph neural networks for extractive document summarization," *arXiv preprint arXiv:2004.12393*, 2020.
- [5] 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*, 2017, vol. 31, no. 1.
- [6] Q. Zhou, N. Yang, F. Wei, S. Huang, M. Zhou, and T. Zhao, "Neural document summarization by jointly learning to score and select sentences," *arXiv preprint arXiv:1807.02305*, 2018.
- [7] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [8] W. Xiao and G. Carenini, "Extractive summarization of long documents by combining global and local context," *arXiv preprint arXiv:1909.08089*, 2019.
- [9] L. Wang, J. Yao, Y. Tao, L. Zhong, W. Liu, and Q. Du, "A reinforced topic-aware convolutional sequence-to-sequence model for abstractive text summarization," *arXiv preprint arXiv:1805.03616*, 2018.
- [10] C. Zheng, K. Zhang, H. J. Wang, L. Fan, and Z. Wang, "Topic-guided abstractive text summarization: a joint learning approach," *arXiv preprint arXiv:2010.10323*, 2020.
- [11] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.
- [12] P. Cui, L. Hu, and Y. Liu, "Enhancing extractive text summarization with topic-aware graph neural networks," *arXiv preprint arXiv:2010.06253*, 2020.
- [13] H. Jin, T. Wang, and X. Wan, "Multi-granularity interaction network for extractive and abstractive multi-document summarization," in *Proceedings of the 58th annual meeting of the association for computational linguistics*, 2020, pp. 6244-6254.
- [14] T.-A. Phan, N.-D. N. Nguyen, and K.-H. N. Bui, "HeterGraphLongSum: Heterogeneous Graph Neural Network with Passage Aggregation for Extractive Long Document Summarization," in *Proceedings of the 29th International Conference on Computational Linguistics*, 2022, pp. 6248-6258.
- [15] T.-A. Phan, N. D. Nguyen, and K.-H. N. Bui, "Extractive Text Summarization with Latent Topics using Heterogeneous Graph Neural Network," in *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, 2022, pp. 749-756.
- [16] Z. Song and I. King, "Hierarchical Heterogeneous Graph Attention Network for Syntax-Aware Summarization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, vol. 36, no. 10, pp. 11340-11348.
- [17] M. Umair, I. Alam, A. Khan, I. Khan, N. Ullah, and M. Y. Momand, "N-GPETS: Neural Attention Graph-Based Pretrained Statistical Model for Extractive Text Summarization," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [18] T. Nguyen, A. T. Luu, T. Lu, and T. Quan, "Enriching and controlling global semantics for text summarization," *arXiv preprint arXiv:2109.10616*, 2021.

- [19] Y. Li *et al.*, "Topic-Aware Abstractive Summarization Based on Heterogeneous Graph Attention Networks for Chinese Complaint Reports," *Computers, Materials & Continua*, vol. 76, no. 3, 2023.
- [20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [21] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77-84, 2012.
- [22] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014.
- [23] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.
- [24] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
- [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] L. Dong *et al.*, "Unified language model pre-training for natural language understanding and generation," *Advances in neural information processing systems*, vol. 32, 2019.
- [27] S. Lamsiyah, A. E. Mahdaouy, S. E. A. Ouatik, and B. Espinasse, "Unsupervised extractive multi-document summarization method based on transfer learning from BERT multi-task fine-tuning," *Journal of Information Science*, vol. 49, no. 1, pp. 164-182, 2023.
- [28] Y. Liu, "Fine-tune BERT for extractive summarization," *arXiv preprint arXiv:1903.10318*, 2019.
- [29] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," *arXiv preprint arXiv:1908.08345*, 2019.
- [30] D. Miller, "Leveraging BERT for extractive text summarization on lectures," *arXiv preprint arXiv:1906.04165*, 2019.
- [31] X. Zhang, F. Wei, and M. Zhou, "HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization," *arXiv preprint arXiv:1905.06566*, 2019.
- [32] M. Zhong, P. Liu, D. Wang, X. Qiu, and X. Huang, "Searching for effective neural extractive summarization: What works and what's next," *arXiv preprint arXiv:1907.03491*, 2019.
- [33] M. Lewis *et al.*, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *arXiv preprint arXiv:1910.13461*, 2019.
- [34] J. Zhang, Y. Zhao, M. Saleh, and P. Liu, "Pegasus: Pre-training with extracted gap-sentences for abstractive summarization," in *International Conference on Machine Learning*, 2020: PMLR, pp. 11328-11339.
- [35] P. Fernandes, M. Allamanis, and M. Brockschmidt, "Structured neural summarization," *arXiv preprint arXiv:1811.01824*, 2018.
- [36] G. Mishra, N. Sethi, and Y.-C. Hu, "Intelligent Abstractive Text Summarization using Hybrid Word2Vec and Swin Transformer for Long Documents," *International Journal of Computer Information Systems & Industrial Management Applications*, vol. 15, 2023.
- [37] T. Chen, X. Wang, T. Yue, X. Bai, C. X. Le, and W. Wang, "Enhancing Abstractive Summarization with Extracted Knowledge Graphs and Multi-Source Transformers," *Applied Sciences*, vol. 13, no. 13, p. 7753, 2023.
- [38] R. Srivastava, P. Singh, K. Rana, and V. Kumar, "A topic modeled unsupervised approach to single document extractive text summarization," *Knowledge-Based Systems*, vol. 246, p. 108636, 2022.
- [39] T. Ma, Q. Pan, H. Rong, Y. Qian, Y. Tian, and N. Al-Nabhan, "T-bertsum: Topic-aware text summarization based on bert," *IEEE Transactions on Computational Social Systems*, vol. 9, no. 3, pp. 879-890, 2021.
- [40] A. Joshi, E. Fidalgo, E. Alegre, and L. Fernández-Robles, "DeepSumm: Exploiting topic models and sequence to sequence networks for extractive text summarization," *Expert Systems with Applications*, vol. 211, p. 118442, 2023.
- [41] Q. Xie, J. Huang, T. Saha, and S. Ananiadou, "Gretel: Graph contrastive topic enhanced language model for long document extractive summarization," *arXiv preprint arXiv:2208.09982*, 2022.
- [42] H. Xu, Y. Wang, K. Han, B. Ma, J. Chen, and X. Li, "Selective attention encoders by syntactic graph convolutional networks for document summarization," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020: IEEE, pp. 8219-8223.
- [43] Y. J. Huang and S. Kurohashi, "Extractive summarization considering discourse and coreference relations based on heterogeneous graph," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 2021, pp. 3046-3052.
- [44] R. Jia, Y. Cao, H. Tang, F. Fang, C. Cao, and S. Wang, "Neural extractive summarization with hierarchical attentive heterogeneous graph network," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 3622-3631.
- [45] Y. Liu, J.-G. Zhang, Y. Wan, C. Xia, L. He, and P. S. Yu, "HETFORMER: heterogeneous transformer with sparse attention for long-text extractive summarization," *arXiv preprint arXiv:2110.06388*, 2021.
- [46] B. Jing, Z. You, T. Yang, W. Fan, and H. Tong, "Multiplex graph neural network for extractive text summarization," *arXiv preprint arXiv:2108.12870*, 2021.
- [47] T. Vo, "An approach of syntactical text graph representation learning for extractive summarization," *International Journal of Intelligent Robotics and Applications*, vol. 7, no. 1, pp. 190-204, 2023.
- [48] H. Zhang, X. Liu, and J. Zhang, "Hegel: Hypergraph transformer for long document summarization," *arXiv preprint arXiv:2210.04126*, 2022.
- [49] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimedia Tools and Applications*, vol. 78, pp. 15169-15211, 2019.
- [50] A. Cohan *et al.*, "A discourse-aware attention model for abstractive summarization of long documents," *arXiv preprint arXiv:1804.05685*, 2018.
- [51] L. Vanderwende, H. Suzuki, C. Brockett, and A. Nenkova, "Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion," *Information Processing & Management*, vol. 43, no. 6, pp. 1606-1618, 2007.
- [52] 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.
- [53] J. Steinberger and K. Jezek, "Using latent semantic analysis in text summarization and summary evaluation," *Proc. ISIM*, vol. 4, no. 93-100, p. 8, 2004.
- [54] R. Nallapati, B. Zhou, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence rnns and beyond," *arXiv preprint arXiv:1602.06023*, 2016.
- [55] J. Cheng and M. Lapata, "Neural summarization by extracting sentences and words," *arXiv preprint arXiv:1603.07252*, 2016.
- [56] M. Zhong, P. Liu, Y. Chen, D. Wang, X. Qiu, and X. Huang, "Extractive summarization as text matching," *arXiv preprint arXiv:2004.08795*, 2020.
- [57] P. Cui and L. Hu, "Sliding selector network with dynamic memory for extractive summarization of long documents," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, pp. 5881-5891.



Ayesha Khaliq received her Master degree in Computer Science from the University of Engineering and Technology Faisalabad. Currently she is doing Ph.D in Computer Science From University of Agriculture Faisalabad Punjab, Pakistan. she is serving as Lecturer in Government College University Faisalabad. Her research interests encompass sentiment analysis, text summarization, and applications of machine learning and Deep Learning.



Atif Khan received his M.Sc. degree in Computer Science from the University of Peshawar, Pakistan, in 2004, and his Ph.D. degree in Computer Science (Text Mining) from Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia, in 2016. Since 2022, he has been working as an Associate Professor at Islamia College Peshawar, KP, Pakistan. He actively participates as a technical committee member in numerous international conferences and serves as a reviewer for various international conferences and journals. His research interests encompass data mining, text mining, opinion mining, recommender systems, Internet of Things (IoT) and applications of machine learning and Deep Learning. With a focus on these areas, he has authored numerous research articles in reputable impact factor journals. During his Ph.D. tenure at UTM, he was honored with the Best Student Award and the Pro-Chancellor Award for his outstanding contributions to the field of text mining. Additionally, he holds the position of Associate Editor for ACM Transactions on Asian and Low-Resource Language Information Processing.



Salaman Afsar is serving as Assistant Professor in Department of Computer Science, University of Agriculture Faisalabad Punjab, Pakistan. He actively participates as a technical committee member in numerous international conferences. His research interests encompass machine learning, Deep Learning, Social Network Analysis, Computer Networks and Data Mining. With a focus on these areas, he has authored numerous research articles in reputable impact factor journals.



Muhammad Umair received his BS degree in Computer Science from the University of Peshawar, Pakistan, in 2018, and his MS. degree in Computer Science (Text Summarization) from the City University of Science and Information Technology (CUSIT), Peshawar, in 2022. Since 2023, he has been working as an Assistant Director I.T/Technical at

Establishment Department Civil Secretariat Peshawar, KP, Pakistan. He actively participates as a technical committee member / Focal Person in numerous Technical and Artificial Intelligence-Based software systems in the Government sector. His research interests encompass data mining, text mining, text summarization, and applications of machine learning, Deep Learning, and Large Language Models (LLMs). With a focus on these areas, he has authored numerous research articles in reputable impact factor journals.



SALMAN JAN received a master's degree in computer science from the University of Peshawar and a PhD degree from the MIIT, Universiti Kuala Lumpur, in 2019. His PhD thesis was on Android malware analysis and its integration with blockchain for preserving the integrity of the behavioral logs and ultimately to secure the classification results of the behavioral logs through the employed Deep Learning models including DCGAN, CNN, and FCNN.

He has acquired skills and published in a number of well-reputed journals in areas of machine learning, deep learning, artificial intelligence, blockchain, IoT, and security Augmented and Virtual Reality. He also is actively working on international projects in the capacity of consultant and successfully completed Four (4) international projects. He is presently serving as a Senior Lecturer of Cyber Security & Technological Conversion at the Malaysian Institute of Information Technology, University Kuala Lumpur, Malaysia. His publications, citations, and ORCID ID can be found at: <https://scholar.google.com/citations?user=4n2MC74AAAAJ&hl=en&oi=ao> and 0000-0002-8250-694X.



Megat F. Zuhairi is an Associate Professor at Universiti Kuala Lumpur, Malaysian Institute of Information Technology. He is currently a faculty member of the Informatics and Analytics Department at the institute. He received his PhD in Electronics and Electrical Engineering from the University of Strathclyde in 2012 and his M.S. in Communication Networks and Software from the University of Surrey, UK in 2002. He is an active researcher and a certified Cisco Network Academy Instructor. His research interests include computer networking, process mining, blockchain, and system performance and modeling.