Research article

The power of graphs in medicine: Introducing BioGraphSum for effective text summarization

Cengiz Hark

İnönü University, Department of Computer Engineering, 44000, Malatya, Turkey

A R T I C L E I N F O

A B S T R A C T

In biomedicine, the expansive scientific literature combined with the frequent use of abbreviations, acronyms, and symbols presents considerable challenges for text processing and summarization. The Unified Medical Language System (UMLS) has been a go-to for extracting concepts and determining correlations in these studies; hence, the BioGraphSum model introduced in this study aims to reduce this UMLS dependence. Through adoption of an innovative perspective, sentences within a piece of text are graphically conceptualized as nodes, enabling the concept of "Malatya centrality" to be leveraged. This approach focuses on pinpointing influential nodes on a graph and, by analogy, the most pertinent sentences within the text for summarization. In order to evaluate the performance of the BioGraphSum approach, a corpus was curated that consisted of 450 contemporary scientific research articles available on the PubMed database, aligned with proven research methodology. The BioGraphSum model was subjected to rigorous testing against this corpus in order to demonstrate its capabilities. Preliminary results, especially in the precision-based and f-score-based ROUGE-(1–2), ROUGE-L, and ROUGE-SU metrics reported significant improvements when compared to other existing models considered state-of-the-art in text summarization.

1. Introduction

Information on biomedicine and healthcare has been expanding at an exponential rate. However, the enormous volume and diversity of textual data in this field can result in information overload for those seeking to digest this information, which also reduces their productivity and prevents them from learning new things properly. The increased cognitive demands and mental effort caused by this information overload may make it more difficult for healthcare professionals to adequately grasp their patients' medical problems and thereby affecting their ability to make the best medical judgments. Clinical researchers are required to expend a considerable amount of time and energy accessing the necessary and pertinent information from the vast amount of regularly updated biomedical data that is now available.

According to Ref. [1], the effects of information overload an even be lethal since the workflow of today's healthcare professionals now includes accessing textual data and interacting with computer-based technologies. Through automatic text summarization studies performed on biomedical data, this scenario of information overload can be significantly reduced, with document summary a crucial component of resolving this issue. Text summarization can aid clinical decision making and raise the standard of patient treatment by enabling healthcare professionals to concentrate on the more essential facts. However, despite recent improvements in document summarization, certain issues still require improved solutions. Automated document summarization seeks to develop techniques that

can identify the most crucial elements of information from a piece of text, and then condenses it in a way that is of benefit to readers. However, certain aspects must be obtained via data processing in order to reduce the time needed to acquire the desired information, hence the field of automatic text summarization systems has received increased attention as a result [2].

Automatic text summarization systems can be divided into extractive, abstractive, and hybrid approaches [3]. In extractive summarization systems, the most relevant and consistent statements are selected consecutively from the text being summarized, and a summary is then created [4], whereas abstractive summarization algorithms can change the document's actual content, word order, and semantic representation in order to express it using different words. In contrast to extractive summaries, abstractive summaries use sentences distinct from those found in the main text(s) [5]. The third group, hybrid document summarization, combines both inferential and abstractive methodologies in order to produce a summary [3]. Text summarization can also be classified as either single-document summarization (SDS), or multi-document summarization (MDS), depending on how many source papers were utilized to produce the final summary. SDS creates a summary from a single text document and aims to condense any given input text whilst preserving the most crucial information [6], whereas a multi-document summary is based on a set of input documents [4]. The summary document may fall under the generic or query-based categories, depending on the information it contains. In order to provide a general knowledge summary of a document's content, a generic text summarizer pulls the essential information together from one or more input texts. In query-based summarizing, a large corpus of uniform documents is culled based on a query and used as input to a multi-document summarizer, and the information about the query then subsequently produced in summary form. A general summary aims to offer a broad overview of a document's content, whereas a query-based summary provides the information that is considered most pertinent to the initial search query. Additionally, there are controlled and unsupervised methods for document summarization. By classifying the units in the text that need to be summarized according to whether or not they are included in the summary, the classification problem of choosing relevant phrases in supervised summarization can be demonstrated. Unsupervised algorithms frequently assign grades to textual sequences based on specified standards and attributes. However, whilst unsupervised techniques do not require the use of educational data, supervised approaches do. Promising outcomes can be gained by transferring learning using trained word-embeddings within a supervised learning framework.

The primary goal of this paper is to present a brand-new system aimed at successfully and efficiently selecting summary sentences from biomedical texts based on their familiarity with the key topics and subthemes of the text. BioGraphSum is a novel, graph-based model designed specifically for the purpose of biomedical text summarization. Its design utilizes an innovative approach that integrates Malatya centrality, a concept derived from graph theory, to prioritize the most influential sentences within a set of biomedical texts. In order to take links between different sequences into consideration, a source document is proposed as an undirected weighted graph. With the modeling steps, the details of which are provided in this paper, similarities between sentences in the text are transferred to the graph according to maximum representation. After having transferred the source texts to the graph, an effective and simple multi-document generic unsupervised document summarization system is used to advance graph-based biomedical text summarization one step further by using the vertex cover algorithm. The vertex cover algorithm served as a source of inspiration for the current study since it represents only a portion of the vertices that can cover an entire network. In real world problems such as text summarization, as in vertex cover, the mainstay of sentence selection is common word groups. This approach determines the level at which a successful and comprehensive summary contains golden information and how often it contains repetitive concepts. As known, a comprehensive and high-quality summary should contain the maximum amount of information possible whilst keeping repetitive information to an absolute minimum.

Building upon the graphical expression of a text set, the current study aimed to produce summaries with a high level of coverage that can be accessed from the nodes subset of each edge in the vertex cover problem. Another innovation of the BioGraphSum model is the first use of Malatya centrality [7] in a text summarization problem since the algorithm offers a polynomial solution for the minimum vertex-cover problem. One of the essential advantages of this method is that in each iteration of the Malatya algorithm, the node carrying the most information can be selected in a polynomial time. Next, the inclusion of the selected node in the summary is removed from the diagram by deleting all edges belonging to the relevant node, and these erased edges represent common text groups. Thus, in the next step, it becomes possible to detect text groups from the main text that are not included in the abstract. Since common edges are deleted prior to new sequences being included in the summary in the following iteration, sequences with only minimal repetition are included in the summary. A quality and comprehensive summary should avoid word and concept repetitions, and this model allows for very comprehensive golden summaries to be created. The BioGraphSum model then selects and combines the most pertinent and instructive sentences from all of the clusters in order to construct the final summary, which covers all of the text's current subtopics. In many studies in the literature on biomedical summarization, UMLS ready-made normalization tools such as MetaMap and OGER have been used [8]. However, in the BioGraphSum model, the outputs from OGER and MetaMap-style tools are not utilized at any stage, meaning that this former dependency on ULMS has been removed.

The current study then tested the effectiveness of the BioGraphSum text summarization approach by generating biomedical summaries based on different dimensions using datasets from NLM's PubMed. The results, based on the Recall-Oriented Understudy for Gisting Assessment (ROUGE) [9] metric, show that the proposed BioGraphSum model achieved scores of 0.2942, 0.1031, 0.1829, and 0.07498 for ROUGE-1, ROUGE-2, ROUGE-S, and ROUGE-SU, respectively. Overall, superior results were reported compared to other competitive state-of-the-art methods. The outcomes presented in the pertinent tables and figures in this paper demonstrate the BioGraphSum model's superiority.

The primary contributions of the current study are as follows.

- Innovative approach: An unparalleled and effective method is introduced for summarizing unstructured documents in the realm of biomedical text summarization.

- Pioneering application: The utilization of Malatya centrality marks a first in addressing text summarization challenges, offering a polynomial-time solution to the minimum vertex-cover problem.
- Training-free model: The proposed BioGraphSum model operates without need for any training or test data, enhancing its applicability and versatility.
- Distinct independence: Despite the uniqueness of biomedical texts, the BioGraphSum model operates independently of domain-specific tools such as UMLS, underscoring the significance of eliminating dependencies on tools such as OGER and MetaMap.

The remaining sections of this paper are presented as follows: The essential work on text summarization is covered in Section 2, whilst the purpose and goals of the study are further explained in Section 3. Section 4 outlines the methodology and each step of the process before Section 5 discusses the corpus, evaluation metrics, and experimental findings of the study, including comparisons with existing cutting-edge techniques. Finally, Section 6 provides comprehensive summary information regarding the study's findings.

2. Related research

2.1. Extractive summarization

Extractive text summarization methods select and extract the sentences and words that are considered to have the most meaning and value from the raw text data, and are then used in order to create valid, comprehensive, and meaningful summaries. These summaries are formed using statistical and heuristic algorithms that review the sentences considered the most essential and valuable in the texts; an approach that can be achieved without first gaining a deep understanding or semantic analysis of the text [10–13].

Inferential text summarization techniques can be categorized into distinct classes based on the techniques employed. These techniques comprise TF-IDF approaches, as well as methods based on artificial neural networks, clustering, queries, machine learning, graphs, and those that utilize fuzzy logic. In fuzzy logic-based text summarization approaches, each feature of a text is used as input to then detect and extract essential sentences from the text. The methods used in fuzzy text summarization can vary according to differences in the feature extraction, fuzzy rules, linguistic variables, membership functions, blurring, and defuzzification methods used [14]. In the first stage of the current research, clustering algorithms were employed to group text phrases before using a teaching-learning-based optimization strategy in the second stage to determine the best weights for the text fragments. In order to compute the final score of the sequences, a fuzzy inference method was simultaneously combined with a comprehensive human-generated base [15].

The definition of TF-IDF-oriented techniques is the assessment of the textual relevance of each word in a string or dataset. In Ref. [15], an inferential artificial text summarization system was introduced for the Turkish language that combined the statistically-based TF-IDF and graph-based PageRank algorithms. In the clustering approach, another inferential text summarization approach calculates the similarity of sentences according to similarity criteria between sentences in the text [13]. Summaries produced by considering the scope of the summaries to be created are classified as query-based or general summaries. Summaries created in query-based summarization systems take into account essential words provided by the user [2]. For a Chinese query-based document summarizing challenge, the Query-Based document summarization (QBSUM) dataset was introduced, and several solutions were provided within the scope of [16]'s study. Summarization models based on machine learning mostly classify the sentences within the given text using popular machine-learning methods such as SVM, DT, LR, RF, and KNN [13]. It is also possible to use a specific variation of KNN (K Nearest Neighbor), in which the similarity between feature vectors is determined by taking into account the similarity between value features. In the current research, both attributes and attribute values were taken into account, and modified by adapting to KNN for use in the text summarization task [17].

Graph-based approaches are also commonly used in inferential text summarization systems. For example, in Ref. [2], graphs consisting of sentences were used to represent the document set, and an automated text summarization system was introduced based on entropy and graphs. In addition, in Ref. [8], a multi-layered graph modeling and text summarization study was conducted for the purpose of biomedical text summarization. Numerous productive studies with similar assumptions are also evident in the current literature [6,18–22].

Latent semantic analysis (LSA) techniques are unsupervised and therefore do not require external information or training data, which is similar to graph-based techniques. The primary goal of text summarization is to ascertain how phrases and words relate to one another in terms of context. In order to overcome the performance problems of previous LSA-based solutions, an innovative approach was introduced that combined the syntactic and semantic information of the text using statistical and algebraic approaches [23]. When the available training data is abundant, neural-based models perform better than conventional models and require less human intervention. However, in order to extract more significant features from phrases using deep learning models, a significant amount of data is first required in order to train the systems using pre-trained models [24]. As a result, very rich summaries can be obtained with approaches in which document summarization is handled as an optimization problem. Hark et al. [4] introduced the multi-objective saplings growing-up algorithm (MO-SGuA) for extractive and generic summaries. In their study, they discussed document summarization as a multi-objective optimization problem, and the reported results clearly demonstrated the success of their approach.

The current research proposes a graph-based, unsupervised, generic, multi-document system to summarize Biomedical texts.

3. Motivation and objectives

A finite set of sentences that encompasses all the concepts addressed by each and every sentence of the input texts was first selected
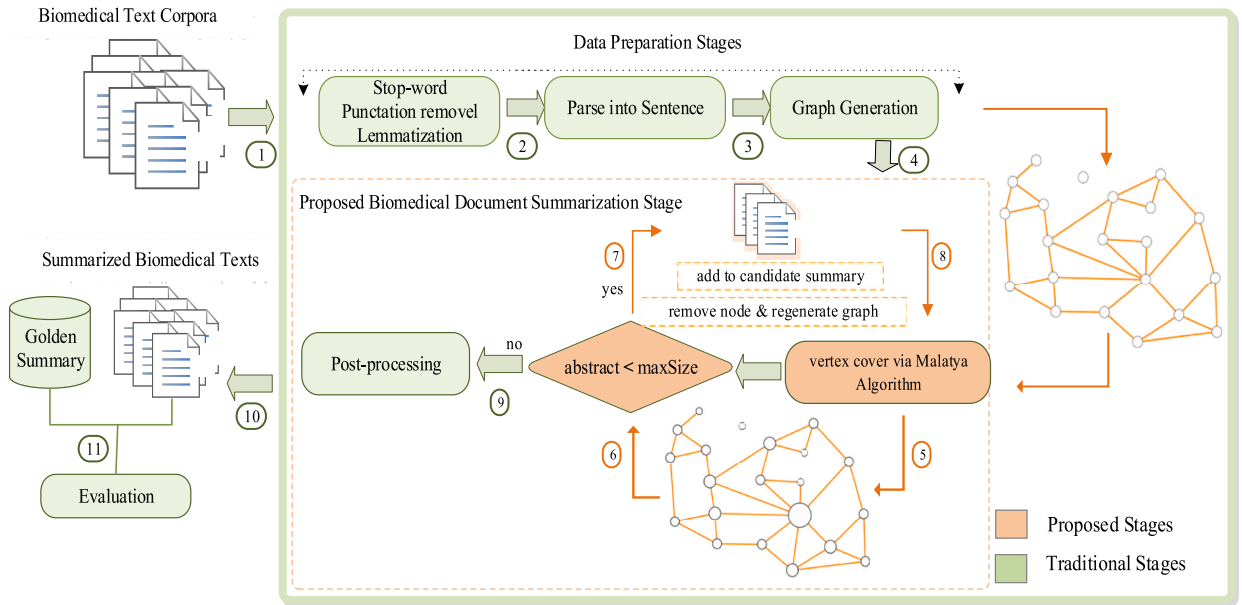
Fig. 1. High-level design of proposed BioGraphSum model.

according to the extractive multi-document summary procedure. Since more than one input document contributing to a summary may contain significant equivalent information, the summary drawn from the main text may contain a certain level of information already covered. Whilst it is an undesirable result to have such conflicts in a summary, the proposed multi-document text summarization system is able to overcome this problem. For the first time, within the scope of the current study, the Malatya algorithm, which is the subject of a real-world problem, provides complete content information and diversity. At the same time, information containing minimum repetitions is included in the summary due to application of the working principles of the Malatya algorithm. Despite having the moniker "vertex cover algorithm," it has served as a source of inspiration for the current study since it only represents a portion of the vertices that can cover the entire network. A finite vertex-cover set of a graph is a collection of chosen vertices such that each edge in the graph coincides with at least one of the chosen vertices, as is detailed in the relevant section. In this way, in order to obtain rich
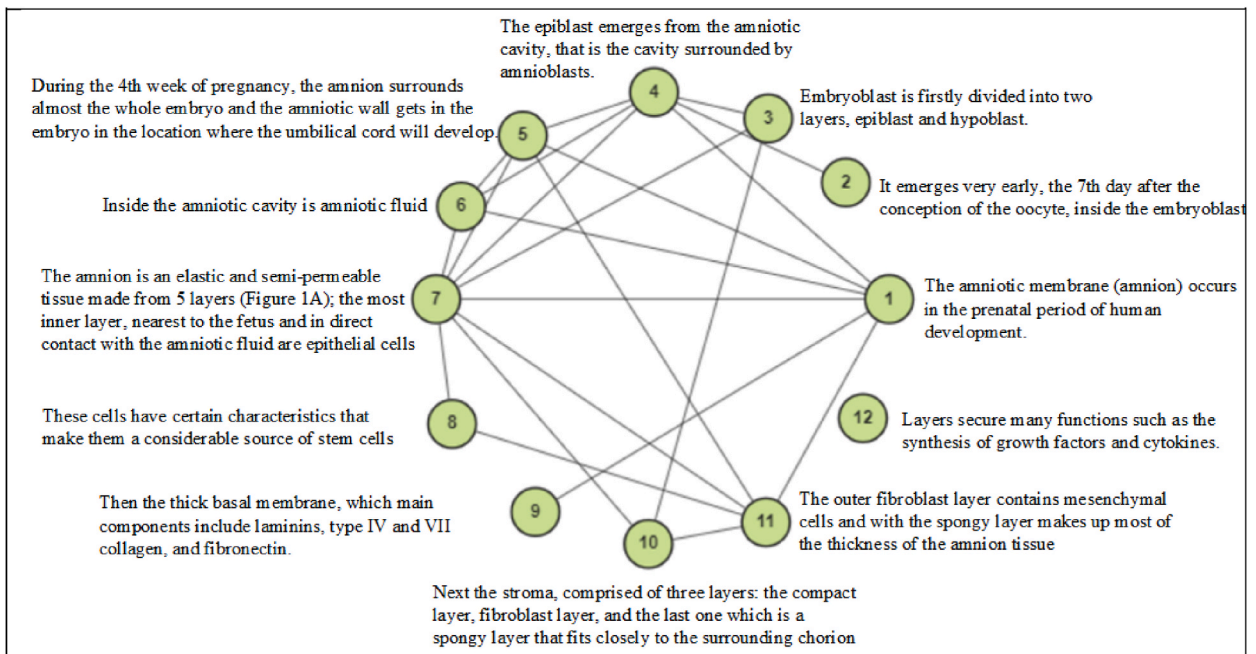

Fig. 2. Example biomedical content.

text summaries, the common opinion and content of the textual diagram are preserved while also achieving minimal repetition. The main source of inspiration for the experimental approaches used in the current study were conceptual parallels drawn between the vertex-cover problem and the multi-document summarization problem.

4. Proposed method

This study proposes an extractive, multi-documented, and generic text summarization system. The performance of this innovative method was tested against real published biomedical text data. Fig. 1 illustrates the proposed biomedical text summarization system which consists of four main stages. The first stage entails preprocessing and/or conversion of unstructured texts. Removing stop-words, lemmatization, and punctuation steps are vital in order that the documents may be used effectively in the subsequent steps. The second stage provides the mathematical and graphical transformations of each sentence that makes up the texts. At this stage, common word intersections between sentences are taken into account.

The third and main stage is where the Malatya algorithm, used for the first time in a text summarization task, is run repeatedly and a subset of the most effective nodes is created to cover all edges. This step is where rich summaries of a desired size are created. The final stage is where post-processing operations are performed such as arranging the selected sentences according to their order in the main document so as to create a more fluent summary. This paper also reports on the performance and efficiency of the proposed Bio-GraphSum model according to various recognized performance metrics.

4.1. Text preprocessing

Preprocessing steps are often necessary for datasets that lack a specific structure. For this, maximizing representation in the textual presentation step is the primary objective. The system's performance is anticipated to improve with the removal of extraneous information from the biomedical text documents used. Prior to processing the natural language text, pronouns, prepositions, and conjunctions, etc. are first removed since these terms do not represent sentences nor make sense on their own and are therefore eliminated from the original datasets. By processing just words with categorical significance, the workload of the whole process is significantly decreased. Next, lemmatization and stemming processes were performed on the texts, where the purpose is to convert words into common basic forms. Since the data used in the current study consisted of actual PubMed articles, this stage was critical because the published articles in the dataset may contain items with vague conceptual meanings [2,4,5,25]. The BioGraphSum text summarization system proposed in this study starts with a text preprocessing phase.

4.2. Textual graph modeling

Different facets of the subject being reported should be included in order to generate a comprehensive summary. The process aims to provide convenience for readers by concentrating the information at the maximum level. Therefore, it is first necessary to remove certain features in order to minimize the time spent accessing information. Representing the texts mathematically and with maximum
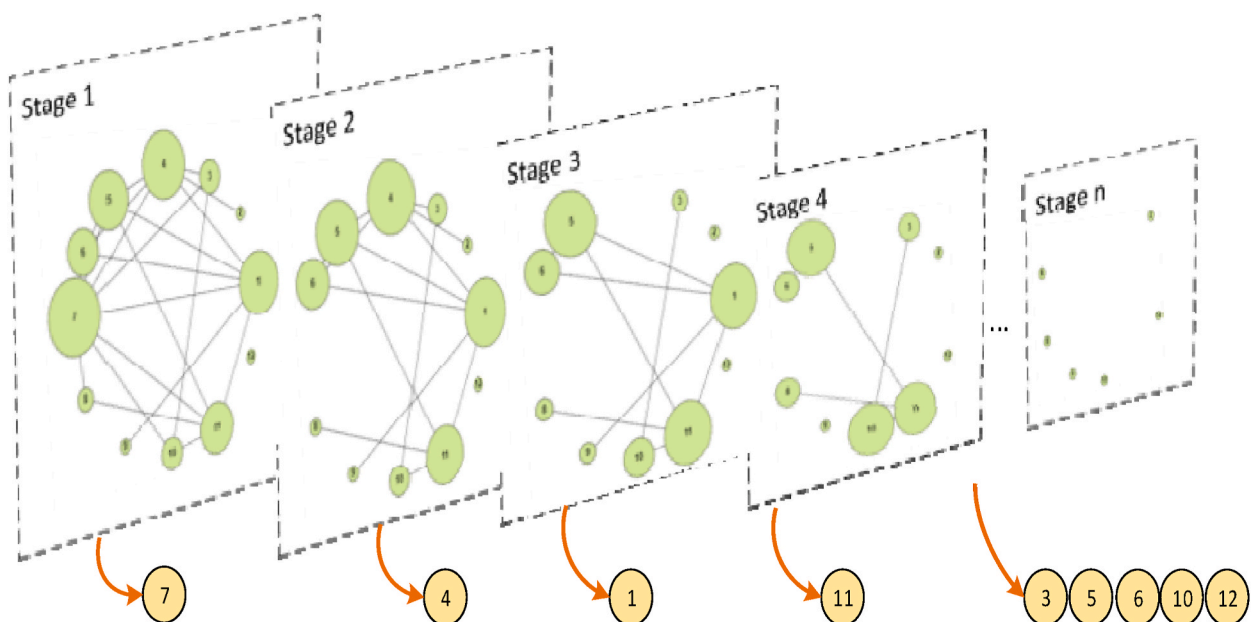

Fig. 3. Selection stages of nodes using the Malatya algorithm.

information is another critical step of text summarization. The textual diagrams to be modeled should be able to derive various relationships as a result. Within the scope of the current study, an effective textual diagram creation process was conducted so as to present the maximum level of representation in light of the analyses performed. The proposed BioGraphSum model creates graphs from input texts that represent various sorts of sentence relationships. The procedures listed in Algorithm 1 are then used to create the textual diagram.

The next step is to model a biomedical text as a weighted, undirected graph, where each text sequence is a node and the edges denote different sentence associations. This is phase 4 of the proposed BioGraphSum architecture. Fig. 2 represents an exemplary Biomedical manuscript [26] and its corresponding diagram.

Algorithm 1. CreateGraph

Input : nodeList
Output : Graph(V, E)

1. edgeList ←∅
2. For node in NodeList do
3.   V ← node
4.   For N in NodeList do
5.   For word in V.LemmatizedWords do
6.   If N.LemmatizedWords contains word
7. $E \leftarrow$ Create New Edge between V and E
8. Output = nodeList, edgeList

4.3. Vertex cover via Malatya algorithm

Many real-world problems can be successfully solved using graph theories, and the vertex cover problem is one such hypothesis. It is considered a strong hypothesis which searches for a subset of vertices that can represent an entire network since it is an optimization problem and optimal solutions are not easy to find. Different approaches have been used in the literature to address the vertex cover problem or to offer near-optimal solutions. One such approach recently introduced the Malatya algorithm [7], which offers practical solution sets to the NP-hard minimum vertex cover issue, and was selected for use in the current study. The Malatya algorithm's rough code is represented as shown in Algorithm 2.

Algorithm 2. Apply Vertex Cover

Input : G(V,E)
Output : Sentence

1. vc ←∅
2. While E ≠∅ do
3.   I ←1, …, |V|
4. $\psi(v_i) = \sum_{\forall v_j \in N(v_i)} \frac{d(v_i)}{d(v_j)}$
5. $V_c = V_c \cup \{max(\psi(v_i))\}$
6. V=V-{v$_i$}, and E=E- $\forall$(v$_i$,v$_j$) ∈ E
7. Output = V_c.Sentence

4.4. Proposed biomedical document summarization

The Malatya algorithm has not previously been used to resolve any real-world problem, hence the current study presents the first such approach. In this study, the algorithm's effectiveness was tested against a significantly important and challenging problem, that of biomedical text summarization.

While calculating the centrality values of the nodes in the diagram, the Malatya algorithm is not affected by local weights. Since the

degrees of all nodes adjacent to the relevant node are included in the calculation, this approach makes the Malatya algorithm significantly effective. All nodes are visited by taking into account both local and neighborhood weights. As a result, it may be possible to look at how inclusive the candidate summary's sentences are. At the end of the first step, the node with the highest node score is added to the minimum vertex cover subset. By deleting the node added to the subset and the edges connecting to this node from the diagram, the same procedure is performed. At the end of the specified steps, the minimum number of nodes with the maximum coverage of nodes is obtained using the Malaya algorithm. The steps of an example biomedical diagram using the Malatya algorithm are schematized as illustrated in Fig. 3, which is based on the example biomedical content presented in Fig. 2 (biomedical text sourced from the corpus used in the current study).

Fig. 3 illustrates the direct proportionality between the centrality value of Malatya and the node sizes. According to the Malatya algorithm, the more valuable nodes are pictured larger in order that the working steps of the algorithm can be more easily understood. For example, in Stage 1, the highest Malatya centrality node score belongs to Node 7. As shown in Fig. 2, Node 7 reads; *"The amnion is an elastic and semi-permeable tissue made from 5 layers (Fig. 1A); the most inner layer represents the sentence Nearest to the fetus and in direct contact with the amniotic fluid are epithelial cells."* Thus, the most comprehensive sentence worth including in the summary was determined. In the subsequent steps, this process is repeated based on the basic principles of the algorithm. This approach ensures that summaries of the desired size can be obtained within the scope of any given study, and dynamic subsets can be obtained according to the length of the summary generated.

Algorithm 3.   Proposed BioGraphSum Algorithm (Generate Summarized Text)

Input : Biomedical Document Corpora (Content, maxSize)
Output: Summarized Document

  1. nodeList ←∅
  2. cumleListe ← Get Sentence from Content
  3. For C in cumleListe do
  4.   D ←Remove alphanumeric characters from C
  5.   D ←Remove stop words from D
  6.   WordList ← Lemmatized words of D
  7.   N ← Create New Node
  8.   N.Sentence = C
  9.   N.LemmatizedWords ← WordList
  10.   nodeList.append(N)
  11. G (V,E) ← CreateGraph(nodeList)   // Algorithm 1
  12. abstract ←∅
  13. While size of abstract < maxSize do
  14.   abstract.append(ApplyVertexCover(G(V,E)))   // Algorithm 2
  15. Output = abstract

Algorithm 3 outlines the fundamental steps of the biomedical document summarizing method. The weight of the nodes on the overall graph that represent the sentence is first determined using the Malatya algorithm. The subsequent phases of the technique then involve removing the pertinent node from the graph in order to identify the candidate summary that is most similar to the golden summary.

The proposed BioGraphSum algorithm in the current study takes the Biomedical Document Corpora as input and provides summarized biomedical texts.

The first stage obtains the sentences that make up the Biomedical Document Corpora (line 2).

After the sentence list is created, lines 3–10 of Algorithm 3 are performed as an iterative step, where all sentences in the corpora are first stripped of alphanumeric characters (line 4). Then, in order to avoid unnecessary repetition and to prevent meaningless stop words from creating unnecessary edges, all stop words in the sentences are deleted (line 5).

In sentences where there is similar and/or the same meaning, these are transferred to the textual diagram as a single element (line 6). A node is then created for each preprocessed sentence (line 7).

Lines 4–6 must offer a strong textual depiction. However, in order to include the most pertinent sequences in the summary in the same order at the end of the process, they must be correctly stored (line 8).

Control is then transferred to Algorithm 1, and highly representative textual graphs are created (lines 12–14).

The final stage adds the nodes (representing the sentences) selected by the Malatya algorithm to the summary according to the order in the biomedical text corpora, based on the steps determined in Algorithm 2.

5. Experimental evaluations

This section presents an assessment of the BioGraphSum algorithm as well as describing the evaluation corpus and evaluation metrics used in the study. Reports are made regarding the performance of the introduced model and then its performance evaluated in comparison to state-of-the-art models from the current literature.

5.1. Evaluation corpus

To the best of the authors' knowledge, no publicly available dataset exists that is frequently used for summarizing biomedical text, with studies using collections of published articles from the biomedical field [27–32]. Gold/reference abstracts are accepted as standalone article abstracts in biomedical libraries, whilst parts of articles other than the abstract are accepted as the texts to be summarized. In order to test the performance of the introduced BioGraphSum text summarization model, a corpus was created consisting of 450 current scientific research articles from the PubMed corpus in accordance with previous studies. The BioGraphSum model was then tested using this 450-article corpus [33].

5.2. Evaluation metrics

The ROUGE (Recall Oriented Understudy of Gisting Evaluation) [9] metric was used to assess the effectiveness of the summarization system. The ROUGE recall-based evaluation approach calculates n-gram-based recall for the candidate text in relation to the references for the text generation task. The following metrics were used to assess the model's performance: ROUGE, f-score (F_{lcs}), recall (R_{lcs}), and precision (P_{lcs}). The recall metric tests the extent to which the system is covered, with its accuracy quantified according to the level of accurate information supplied by the system. The current study employed the ROUGE-N, ROUGE L, ROUGE-W-1.2, and ROUGE-SU metrics to assess the BioGraphSum model. ROUGE-N was used to assess the n-gram number of the summary generated by the BioGraphSum model.

$$ROUGE - N = \frac{\sum\limits_{C \in \{Reference\ Summaries\}} \sum\limits_{gram_n \in S} Count_{match}(gram_n)}{\sum\limits_{C \in \{Reference\ Summaries\}} \sum\limits_{gram_n \in S} Count(gram_n)} \tag{1}$$

$$R_{lcs} = \frac{LCS(X, Y)}{m} \tag{2}$$

$$P_{lcs} = \frac{LCS(X, Y)}{n} \tag{3}$$

$$F_{lcs} = \frac{(1 + \beta^2) R_{lcs} P_{lcs}}{R_{lcs} + \beta^2 P_{lcs}} \tag{4}$$

The recall metric is based on the number of correct sentences in the candidate summary ($LCS(X, Y)$) and the total number of sentences in the model summary (m). The precision metric uses the number of correct sentences in the candidate summary ($LCS(X,Y)$), whilst the total number of sentences in the candidate summary (n), and F_{lcs} uses parameter β to determine the balance between the recall and precision metrics.

Countmatch, where n is the length of the n-gram, or the number of n-grams that cross paths in both the draft summary and the reference summary.

5.3. Experimental results

The current study's experimental procedures were conducted using Python on a desktop computer with 16 GB of memory and an AMD Ryzen 7 PRO 4750U CPU with Radeon Graphics running at 1.70 GHz.

Since the texts utilized to gauge the performance of the proposed BioGraphSum model are actual published articles, the application of data pretreatment and preparation phases were considered essential. The performance of the developed innovative paradigm critically depends on high quality representation. During the data preprocessing and preparation stages, stop words and punctuation were first removed from the texts, then Lemmatization etc., transactions were applied. After the preprocessing stage, the relationships between the word units were accurately transferred to textual graphics, with source documents modeled in the form of weighted and undirected graphs. As a result, graphs with a high level of textual expression representation were created by the model.

Next, the Malatya algorithm was applied, which is an approach not previously used to resolve any real-world problem, with each step of the algorithm followed. Through application of the algorithm's iterative steps, all nodes were scanned to cover all edges forming the representative graph. A subset containing the most efficient nodes was then obtained. In the experimental studies, the most information-bearing node of each iteration was selected in polynomial time, since the Malatya algorithm, unlike prior algorithms, offers solutions in polynomial time. By including the selected node, in summary, all adjacent edges to the relevant node were then deleted and removed from the graph. Erased edges represent common text groups; thus, in the next stage, text phrases in the main text were determined that had yet to be included in the abstract, thereby preventing repetitious addition of the same text phrases to the

summary. At this stage, the units (sentences) that made up the text were selected according to the level of their content. As such, very rich extracts were obtained at the desired compression ratios of 90 %, 80 %, 70 %, and golden summary size. After expressing a text set using a graph, the fact that every feature was accessible from a subset of nodes in the vertex cover problem enabled the production of summaries with a high level of coverage. Finally, the generated summary sentences were arranged according to their order in the main text so as to ensure their integrity and the overall fluency of each generated summary. In order to cover all of the text's existing subtopics, the most informative sentences were chosen and combined to generate the final summary.

In addition, the performance and efficiency of the BioGraphSum model were also measured according to various recognized performance metrics. ROUGE 1–4, ROUGE-L, ROUGE-W, ROUGE-S, and ROUGE-SU measures were used to thoroughly analyze the BioGraphSum model's performance and accuracy based on system recall, precision, and f-score values. Since the average size of the abstracts were between 15 % and 35 % of the original text size, the experimental method was repeated in order to produce summaries of various sizes, with four distinct compression ratios [8,25,30]. Thus, the variation of the introduced model's performance according to compression ratio could also be reported. The results achieved confirm the authors' hypothesis that successful biomedical text summarization can be achieved without use of a domain-specific database such as UMLS, SemRep, MetaMap, or OGER.

It is generally accepted that summarizing biomedical texts can be very challenging due to the regular use of abbreviations, acronyms, and symbols. However, as can be seen from the current study's experimental procedures, use of the Malatya centrality algorithm on biomedical texts can generate high-quality summaries through accurate extraction of the most meaningful sentences in the text. In the current study's experimental process, the success of the BioGraphSum model was replicated for the golden summary size, and for compression ratios of 90 %, 80 %, and 70 %. The ROUGE performance metrics were used to gauge the model's effectiveness, as shown in Table 1 and Table 2. Separate findings are presented for the ROUGE-1-2-3-4, ROUGE-L, ROUGE-W, ROUGE-S, and ROUGE-SU metrics, with recall, precision, and f-score measures also presented for each metric. In Tables 1 and 2, the performance metric is displayed in the first column, while the scores returned by the BioGraphSum model at the 95 % confidence interval are displayed in the other columns according to summary size (golden summary and 90 % compression ratio in Tables 1 and 80 % and 70 % compression ratios in Table 2).

As can be seen from Table 1 and Table 2, the highest ROUGE (1–4), ROUGE-L, ROUGE-W, ROUGE S, and ROUGE-SU scores were calculated as 0.77344, 0.3117, 0.1332, 0.0666, 0.5165, 0.1131, 0.5847, and 0.58666, respectively (recall-based). The most successful results based on recall were obtained without exception when the compression ratio was 70 %. Again, in Table 1 and Table 2, the highest precision-based ROUGE (1–4), ROUGE-L, ROUGE-W, ROUGE S, and ROUGE-SU scores obtained were 0.3722, 0.1025, 0.04, 0.0195, 0.2135, 0.12888, 0.1356, and 0.13812, respectively. The highest performance values based on precision were reported for the golden summary size dimension. In Table 1 and Table 2, an f-score-focused evaluation also reported ROUGE (1–4), ROUGE-L, ROUGE-W, ROUGE S, and ROUGE-SU scores as 0.3709, 0.105, 0.0434, 0.0225, 0.2103, 0.0759, 0.133, and 0.13513, respectively.

Upon examining these performance results in general, it can be seen that the most successful results in terms of recall scores were obtained for the compression ratio of 70 %, without exception. Although high-performance values were reported for all summary dimensions based on an f-score evaluation, most were obtained for the golden summary size. In the precision-based evaluation, without exception the most successful results were obtained for the golden summary size.

In general, the recall results improved when there were more sentences in the output. As can be seen in the related tables, the recall value showed a net increase as the size of the summary grew. There was an inverse relationship established between the summary sizes of the BioGraphSum model and the precision and f-score values. As the compression ratio decreased (summary size increased), the precision and f-score values were seen to decrease.

5.4. Comparison with state–OF–the-art methods

State-of-the-art methods techniques were then compared to the performance outcomes obtained using the introduced BioGraphSum model. In addition, comparisons were made between summaries generated with an 80 % compression ratio in order to provide a fair comparison with previous similar summarization approaches reported in the literature.

There are other current and successful studies on biomedical texts in the literature that are graph-based and that provide good results [2,4]. However, these studies were excluded from the current study's comparison in order to provide a fair and realistic

Table 1
ROUGE scores of BioGraphSum summaries for 90 % compression ratio.

ROUGE evaluation methods	BioGraphSum summaries (95 % confidence interval)					
	Golden summary size			Compression ratio 90 %		
	recall	precision	f-score	recall	precision	f-score
ROUGE-1	**0.3905**	0.3722	0.3709	**0.5479**	0.284	0.3551
ROUGE-2	**0.1077**	0.1025	0.1022	**0.1587**	0.085	0.105
ROUGE-3	**0.042**	0.04	0.0398	**0.0589**	0.0337	0.0407
ROUGE-4	**0.0208**	0.0195	0.0196	**0.0282**	0.0171	0.0201
ROUGE-L	**0.2208**	0.2135	0.2103	**0.3051**	0.1649	0.2023
ROUGE-W	0.0509	**0.12888**	0.0709	0.0677	**0.09608**	0.0751
ROUGE-S	**0.1545**	0.1356	0.133	**0.2942**	0.0864	0.1139
ROUGE-SU	**0.15675**	0.13812	0.13513	**0.29676**	0.08749	0.11532

Table 2
ROUGE scores of BioGraphSum summaries for 80 % and 70 % compression ratio.

ROUGE evaluation methods	BioGraphSum summaries (95 % confidence interval)					
	Compression ratio 80 %			Compression ratio 70 %		
	recall	precision	f-score	recall	precision	f-score
ROUGE-1	**0.691**	0.1985	0.2942	**0.77344**	0.1513	0.2439
ROUGE-2	**0.2417**	0.0696	0.1031	**0.3117**	0.0608	0.098
ROUGE-3	**0.0961**	0.0292	0.0426	**0.1332**	0.0271	0.0434
ROUGE-4	**0.0464**	0.0148	0.0212	**0.0666**	0.0142	0.0225
ROUGE-L	**0.4242**	0.1244	0.1829	**0.5165**	0.1024	0.1645
ROUGE-W	**0.0927**	0.07152	0.0759	**0.1131**	0.05892	0.0734
ROUGE-S	**0.4673**	0.0453	0.0742	**0.5847**	0.0271	0.0486
ROUGE-SU	**0.46956**	0.04581	0.07498	**0.58666**	0.02734	0.04903

evaluation. The selected state-of-the-art methods in the current study were biomedical and graph-based approaches similar to the proposed BioGraphSum model. Recent studies on biomedical summarization systems show that the LexRank algorithm produces superior outcomes to alternative graph-based techniques [8,25]. Therefore, the LexRank algorithm was included in the comparisons as a precursor to graph-based biomedical summarization methods [34]. By assigning points to sentences based on the LexRank method, inferential summaries were provided. In the framework of NLP, Erkan et al. suggested a stochastic and graph-based method to determine the relevance of text units. They assessed the significance of sentences in the representative graphs using an eigenvector centrality metric (based on node centrality). The authors showed that LexRank frequently outperformed both degree-based and centroid-based algorithms through their experimental approach. Another competitive and successful method is known as "Leveraging BERT," where BERT and K-Means clustering are used together [35].

Additionally, the baselines used the unsupervised MultiRank algorithm, which successfully models networks by using word similarity, semantic similarity, and common reference similarity. Using SemRep, MetaMap, and OGER, the MultiGBS algorithm extracts concepts and relationships from UMLS data [8].

The proposed BioGraphSum model was tested against a corpus created from 450 current scientific research articles published in PubMed. In Table 3 the highest value for each row is presented as bold text. Based on the values for the ROUGE-1, ROUGE-2, ROUGE-L, and ROUGE-SU metrics, the BioGraphSum model's results with a compression ratio of 80 % were compared with state-of-the-art models for each performance parameter. When Table 3 and Fig. 4 are examined, it can be seen that LexRank produced the highest recall-based value for ROUGE-1, ROUGE-2, and ROUGE-L metrics, with the proposed BioGraphSum model producing the second highest results to LexRank for both ROUGE-1 and ROUGE-2.

As can be seen from the corresponding graph presented as Fig. 5, the BioGraphSum model reported ROUGE-1, ROUGE-2, and ROUGE-L values with the highest precision focus compared to the state-of-the-art models.

Similarly, the BioGraphSum model also reported the highest ROUGE-1, ROUGE-2, and ROUGE-L values focused on f-scores compared to the state-of-the-art models.

As can be seen from Table 3 and Figs. 4–6, although the BioGraphSum model did not produce the highest results for every metric, the model achieved the most highest values in the comparisons, and the second-best values were generally achieved in cases where the highest values could not have been obtained in the experimental process applied that was.

The comparison results show that the proposed and innovative BioGraphSum model generally outperformed the state-of-the-art

Table 3
ROUGE results for state-of-the-art methods.

| ROUGE evaluation methods	Compression ratio 80 %			
	summary methods			
	Leveraging BERT	LexRank	MultiGBS	Proposed BioGraphSum
recall				
ROUGE-1	0.649	**0.720**	0.586	0.691
ROUGE-2	0.176	**0.256**	0.151	0.2417
ROUGE-L	0.403	**0.559**	0.452	0.4242
ROUGE-SU	0.345	0.425	0.301	**0.46956**
precision				
ROUGE-1	0.101	0.064	0.111	**0.1985**
ROUGE-2	0.034	0.027	0.035	**0.0696**
ROUGE-L	0.062	0.083	0.103	**0.1244**
ROUGE-SU	0.045	0.033	**0.052**	0.04581
f-score				
ROUGE-1	0.155	0.108	0.164	**0.2942**
ROUGE-2	0.051	0.0455	0.052	**0.1031**
ROUGE-L	0.095	0.129	0.146	**0.1829**
ROUGE-SU	0.070	0.057	**0.075**	0.07498
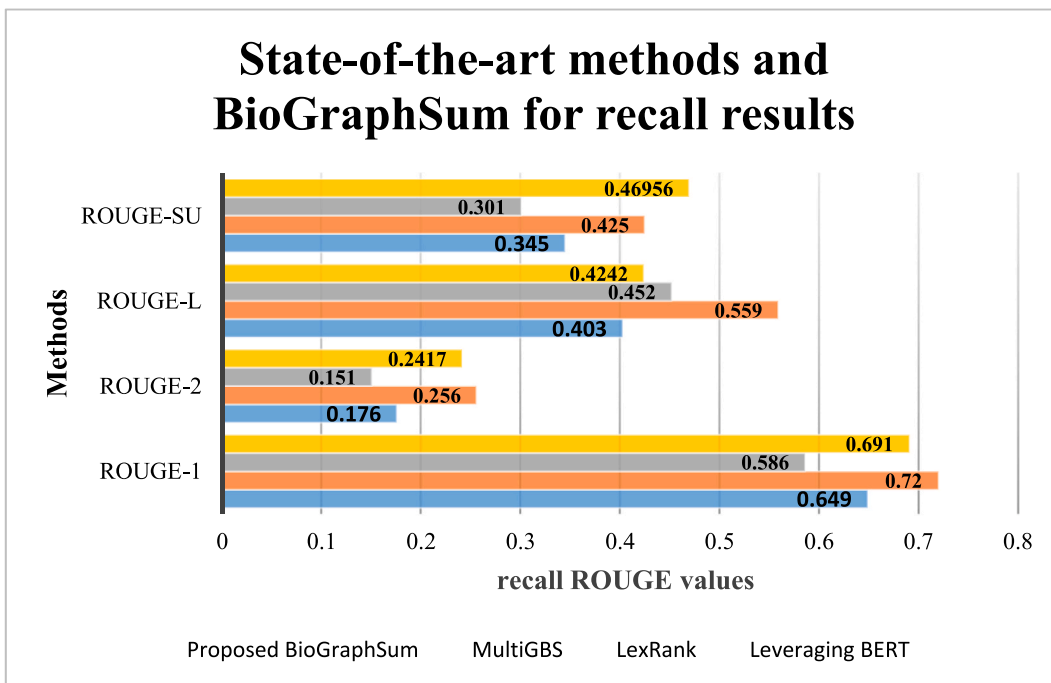
Fig. 4. Comparison of BioGraphSum and state-of-the-art methods (recall-based ROUGE).
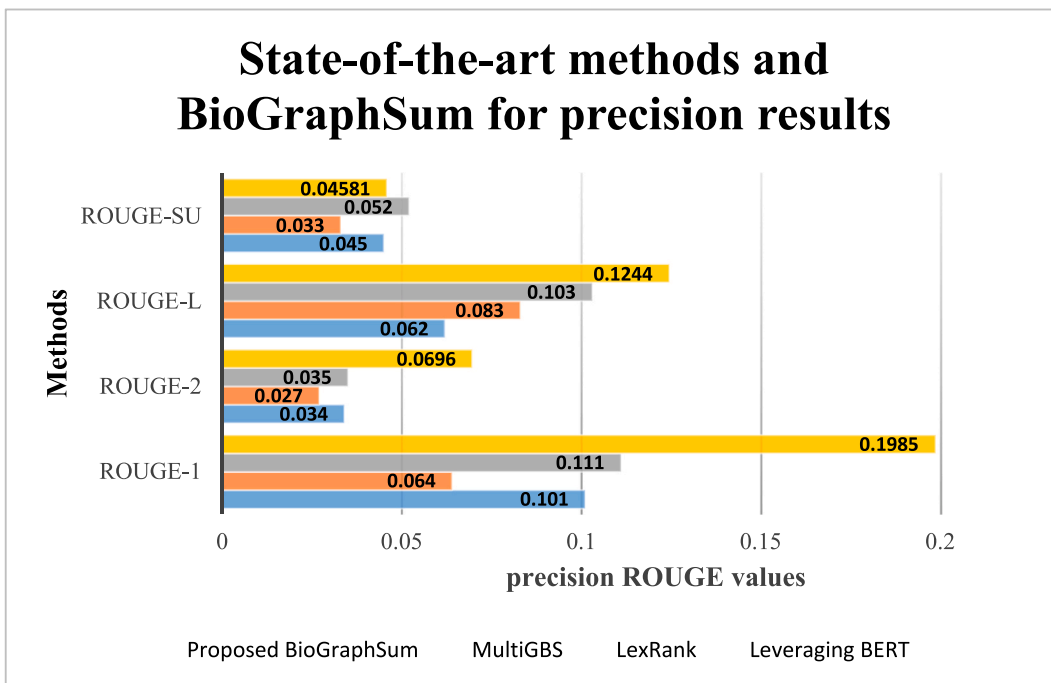

Fig. 5. Comparison of BioGraphSum and state-of-the-art methods (precision-based ROUGE).

studies.

Precision-based ROUGE values provide information on the detection of redundancy in general in the text, whereas recall-based ROUGE values represent coverage, and f-score values reflect both redundancy and coverage at the same time. Therefore, a high f-score value signifies a lower level of repetition and more meaning (i.e., a higher quality summary). Upon examination of Tables 3 and it can be seen that the state-of-the-art methods need to be included in terms of f-scores (except ROUGE-SU), indicating that the
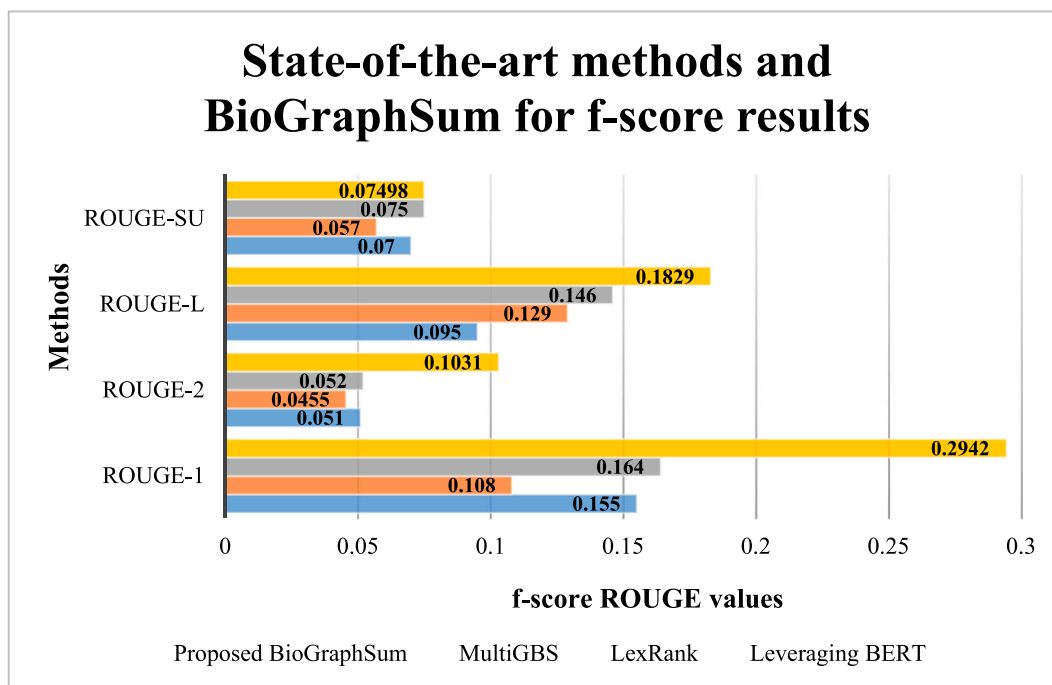
Fig. 6. Comparison of BioGraphSum and state-of-the-art methods (f-score-based ROUGE).

BioGraphSum model produced only a low level of repetition and the most meaning.

6. Discussion

The innovative BioGraphSum model introduced in this scholarly article is assessed in this section for its efficacy as a novel biomedical text summarization method. The unique feature of the BioGraphSum model is that it has no reliance upon the use of conventional biomedical databases such as the Unified Medical Language System (UMLS). Instead, it utilizes the novel approach of Malatya centrality, which is based on graph theory. Text sentences are viewed as nodes on a graph, and in order to improve the quality of the summaries generated, the most influential nodes are given priority. According to the findings presented in this study, the BioGraphSum model performed noticeably better than state-of-the-art approaches such as LexRank in two important ROUGE measures: precision and f-score, which gauge the summary's relevance and information redundancy, respectively.

The presented findings have significant ramifications for the field of biomedical text summarization, with BioGraphSum showing adaptability and usefulness across different biomedical contexts by eschewing dependency on UMLS and other related tools. This innovative model could democratize text summary methods in settings where access to domain-specific databases is either scarce or nonexistent. Additionally, by avoiding the requirement for significant computational resources and domain-specific training, the BioGraphSum method helps in the scaling of text summarization applications.

The model BioGraphSum does have certain limitations, though. The intricacy of biomedical language and the profusion of acronyms within biomedical texts presents formidable obstacles that may periodically affect the summarization's effectiveness. Therefore, future work should concentrate on improving the model's capacity to comprehend and analyze the intricacies of biomedical language. Furthermore, there is significant advancement still required in order to handle redundancy and potential overlap in information content, especially in multi-document generated summaries. In addition, the abundance of biomedical terms and abbreviations found in biomedical texts was found to negatively affect the performance of the BioGraphSum model. Another limitation of the current study is that not all of the results for the BioGraphSum model (ROUGE 1-2-3-4-L-W-S-SU) could be compared to the state-of-the-art methods since their results were limited only to ROUGE 1-2-L-SU. However, despite these limitations, it is considered an important attribute that the BioGraphSum model does not require any training-test data, nor does it require heavy hardware infrastructure in order to summarize biomedical texts.

The introduced BioGraphSum model offers a promising new direction for biomedical text summarization, suggesting a shift towards more adaptable, resource-efficient technologies in natural language processing. Its success in leveraging graph-based methodologies to produce high-quality summaries underscores its potential utility and marks a significant step forward in the ongoing evolution of text summarization techniques.

7. Conclusion

This article introduced a new and original graph-based approach for summarizing biomedical texts. The BioGraphSum model allows for summaries to be generated that include only effective and efficient sentences based on their similarity with the main subthemes of the overall text. In the proposed model, biomedical texts are modeled as undirected and weighted graphs based on the relationships between sentences. After transferring the texts to the graphs, a multi-document, generic, unsupervised document summary system takes graph-based biomedical text summarization one step further through use of the vertex cover algorithm. After expressing a text set with a graph, the fact that each edge can be accessed from the subset of the nodes in the vertex cover problem led to the production of summaries with high coverage. Contrary to other approaches to resolving the minimum vertex-cover problem, the Malatya algorithm, which solves in polynomial time, was used in the current study for the first time in a challenging real-world problem such as biomedical text summarization. The minimum vertex cover issue has a polynomial solution provided by the Malatya algorithm. One of the essential advantages of the proposed method is that for each iteration of the Malatya algorithm, the node carrying the most information can be selected in a polynomial time.

Whilst numerous prior studies utilized a domain-specific dictionary database, the BioGraphSum model did not use UMLS since it was considered important to remove dependency on OGER and MetaMap-style tools in domain-specific text summarization. The BioGraphSum model was then assessed using a collection of 450 biomedical publications. LexRank was also contrasted with other approaches such as MultiGBS and Leveraging-BERT. The results show that the BioGraphSum model performed better than rival tools in terms of certain ROUGE criteria. The vertex cover algorithm's ability to access every edge from a subset of nodes is how the authors of the current study explain the superior performance of the proposed BioGraphSum model performs compared to other highly effective techniques like LexRank, MultiGBS, and Leveraging BERT. Whilst the BioGraphSum model was initially developed to produce text summaries with a high level of coverage and rich abstracts, the model's robust infrastructure could also be used outside of the biomedical field in order to resolve similar issues in line with the needs of different fields.

Finally, since the BioGraphSum model does not require any training or test data, there is no need for very large datasets or heavy hardware infrastructure.

Data availability statement

The data used in this study is available from an online data repository located at https://github.com/cengizhark/Dataset-for-Pubmed-s-Biomedical-Articles. This resource may be utilized for the purpose of validating the results of the current study, for conducting further analyses, or for use in related studies. Access to the data repository will be permanently provided starting from the publication of this study.

CRediT authorship contribution statement

Cengiz Hark: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. M, W. M, Y. F, Y. Y, W. J, M. J, A systematic review of automatic text summarization for biomedical literature and EHRs, J. Am. Med. Inf. Assoc. (Aug. 2021), https://doi.org/10.1093/JAMIA/OCAB143.

[2] C. Hark, A. Karcı, Karcı summarization: a simple and effective approach for automatic text summarization using Karcı entropy, Inf. Process. Manag. 57 (3) (May 2020) 102187, https://doi.org/10.1016/j.ipm.2019.102187.

[3] W.S. El-Kassas, C.R. Salama, A.A. Rafea, H.K. Mohamed, Automatic text summarization: a comprehensive survey, Expert Syst. Appl. 165 (Mar. 2021) 113679, https://doi.org/10.1016/J.ESWA.2020.113679.

[4] C. Hark, T. Uçkan, A. Karcı, A new multi-document summarisation approach using saplings growing-up optimisation algorithms: simultaneously optimised coverage and diversity, J. Inf. Sci. (Jun. 2022), https://doi.org/10.1177/01655515221101841, 016555152211018.

[5] A. Givchi, R. Ramezani, A. Baraani-Dastjerdi, Graph-based abstractive biomedical text summarization, J. Biomed. Inf. 132 (Aug. 2022) 104099, https://doi.org/10.1016/J.JBI.2022.104099.

[6] M. Joshi, H. Wang, S. McClean, Dense semantic graph and its application in single document summarisation, Stud. Comput. Intell. 746 (2018) 55–67, https://doi.org/10.1007/978-3-319-68392-8_4/COVER.

[7] A. Karcï, S. Yakut, F. Öztemïz, A new approach based on centrality value in solving the minimum vertex cover problem: Malatya centrality algorithm, Comput. Sci. 7 (Issue:2) (Dec. 2022) 81–88, https://doi.org/10.53070/BBD.1195501.

[8] E. Davoodijam, N. Ghadiri, M. Lotfi Shahreza, F. Rinaldi, MultiGBS: a multi-layer graph approach to biomedical summarization, J. Biomed. Inf. 116 (Apr. 2021) 103706, https://doi.org/10.1016/j.jbi.2021.103706.

[9] C.-Y. Lin, ROUGE: A Package for Automatic Evaluation of Summaries (2004) 74–81 [Online]. Available: https://aclanthology.org/W04-1013. (Accessed 17 November 2022).

[10] T. Uçkan, C. Hark, A. Karcı, SSC: Clustering of Turkish Texts by Spectral Graph Partitioning, J. Polytech. (2020), https://doi.org/10.2339/politeknik.684558.

[11] C. Hark, T. Uçkan, E. Seyyarer, A. Karci, Araştırma Makalesi/research article Met in özetlemesi için düğüm merkez liklerine dayalı denetimsiz bir yaklaşım an unsupervised approach based on node centralization for text summarization 8 (3) (2019) 1109–1118.

[12] C. Hark, A. Seyyarer, T. Uckan, A. Karci, Doğal dil işleme yaklaşimlari ile yapisal olmayan dökümanlarin benzerliği, in: IDAP 2017 - International Artificial Intelligence and Data Processing Symposium, 2017, pp. 1–6, https://doi.org/10.1109/IDAP.2017.8090306.

[13] Uçkan Taner, C. Hark, Çikarimsal Metin özetleme yöntemleri, in: ENDÜSTRİDE DİJİTALLEŞME ÖRNEKLERİ, 2022, pp. 31–46.

[14] S.R. Rahimi, A.T. Mozhdehi, M. Abdolahi, An overview on extractive text summarization, in: 2017 IEEE 4th Int. Conf. Knowledge-Based Eng. Innov. KBEI 2017, 2018-January, Mar. 2018, pp. 54–62, https://doi.org/10.1109/KBEI.2017.8324874.

[15] P. Verma, A. Verma, S. Pal, An approach for extractive text summarization using fuzzy evolutionary and clustering algorithms, Appl. Soft Comput. 120 (May 2022) 108670, https://doi.org/10.1016/J.ASOC.2022.108670.

[16] M. Zhao, et al., QBSUM: a large-scale query-based document summarization dataset from real-world applications, Comput. Speech Lang 66 (Mar. 2021) 101166, https://doi.org/10.1016/j.csl.2020.101166.

[17] T. Jo, K nearest neighbor for text summarization using feature similarity, Proc. - 2017 Int. Conf. Commun. Control. Comput. Electron. Eng. ICCCCEE 2017 (Feb. 2017), https://doi.org/10.1109/ICCCCEE.2017.7866705.

[18] P. Goyal, E. Ferrara, Graph embedding techniques, applications, and performance: a survey, Knowl. Base Syst. 151 (Jul. 2018) 78–94, https://doi.org/10.1016/J.KNOSYS.2018.03.022.

[19] M.R. Hussain, A. Qahmash, S. Alelyani, M.S. Alsaqer, Effective graph based vertex cover algorithm using multi-document node summarizations, in: 2020 IEEE Asia-Pacific Conf. Comput. Sci. Data Eng. CSDE 2020, Dec. 2020, https://doi.org/10.1109/CSDE50874.2020.9411596.

[20] N. Alami, M. Meknassi, S. Alaoui Ouatik, N. Ennahnahi, Arabic text summarization based on graph theory, Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. AICCSA 2016 (July) (2016), https://doi.org/10.1109/AICCSA.2015.7507254.

[21] H. Rashidghalam, M. Taherkhani, F. Mahmoudi, Text summarization using concept graph and BabelNet knowledge base, in: 2016 Artif. Intell. Robot. IRANOPEN 2016, 2016, pp. 115–119, https://doi.org/10.1109/RIOS.2016.7529500.

[22] S.F. Mousavi, M. Safayani, A. Mirzaei, H. Bahonar, Hierarchical graph embedding in vector space by graph pyramid, Pattern Recogn. 61 (2017) 245–254, https://doi.org/10.1016/j.patcog.2016.07.043.

[23] K. Al-Sabahi, Z. Zhang, J. Long, K. Alwesabi, An enhanced latent semantic analysis approach for Arabic document summarization, Arabian J. Sci. Eng. 43 (12) (Dec. 2018) 8079–8094, https://doi.org/10.1007/s13369-018-3286-z.

[24] X. Cai, et al., COVIDSum: a linguistically enriched SciBERT-based summarization model for COVID-19 scientific papers, J. Biomed. Inf. 127 (Mar. 2022) 103999, https://doi.org/10.1016/J.JBI.2022.103999.

[25] M. Nasr Azadani, N. Ghadiri, E. Davoodijam, Graph-based biomedical text summarization: an itemset mining and sentence clustering approach, J. Biomed. Inf. 84 (Aug. 2018) 42–58, https://doi.org/10.1016/j.jbi.2018.06.005.

[26] I. Schmiedova, A. Dembickaja, L. Kiselakova, B. Nowakova, P. Slama, Using of amniotic membrane derivatives for the treatment of chronic wounds, Membranes 11 (12) (Nov. 2021) 941, https://doi.org/10.3390/membranes11120941.

[27] L. Plaza, A. Díaz, P. Gervás, A semantic graph-based approach to biomedical summarisation, Artif. Intell. Med. 53 (1) (Sep. 2011) 1–14, https://doi.org/10.1016/J.ARTMED.2011.06.005.

[28] R. Mishra, et al., Text summarization in the biomedical domain: a systematic review of recent research, J. Biomed. Inf. 52 (Dec. 2014) 457–467, https://doi.org/10.1016/J.JBI.2014.06.009.

[29] M. Moradi, N. Ghadiri, Quantifying the informativeness for biomedical literature summarization: an itemset mining method, Comput. Methods Progr. Biomed. 146 (Jul. 2017) 77–89, https://doi.org/10.1016/J.CMPB.2017.05.011.

[30] H.D. Menéndez, L. Plaza, D. Camacho, A genetic graph-based clustering approach to biomedical summarization, ACM Int. Conf. Proceeding Ser. (2013), https://doi.org/10.1145/2479787.2479807.

[31] M. Moradi, CIBS: a biomedical text summarizer using topic-based sentence clustering, J. Biomed. Inf. 88 (Dec. 2018) 53–61, https://doi.org/10.1016/J.JBI.2018.11.006.

[32] E. Lloret, L. Plaza, A. Aker, The challenging task of summary evaluation: an overview, Comput. Humanit. 52 (1) (Mar. 2018) 101–148, https://doi.org/10.1007/S10579-017-9399-2/TABLES/6.

[33] C. Hark, Dataset for pubmed's biomedical articles. https://github.com/cengizhark/Dataset-for-Pubmed-s-Biomedical-Articles. (Accessed 26 March 2024).

[34] G. Erkan, D.R. Radev, Lexrank: graph-based lexical centrality as salience in text summarization, J. Artif. Intell. Res. 22 (2004) 457–479.

[35] D. Miller, Leveraging BERT for Extractive Text Summarization on Lectures, Jun. 2019, https://doi.org/10.48550/arxiv.1906.04165.