IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 11 January 2025, accepted 30 January 2025, date of publication 4 February 2025, date of current version 13 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3538886

TOPICAL REVIEW

Current Trends and Advances in Extractive Text Summarization: A Comprehensive Review

MARYAM AZAM[1], SHAH KHALID[1], SULAIMAN ALMUTAIRI[2], HASAN ALI KHATTAK[1,3], ABDALLAH NAMOUN[4], (Member, IEEE), AMJAD ALI[5], AND HAFIZ SYED MUHAMMAD BILAL[1]

[1]School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
[2]Department of Health Informatics, College of Public Health and Health Informatics, Qassim University, Buraydah, Qassim 51452, Saudi Arabia
[3]Peter Faber Business School, Australian Catholic University, North Sydney, NSW 2060, Australia
[4]AI Center, Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, Saudi Arabia
[5]Department of Computer and Software Technologies, University of Swat, Swat, Khyber Pakhtunkhwa 19130, Pakistan

Corresponding authors: Shah Khalid (shah.khalid@seecs.edu.pk) and Sulaiman Almutairi (ssmtiery@qu.edu.sa)

The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2025).

ABSTRACT Given the rapid increase of textual data in various fields, text summarization has become essential for efficient information handling. Over recent decades, numerous methods have been proposed to enhance summarization processes, and various review papers and books have been published to encapsulate these methodologies and discuss their implications. However, existing reviews often fail to provide a comprehensive retrospective of recent advancements, particularly concerning detailed architectural frameworks, the field's current state, evaluation methodologies, and unresolved challenges. This paper addresses this gap by presenting a detailed analysis of the extractive approaches, encompassing their inherent strengths, limitations, and underlying mechanisms. We present a detailed, multi-layered architectural framework designed to advance and develop summarization models, thereby supporting researchers in their endeavors. The text summarization framework consists mainly of text preprocessing, feature extraction, sentence scoring, use of a base model, sentence selection and output summary, and post-processing. Furthermore, this review of 145 research articles categorizes domain-specific summarization techniques, focusing on unique challenges and tailored strategies for news, scientific articles, and social media. These techniques include statistical, fuzzy logic, rule, optimization, graph, clustering-based, machine learning, and deep learning. We emphasize the impact of evaluation metrics and benchmark datasets in performance assessment, providing a detailed analysis of the commonly utilized datasets and metrics (mainly ROUGE-1, ROUGE-2, ROUGE-L, and ROUGE-S) in the current literature. This review article is a valuable resource for advancing text summarization techniques in natural language processing and machine learning by identifying future research directions and open challenges. Notable challenges include expanding summarization for complex tasks, multiple documents, multimodal user input, multi-format and multilingual data, refining the stopping criteria, and improving the evaluation metrics.

INDEX TERMS Survey, text summarization, transformer-based models, domain-specific summarization, generic architecture, datasets and evaluation measures.

I. INTRODUCTION

In the modern era, data is exploding at the fastest pace ever, generated by people mainly on social media networks, websites, blogs, and news platforms [1]. However, this surge has resulted in a challenge: When people search for information, they tend to find a deluge of data, making it challenging to retrieve relevant results. The results must be compressed to give people access to the necessary information to solve this problem. However, another subsequent challenge arises in the difficult task of summarising the enormous amount of data manually [2].

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

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

FIGURE 1. **Model architecture of an automatic text summarizer.**

Researchers have worked on automated text summarization techniques to address the issue of manual summarization. These approaches create summaries by determining the keywords and important information within the text while preserving the actual meaning of the sentences [3]. The basic structure of the Automatic Text Summarization (ATS) system is given in Figure 1.

ATS includes a variety of techniques. The source documents have been categorized as a single-document or multi-document text summarization [4]. The output of the summaries has been divided into two categories: extractive and abstractive summarization [5]. Extractive summarization holds a significant position in the field of Natural Language Processing (NLP). Its primary function is to condense lengthy texts into concise summaries [6]. The importance of sentences depends on the linguistic and statistical characteristics of sentences [7].

ATS is considered one of the hardest challenges in the field of NLP. Researchers have been frequently working in the domain of ATS since 1958 [8]. Researchers are still working to develop a better ATS system. The field continues to advance due to this attempt to improve the ATS system [9].

In 2018, Kirmani et al. proposed various extraction methods along with key statistical characteristics [10]. Dutta et al. focused on the in-depth study of extractive techniques utilized in the summarization of micro-blogs [11]. A comprehensive review of extractive ATS systems that employ fuzzy logic techniques is proposed by Kumar et al. [12]. In 2019, Mosa et al. conducted a survey exploring the application of Swarm Intelligence Optimization strategies in ATS [13].

A study by Suleiman et al. focused on Extractive Text Summarization (ETS) using Deep Learning (DL) techniques [14]. In 2019, the research by Bhattacharya et al. focused on a particular type of summarization, especially summaries made for legal documents [15]. In 2020, Singh et al. formulated an ensemble approach for ETS based on several Machine Learning (ML) models [16]. Gupta et al. used the Elmo embedding technique to generate an extractive summary in 2020 [17]. Elbarougy et al. performed a study in which the summary was extracted using a Modified PageRank technique [18].

In 2020, Haider et al. proposed an approach to ETS by combining K-means clustering and the Gensim Word2Vec method [19]. Jugran et al. introduced a technique for ETS that provides several algorithm choices for a specific task [20]. In 2021, Muthu et al. proposed a Deep Learning Modifier Neural Network (DLMNN) classifier for processing documents [21]. Aljevic et al. developed a unique graph-based technique for ETS [22].

In 2022, Yadav et al. started their research by implementing a reinforcement learning-based fundamental model [23]. In order to minimize topic bias, Srivastava et al. proposed an unsupervised technique for ETS that included topic modeling and clustering [24]. Verma et al. proposed a hybrid approach for ETS that combines clustering, evolutionary, and fuzzy methods [25]. In this study [26], a neural model called N-GPETS that combines a Bidirectional Encoder Representations from Transformers (BERT) model with a graph attention model is introduced by Umair et al. In 2022, Gupta et al. proposed a method that is based on sentence raking for ETS [27].

In 2023, Joshi et al. presented the DeepSumm technique for ETS [28]. Khassawneh et al. used a textual graph technique to produce a cohesive extractive summary [29]. Thirumoorthy et al. proposed a social mimic hybrid optimization method for extractive summarization of a single document [30]. In 2023, Ghadimi et al. presented SGCSumm (Sub-modular Graph Convolutional Summarizer) for extractive summarization of multi-documents [31].

In 2024, Vo et al. utilized a meta-learning approach to develop a DL model and employed explanatory methods like input modification, SHAP, decision trees, and linear regression in order to comprehend the model's process of decision-making [32]. Yadav et al. presented a TGETS model which is a graph-based summarization method that creates summaries based on the average weight of the graph and an aggregate of sentence weights [33].

In the review paper by Adhika et al., various methods for abstractive and extractive summarization are described up to the year 2020 [5]. Moratanch et al. have written a review on extractive summarization and have examined developments up to the year 2016 [7]. Another review paper authored by Sharma et al. includes different methods and comparisons of extractive and abstractive summarization [34]. Similarly, the review paper by Kassas et al. [1] only includes research conducted up until 2020; omitting innovative techniques and developments that were made beyond that year. This gap emphasizes the need for a current, comprehensive review that overcomes these temporal limitations and encompasses the most recent developments, challenges, and trends in extractive summarization. On the other hand, our review paper aims to cover methodologies and developments specifically in the field of extractive summarization up to the year 2024. This unique aspect distinguishes our work, as we present a comprehensive and up-to-date review of ETS. Our work offers researchers valuable information about recent research and developments in the field of extractive summarization to help them comprehend the trends, challenges, and advancements in it [35].

Contributions of this manuscript include a comprehensive study of extractive summarization as mentioned below:

- The evolution and current state of extractive text summarization.
- The classification of extractive text summarization approaches.
- Applications of extractive text summarization.
- Architecture for extractive text summarization.

- Evaluation metrics that are used to assess the quality of the summary.
- The benchmark data sets that are currently in use for the research of extractive text summarization.
- Open research issues and challenges in the field of extractive text summarization.

The paper is structured in the following manner for the remaining sections: Section II discusses the selected papers and outlines the survey methodology. Section III discusses the evolution of extractive text summarization over time and the current state of extractive text summarization approaches. Section IV presents the classification of extractive text summarization approaches. Section V presents the applications of the extractive text summarization system. Section VI presents a generic and layered architecture for text summarization. Section VII explains the metrics for the evaluation of text summarization and presents the benchmark data sets used for text summarization. Section VIII discusses the open issues and challenges in the field of extractive text summarization. Section IX provides the conclusion, summarizing the main concepts and insights presented in the prior sections. For clarity, the structure of the paper is also illustrated in Figure 2.

II. SURVEY METHODOLOGY AND SELECTED PAPERS

Although prior comprehensive review papers offer insightful information about the field of extractive text summarization, they are limited in their coverage of the most current developments. In this paper, our approach includes a thorough review of all the papers from 2017 to 2024 in order to provide a comprehensive and up-to-date perspective. While all the papers before 2017 are not included, significant papers that are relevant to the topic are selected. We found almost all of the papers by searching on Google Scholar,[1] semantic scholar[2] and dblp[3] using specific keywords such as ''Extractive text summarization'', ''extractive summarization techniques'' and ''automatic text summarization''. Furthermore, we extended our search by looking into the reference lists of selected papers. After downloading, each paper was examined manually to ensure it was relevant to the topic.

A. SELECTED PAPERS

Throughout our process of paper selection, we have found about 145 publications that meet our specific criteria. Therefore, this paper will cover these 145 publications that are directly related to extractive text summarization and provide insights into the latest state-of-the-art approaches. Figure 3 depicts the comprehensive selection procedure for articles.

III. THE EVOLUTION AND CURRENT STATE OF EXTRACTIVE TEXT SUMMARIZATION

Early algorithms play an important role in the development of ETS field. These primitive algorithms were introduced in various years and each algorithm brings its own distinct approach for the summarization process. In 1958, Luhn et al. introduced a frequency-based algorithm for text summarization that identifies significant sentences by analyzing how frequently they occur in a document and assigning scores based on structure and content [8]. This paper established the foundation for a frequency-based algorithm and also proposed an extraction-based technique that selects the important sentences and then puts them together to form a summary.

H.P. Edmundson proposed a new technique known as the rule-based algorithm in the late 1960s [36]. This approach not only emphasized significant words and sentences but also considered title and headings, pragmatic words, and structural indicators. These three proposed components outweighed the frequency component in the process of producing better extracts.

In the late 1990s, a novel method for multi-document text summarization, employing a graph-based algorithm, was introduced by Mani et al. [37]. This approach involved analyzing the differences and similarities between two related papers by using a graph representation of the text. Another summarization system based on robust NLP was proposed by Aone et al. in the late 1990s. This system employed robust NLP technologies, including information technology, corpus-driven statistical NLP, and online resources [38]. It tried to address the limitations of traditional knowledge-based, frequency-based, and discourse-based summarization techniques by integrating features derived from these cutting-edge techniques.

Mani et al. introduced a technique to use ML from corpora to create user-specific and general summaries in the late 1990s [39]. This technique worked well and presented comprehensible rules. In addressing customized interests, these rules used specific location details from the generic rules and user-specific keywords. This method is widely used because it does not require manual tagging.

In the early 2000s, Y Gong et al. presented two ways to summarize texts: ranking and selecting sentences from the documents [40]. The first method utilized Information Retrieval techniques for ranking the importance of sentences. In contrast, the other method employed the semantic latent analysis to find the sentences which are important for the creation of a summary. Both of these methods aim to choose highly significant and distinct sentences. In this way a summary is created that covers a broader range of the document's content and has either less or no redundancy.

In the mid 2000s, K Khaikhah introduced a new approach based on neural networks for text summarization [41]. This approach involves training a Neural Network (NN) to identify which sentences are important for the summary. Then, the NN is modified to understand and combine the important features found in summary's sentences. Eventually, this modified NN acts as a filter to create concise summaries.

Arman et al. presented a hybrid method for ETS in the mid-2000s [42]. This novel approach combines Genetic

[1] https://scholar.google.com/
[2] https://www.semanticscholar.org/
[3] https://www.semanticscholar.org/

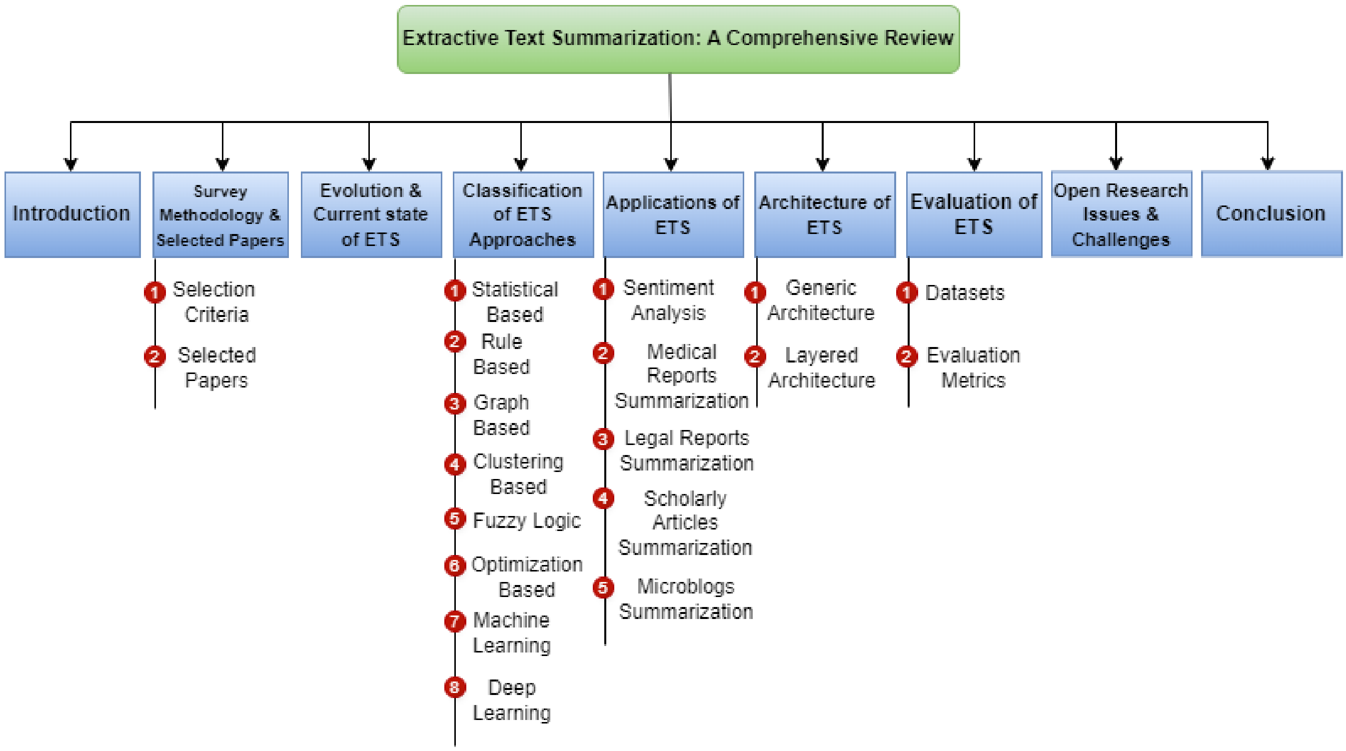


FIGURE 2. **Paper Structure: Main content of the paper.**

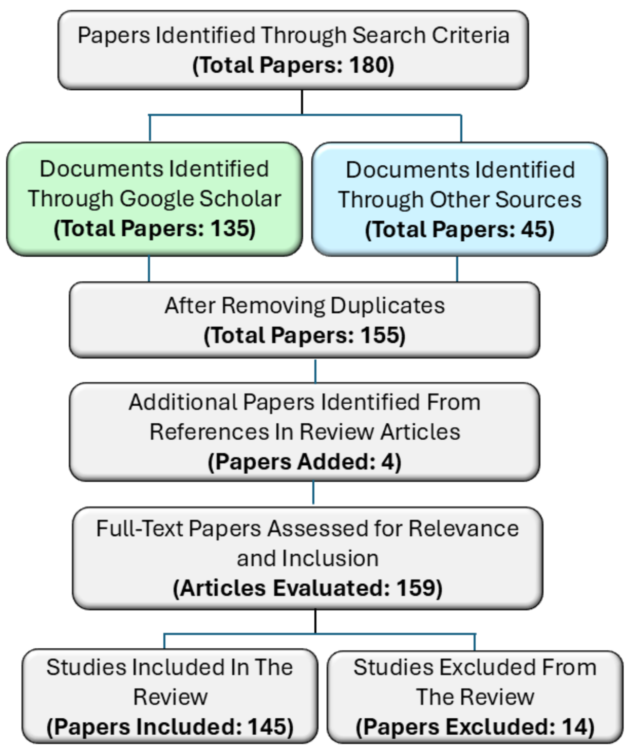


FIGURE 3. **The flowchart depicts the comprehensive selection process, from the initial identification of articles to final inclusion for detailed analysis.**

Algorithms (GA) and Genetic Programming (GP) to improve fuzzy systems membership functions and rule sets. In this approach, the string portion is handled by GA and the structural aspect is dealt with by GP.

In the mid-2010's, Ryang et al. introduced a reinforcement learning based method for creating a summary [43]. By using the summary's specific feature representation, the particular score function can be optimized. The study shows that the reinforcement learning technique can be efficiently applied to ATS problems.

In the late 2010's, Egonmwan presented a transformer based method for ETS [44]. The framework first uses a transformer to encode the original text, and then it applies a sequence-to-sequence model. Their study demonstrated that the sequence-to-sequence model and transformer work well together to produce a more comprehensive encoded vector representation.

In the late 2010's, Liu et al. demonstrated the practical application of BERT in text summarization and presented a framework which applies to both abstractive and extractive models [45]. They introduced a BERT-based document-level encoder which is capable of capturing a document's semantics and generating sentence representations. They built an extractive model by using many inter-sentence Transformer layers. The Figure 4 illustrates the evolution of ETS techniques over time.

IV. CLASSIFICATION OF EXTRACTIVE TEXT SUMMARIZATION APPROACHES

There are various methods for ETS. The given figure 5 illustrates some of these methods. A brief discussion of these methods is also included in this section.

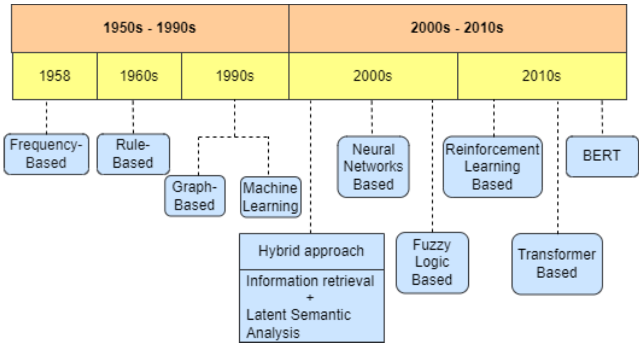


FIGURE 4. Evolution of extractive text summarization techniques.

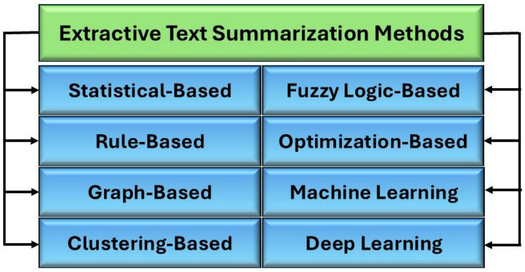


FIGURE 5. Different methods of extractive text summarization.

A. STATISTICAL-BASED APPROACHES

These approaches use statistical models to identify and extract important sentences from a document. The weighted frequency of each word is calculated in the document and then these assigned weights are used to rank the sentences. Then the sentences with the highest ranks are selected to form the final summary [46]. The TF-IDF algorithm is used for ETS [47]. The TF-IDF score of all the words is calculated and the ranking of sentences is made accordingly. The overall sentence score is derived from this TF-IDF score. Hence, all the sentences with a score greater than a certain threshold value are included in the summary. The TF-IDF algorithm can be used to capture semantic information along with linguistic and statistical features from the input document [48]. A sentence ranking approach is applied to create extractive summaries from a set of multiple documents [27].

B. RULE-BASED APPROACHES

Rule-based algorithms for extractive summarization have some predefined rules and guidelines to determine the most important and relevant information and then extract it from the given text document. These algorithms rank and score sentences based on predetermined criteria like keyword weight, title feature, sentence position, sentence length, sentence-to-sentence similarity, numerical value, and proper noun [49]. A fuzzy rule-based method employs the analysis of correlated features to diminish dimensionality, allowing text summarization efficiently and enhancing expert systems for automated text evaluation [50].

C. GRAPH-BASED APPROACHES

Sentence based graphs are used to represent documents in the ETS. The CoRank model integrates graph-based model with word-to-sentence relationships for ETS [51]. A system uses a greedy approach and submodularity framework to produce extractive summaries [52]. There exists a lexical association between words, creating a graph-representation to extract theme-conveying words which are used to identify important sentences within the text document [53]. The neural summarization for multiple documents uses sentence relation graphs and Graph Convolutional Networks, incorporating Recurrent Neural Networks (RNNs) to extract word embeddings and greedy heuristics to select important sentences [54]. An unsupervised method uses a weighted graph, where sentences are represented as vertices, and edges are determined based on the similarity or dissimilarity of sentences [55]. The triples, like subject-object-verb are extracted to form a semantic graph, and PSO is used to train a classifier that generates a sub-graph and produces a summary [56]. Another approach utilizes semantic relationships by PAS extraction, PageRank algorithm, semantic distance measurement, Maximum Marginal Relevance, and choosing top-n sentences [57]. A graph-based unsupervised approach uses Fuzzy logic for single as well as multiple documents [58]. A novel approach uses hypergraph traversals and semantic term clustering-based topic models to determine corpus topics [59]. Another method uses modified TextRank algorithm, incorporating sparse graph partitioning with weighted edges [60]. There exists a technique that considers both sentence similarity and relationship with the overall document, including topic modeling and semantic measure [61]. Multiplex-Graph Summarization model (Multi-GraS) is based on a Multiplex-Graph Convolutional Network (Multi-GCN) approach to simultaneously model diverse relationships involving sentences and words [62]. "EdgeSumm" is a novel unsupervised graph-based approach which is a combination of statistical-based, graph-based, centrality-based, and semantic-based methods [63]. Another graph-based method uses maximum independent sets and a text processing tool called "KUSH" to preserve semantic cohesion [64]. An unsupervised graph-based approach selects important sentences based on distances and similarities [65]. There is a framework which transforms the text into a sentence graph and utilizes selectivity measures to determine important nodes while the edges are determined by Cosine similarity, Jaccard and Mihalcea's measures [22]. MuchSUM is a multiple channels graph convolutional network that specifically includes several salient features by introducing three graph channels to represent the position, centrality, and textual features of the node with Bipartite graphs integrating words and sentences [66]. TGETS is a query-based summarization method based on a sense-aware semantic similarity metric, where a proposed WSD (word sense disambiguation) technique determines the exact sense of words to find relevant sentences, and feature based approach computes semantic similarity scores, with redundancy removal ensuring an informative and redundancy-free summary [33].

D. CLUSTERING-BASED APPROACHES

Clustering-based extractive text summarizing methods require placing related sentences together into clusters and then picking key sentences from each cluster to produce the summary. A single-document extractive summarization model is presented which extracts informative features, scores sentences based on similarity measures, forms clusters, and includes highly-ranked sentences from all of the clusters to form a summary [67]. There exists another clustering approach for a single document that leverages semantic analysis and topic modeling to produce an extractive summary of the text [68]. There is a clustering algorithm based on sentences (K-means) designed for single-document summarization, using Gensim word2vec for efficient feature extraction of semantic topics [19]. Another technique for summarizing multiple Arabic documents involves using Fuzzy clustering methods and LDA algorithms to cluster the documents by topic, followed by extracting key sentences [69].

E. FUZZY-LOGIC-BASED APPROACHES

These techniques use Fuzzy-Logic, similar to human reasoning which describes sentence feature values that can be characterized as zero or one [12]. There is a fuzzy-logic based approach for ETS, considering many factors to determine the most relevant sentences for creating a summary [70]. There is a statistical approach based on features that uses fuzzy logic to handle uncertainty and imprecision in feature weight, with redundancy removal through cosine similarity [71]. A fuzzy logic-based model utilizing shark smell optimization is proposed, where a meta-heuristic function optimizes feature weights, and the summary is generated by applying the dot product of feature scores and these weights [72].

F. OPTIMIZATION-BASED APPROACHES

These approaches reframe the summarization problem as an optimization task. An optimized ensemble technique uses the idea of voting classifiers and tries to increase the accuracy and quality of summaries by using the strengths of different summarization models [73]. A multi-objective technique based on decomposition is presented for multi-document ETS. The MOABC/D uses a multi-core architecture through an asynchronous parallel implementation [74]. MTSQIGA, a multi-document text summarization technique uses a modified quantum-inspired genetic algorithm to extract important sentences and optimize redundancy, relevance, and coverage within a preset length limit using a binary optimization problem [75]. A novel extractive text summarizer employs a discrete differential evolution method to select the optimal group of sentences [76]. A novel extractive multi-document summarising method employs Modified Normalised Google Distance and Word Mover Distance for content coverage and non-redundancy, and Dolphin Swarm Optimization for feature weight optimization [77]. A firefly algorithm uses fitness function based on cohesion, readability, and topic relation factors for multi-document summarization [78].

G. MACHINE LEARNING-BASED APPROACHES

These approaches transform the summarization task into a supervised sentence-level classification task, in which the system uses examples for training to categorize every sentence as either summary or not, based on a dataset of documents paired with their corresponding human created summaries. A supervised text summarization method uses a NN trained on ten features, such as word vector embedding, to effectively extract relevant features [79]. Another model uses a supervised method for generating summaries with discriminative, robust, and minimalistic features [80]. Researchers worldwide have used various ML models with limited success, prompting this work to experiment with Logistic Regression, Decision Tree, Neural Network, Random Forest, SVM models, XGBoost, and Naive Bayes, compared their results, and ultimately proposed an ensemble approach that achieves better accuracy [16]. There exists the application of two Additive models alongside interactions, GAMI-NET and Explainable Boosting Machine, to the problem of extractive summarization using binary classification and linguistic features [81].

H. DEEP LEARNING-BASED APPROACHES

These approaches use models such as RNNs, Transformers, and GNNs to identify the most significant sentences. An unsupervised deep auto-encoder for feature learning can be used to enhance these features with a Restricted Boltzmann Machine. The process includes three phases: feature extraction, enhancement, and summary generation [82]. A hierarchical document encoder and an attention-based extractor with side information are utilized [83]. Combining multilayered bidirectional long-short-term networks and memory networks, the data-driven neural network captures sentence-level information, and the n-gram features [84]. Various DL approaches for ETS include feedforward neural networks [85], attentive encoder based models with RNNs [86], the unsupervised SummCoder framework using sentence embeddings and auto-encoders [87], a combined model for extractive and compressive summarization [88], and a study using DBNs and autoencoders with defined sentence feature vectors [89]. Some more techniques include the use of BERT [45], a Contextualised-Representation of Hierarchical-Attention network for extractive summarization [90], Elmo Embeddings for encoding text into contextual vectors for sentence ranking [17], a Deep Learning-Modifier-Neural Network classifier using entropy values [21], and a topic-aware T-BERTSum model leveraging BERT for extractive summarization [91]. Combining BERT, sentence keyword extraction, and LSA topic modeling on a document effectively extracts informative sentences related to the

TABLE 1. Summary of ETS methods from selected research papers.

Sr.	Method	Algorithm	Strengths	Limitations
1.	Statistical-Based [47] [48]	TF-IDF	This method works well to generate summaries upon user request.	It can not effectively manage redundancy.
2.	Rule-Based [49]	Feature-Matching	It picks the best sentences to provide a summary of the content that covers more information.	Creating rules takes time and requires expert knowledge.
3.	Graph-Based [33] [56] [57] [61]	CoRank, TextRank, PageRank, GCNs, EdgeSumm, MuchSum, TGETS	It improves coherence and identifies redundant data.	It may not identify sentences with semantic equivalents.
4.	Clustering-Based [67] [68]	K-Means, DBSCAN	This approach addresses the absence of coherence by minimizing the impact of sentence order on the summary.	Pre-specification of the quantity of clusters is necessary.
5.	Fuzzy-Logic-Based [70] [71] [72]	Statistical Feature-based-modeling with fuzzy logic, Fuzzy-driven SSO	This method facilitates the acquisition of diverse information and substantial content coverage.	Many redundancy removal approaches are needed.
6.	Optimization-Based [73] [75] [76] [77] [78]	Ensemble Optimization, MOABC/D, MTSQIGA, Discrete-Differential-Evolution, Dolphin-Swarm-Optimization, Swarm Intelligence-based Algorithm	Ensemble optimization integrates the advantages of several algorithms to improve the summarization.	It is time-consuming and computationally costly.
7.	Machine-Learning [79] [80] [81]	Neural Networks, Ensemble approach, Explainable Boosting-Machine, GAMI-Net	It produces robust summaries with minimal features that also manage class imbalance.	For training, it needs a large number of manually produced summaries.
8.	Deep Learning [95] [28] [99] [31] [100]	Ensemble-Noisy Auto-Encoder, EV and D-EV, Hybrid MemNet, AES Model, BERT, CRHASum, Elmo embedding, DLMNN, HiStruct+ model, WL-AttenSumm, KeBioSum, DeepSumm, BERTSum, MFMMR-BertSum, SGCSumm, TGA4ExSum	Features don't need to be manually extracted. The set of features can be modified to suit the needs of the user.	The need for proper tuning of the hyperparameters and the computational expense of training.

document's topic [92]. An unsupervised approach utilizes deep document representations derived from positional encoding, self-attention, and pre-trained sentence vectors, followed by PCA-based feature extraction for sentence importance scoring and ILP-based sentence selection [93]. Recent advancements in extractive summarization leverage various DL models: the integration of Seq2Seq model and a Bidirectional Long Short-Term Memory (LSTM) model with an attention layer [23], the HiStruct+ model injects hierarchical structure details into a pre-trained Transformer, achieving state-of-the-art results on arXiv and PubMed [94]; KeBioSum enhances biomedical summarization with medical evidence data and minimal fine-tuning [95]; DeepSumm uses language and topic vector encodings to capture semantic and structural features [28]; BERTSUM-based approaches handle long documents and multiple domains texts [96]; deep feedforward networks focus on saliency and diversity [97]; MFMMR-BertSum reduces redundancy with modified sentence scoring [98]; and an integrated BERT and BiGRU model improves global context capturing for long documents [99]. The SGCSumm method, which uses the BERT pre-trained language model to represent documents, applies several transformations to turn these representations to a normalized, non-negative, submodular, and non-reducing monotone form and employs a GCN for feature learning [31]. A novel approach, termed TGA4ExSum, integrates attention mechanism based on graph neural networks (GAT) alongside BERT [100]. A BERT-based summarization approach uses multiple routes for feature learning, with BERT outputs processed through RNN, GRU, and LSTM to extract the summary. [101] MODE/D-WS is a novel methodology that enhances ATS by integrating Multiple Objective Differential Evolution with a weighted aggregate strategy and an advanced ATS repair mechanism [102]. The DeepExtract framework leverages Generative Pre-trained Transformer (GPT)-4, semantic clustering, and layered positional encoding to produce contextually precise summaries, optimizing novelty, relevance, and coherence [103]. The DCDSum framework utilizes contrastive learning transforming the task into sentence reranking to minimize redundancy and enhance interpretability [104].

Table 1 presents a comparative analysis of various summarization methods and the specific algorithms used in each method. It provides a comprehensive understanding of the effectiveness of each approach by highlighting its strengths and limitations. The table also includes citations to reference papers that discuss these techniques and their applications in detail.

V. APPLICATIONS OF EXTRACTIVE TEXT SUMMARIZATION

ETS is utilized in analytics and text mining for applications like question answering, extraction, and information retrieval. A multi-modal extractive summarising system is proposed by Li et al. to generate textual summaries from asynchronous inputs, including images, audio, video, and text [105]. The utility of extractive summarization extends prominently to domains such as sentiment analysis [106], the concise summarization of scholarly articles [99], and the synthesis of content from the microblogs [107]. Summarization systems also find use in diverse areas including email, news and other domains such as biomedical [108] or legal documents summarization [109]. The various applications of the summarization system are given below:

- Sentiment Analysis involves examining opinions, judgments, and emotions of people about products and events. It is a significant application of extractive text summarization, wherein summarization techniques are employed to enhance the interpretability of models and to improve classification accuracy. A summarization system that uses fuzzy c-means clustering to provide summaries of user reviews is proposed in this paper [106]. It creates automatic summaries of electronic product reviews, considering both the review content and the credibility of the author, by comparing the semantic and content similarity of each set of sentences. Transformer-based methods have been utilized in sentiment analysis to distill key sentences as summaries, thereby improving the interpretability of the model by harnessing attention weights within cascaded transformers or integrating sentence-level probabilities for document-level classification [110]. A hybrid model integrating NLP and LSTM has been proposed for the summarization of consumer reviews and sentiment analysis. This model incorporates pre-processing steps, hybrid feature extraction that combines review-related and aspect-related features, and LSTM-based sentiment classification to derive comprehensive consumer insights. [111]

- Extracting crucial data from medical reports and conversations and sharing it with patients and doctors can be helpful to prevent infectious diseases on time. Deep learning and transformer models were trained on clinical text data from PubMed to improve the efficiency of information retrieval within medical literature [112]. A lightweight extractive summarization system leveraging medical word embeddings and basic features delivers performance comparable to state-of-the-art models, enabling quick access to research evidence [113]. There is a summarization technique which combines Deep Dense LSTM with CNN for automated compilation of medical documents from biomedical records [108]. It turns poorly punctuated discussion transcriptions into coherent summaries using phrase selection and topic modeling.
- Retrieving essential details from legal government reports is the aim of the government-news summarization system. It helps the readers to understand the news rapidly in this era of excessive information. A text summarization model is introduced which utilizes multiple features. Initially, the TF-IDF technique and word vector embeddings from the BERT model [114] are used to extract features, sentences are then scored based on similarity, keywords, and position, with the highest-ranked sentences forming the summary [109]. BERT and TextRank have been utilized for extractive summarization of news datasets, with evaluations indicating that TextRank outperforms BERT in Recall-Oriented Understudy for Gisting Evaluation (ROUGE) score, recall, and F-measure, whereas BERT exhibits higher precision, highlighting their respective strengths in summarization tasks [115]. Lexical chain-based approaches have been employed to summarize news articles, incorporating advanced scoring techniques, pronoun resolution, and WordNet to identify key sections of text and create coherent and efficient summaries tailored to the layout of news content [116].
- Scholarly articles must be concisely summarised in order to effectively convey the primary points and importance of the research, helping readers to grasp the main ideas without having to read the complete work. In a novel method, sentence level embeddings are generated using BERT, pre-trained on large datasets which are self-supervised. BiGRU subsequently processes the embeddings to extract significant information and capture sequential relationships [99]. Wikipedia-based text summarization systems employ advanced pre-processing, feature engineering, and statistical methodologies to extract meaningful information. By utilizing fuzzy logic, cosine similarity, and LDA modeling, these systems produce precise and effective summaries while addressing the shortcomings of earlier algorithms in retaining important content [117]. For scientific article summarization, a greedy extractive approach incorporating Variable Neighborhood Search (VNS) achieves competitive ROUGE scores on PubMed and arXiv datasets [118]. This technique chooses sentences with high TF-IDF values, offering a resource-efficient alternative to complex neural models.
- Millions of posts are shared on social media sites like Twitter and Facebook. In emergencies, Twitter provides crucial real-time updates as countless tweets are posted. As a result, summarizing microblogs has gained significant importance recently. An example of a summarization system for summarizing tweets can be found in the given paper [107]. The MFMMR-BertSum model has been employed for sentence-level extractive summarization utilizing BERT with an added classification layer and leverages Maximal Marginal Relevance (MMR) to minimize redundancy and enhance the quality of summaries. This approach demonstrates exceptional performance on the CNN/DailyMail dataset, especially in the context of summarizing social media texts [98]. For social media summarization, advanced models integrating BERT, sequence-to-sequence techniques, and reinforcement learning have been effectively utilized on the LCSTS dataset, a high-quality corpus of Chinese short texts from 'Sina Weibo.' [119]. These models have attained notable improvements in ROUGE scores, highlighting their effectiveness in summarizing social media content.

VI. ARCHITECTURE FOR EXTRACTIVE TEXT SUMMARIZATION

A. GENERIC ARCHITECTURE OF ETS

Figure 6 illustrates the framework of an ETS system. It consists of several steps, as discussed below:

1) **Text-Preprocessing:**
Preprocessing the input document(s) to clean and prepare the text for further analysis such as sentence segmentation, tokenization, stop words removal, stemming, bag-of-words, etc [46].
Sentence Segmentation: Sentence segmentation is the method of dividing text into distinct sentences. This can help in recognizing the boundaries of words, facilitating further processing of each phrase. Segmentation occurs upon the existence of punctuation or full stop through the use of a sentence tokenizer [120].
Tokenization: Tokenization signifies the division of sentences into characters, punctuation, and words, referred to as tokens [120]. For example, the sentence ''NLP is reshaping search engines'' can be tokenized into ''NLP'', ''is'', ''reshaping'', ''search'', ''engines''
Stop words Removal: Stop words are the terms which occur most commonly in a given language and generally lack any significant meaning (for example: ''a'', ''is'', ''the'') [121]. The removal of these words can significantly diminish the volume of the dictionary of distinct terms.
Stemming: It is a significant preprocessing method employed for minimizing words down to their base form or root by eliminating either prefixes or suffixes [122]. The terms ''ran'', ''runner,'' and ''running'' are all condensed to the word ''run'' through the process of stemming.
Bag-of-Words: Using the bag-of-words approach, features are extracted from the written text and subsequently given to a classifier. It simply organizes the features depending on the frequent use of the words within the text [120].
2) **Feature Extraction:**
Extracting features from text to help in identifying sentence importance [48]. Traditional approaches frequently include complex ranking or scoring methodologies. Recent methodologies have incorporated deep learning techniques to advance the process of feature extraction. For instance, sentence similarity and word feature vectors can be computed using Global Vectors for Word Representation (GloVe). Furthermore, the Bi-Gated Recurrent Unit, combined with attention mechanisms and sliding windows, is employed to enhance feature extraction. This approach captures more refined textual features and addresses the limitations of conventional B-GRU methods, which often miss critical information [123].
3) **Sentence Scoring:**
To identify the significance of sentences, score them using the feature vectors derived from local information [31]. Moreover, advanced scoring techniques integrate inter-sentence relationships through textual entailment which can further refine sentence importance [124].

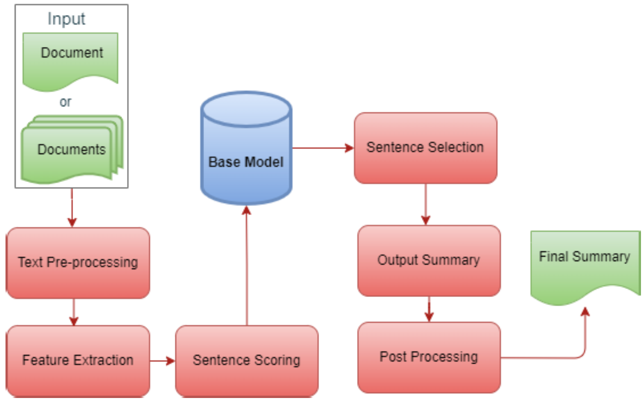


FIGURE 6. The Framework of an extractive text summarization system.

4) **Use of Base model:**
A base model can be utilized to comprehend the text as well as structure, and further validate or redefine the scoring. [27]. For example, a Restricted Boltzmann Machine (RBM) is utilized to abstract and refine the features, thereby enhancing accuracy while retaining essential information [82].
5) **Sentence Selection and output summary:**
The next step involves selecting the most significant sentences on the basis of their scores [49] and generating the primary summary from the selected sentences. Selection can be performed using methods such as integer linear programming or greedy algorithms, which include the highest-scoring sentences while maintaining coherence and eliminating redundancy [124].
6) **Post Processing:**
After generating the summary, post-processing guarantees that the final output is coherent, grammatically accurate, and conforms to the desired format. This step involves refining the output summary by removing redundancy to improve readability and coherence [63].

B. LAYERED ARCHITECTURE OF ETS

Figure 7, a layered architecture, depicts the detailed steps of a summarization system, employing a ML method to demonstrate the use of the generic architecture described above in the figure 6. The process starts with text preprocessing, which can involve various techniques based on the task's specific requirements. In this example, we're employing stemming, tokenization, stopwords removal, and sentence segmentation [16].

Next, the feature extraction process involves determining relevant features depending on the system's requirements. In this scenario, TF-ISF, position, length, proper nouns, numerical, and sentence-sentence similarity features are selected, resulting in the creation of feature vectors [81]. These vectors are subsequently utilized to score the sentences based on their significance to the summarizing task.

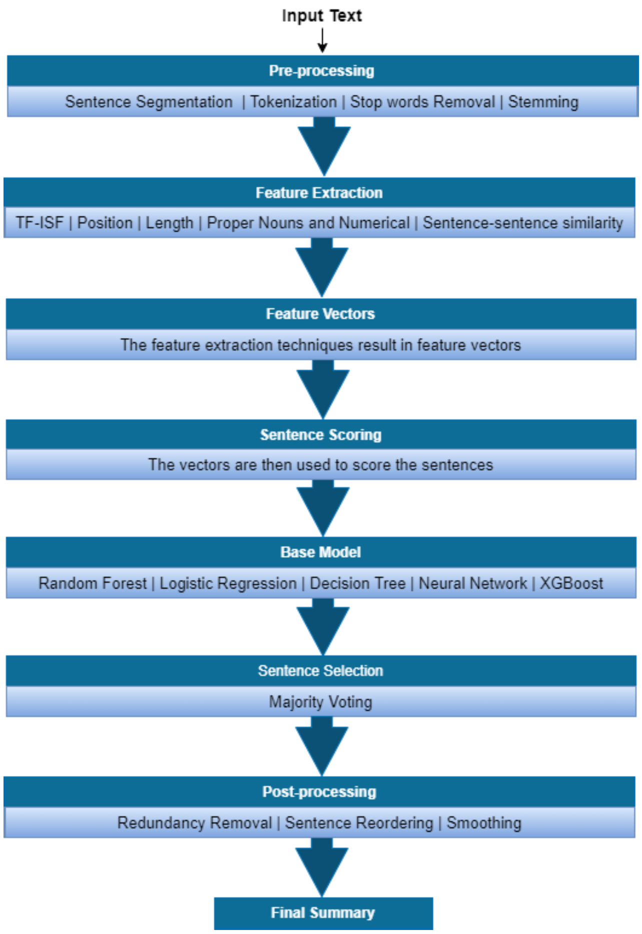


FIGURE 7. Layered architecture of Machine-Learning based summarization system.

The scored sentences are then fed into a base model. In this case, an ensemble technique is used, combining various models to improve the summarization's accuracy and robustness [16]. The following step is sentence selection, which involves selecting the most relevant sentences using a voting method. Finally, the selected sentences are subjected to post-processing, which involves techniques such as redundancy removal and sentence ordering to improve the result and produce a coherent and concise summary.

VII. EVALUATION OF EXTRACTIVE TEXT SUMMARIZATION APPROACHES

This section outlines the key resources utilized for evaluating the ETS systems, including standard datasets and evaluation criteria.

A. DATASETS

El-Kassas et al. offer a comprehensive overview of various corpora employed in summarization tasks [1]. Table 2 summarises the most frequently used benchmarking datasets for evaluating ETS systems. These include:

1) DUC Datasets: The National Institute of Standards and Technology (NIST) provides some of the most widely used datasets for text summarization research. Released during the DUC conferences from 2001 to 2007, these datasets include documents paired with three types of summaries: manually created, automatically generated baselines, and participant-generated summaries from the challenges. These datasets are widely used in the field of extractive summarization system [48], [49] [51], [53] [54]. Access to DUC datasets requires completing application forms available on the DUC website 1.
2) TAC Datasets: The DUC summarization track transitioned to being part of the TAC in 2008. These datasets are frequently used for extractive summarization tasks [72], [125]. To access these datasets, one must fill out application forms available on the website of TAC 2.
3) CNN/Daily Mail [126]: This dataset has been frequently used to evaluate the ETS system [27], [61], [65], [81].
4) CNN-corpus Dataset [127]: This dataset is suitable for single document summarization since it includes texts, summaries, and highlights [27], [128]. Researchers can freely obtain the entire annotated corpus by contacting the authors.
5) PubMed Dataset [129]: It includes biological and medical research publications from the PubMed database. It is typically used for extractive summarization tasks [81], [95], [99], [130].
6) arXiv Dataset [129]: It contains scientific publications from numerous fields such as computer science, physics, and mathematics etc. It is suitable for single document summarization task [94], [99].
7) BBC News Dataset [131]: It contains news articles in a variety of areas, including sports, business, technology, politics, and entertainment. This dataset is used for single and multiple documents summarization [19], [27], [57], [70].
8) Opinosis Dataset [132]: It has 51 files, where each file focuses on a specific product feature, and contains customer reviews about that feature. It represents 51 distinct topics, where each file includes around 100 sentences and features 5 gold summaries per topic which are written manually. It is used to evaluate extractive summarization system [61], [133].
9) EASC Dataset [134]: It includes Arabic articles along with their human-generated extractive summaries. EASC uses content that is protected by copyright.
10) LCSTS Dataset [135]: It includes about two million texts and summaries. LCSTS dataset is derived from the Chinese microblogging platform SinaWeibo.

Table 2 outlines several characteristics for all datasets. 1) Dataset (name), 2) the language in which the data is written, 3) the data domain, 4) its size (total count of documents), and 5) whether it supports multi-document and/or single-document summarization. For multi-document datasets, the document count is represented as "60 × 10,"

TABLE 2. Overview of standard datasets for extractive text summarization.

Dataset	Language	Discipline	size	Single/Multi-document
DUC 2001	English	News	60 x 10	Both
DUC 2002	English	News	60 x 10	Both
DUC 2003	English, Arabic	News	60 x 10, 30 x 25	Both
DUC 2004	English	News	100 x 10	Both
DUC 2005	English	News	50 x 32	Multi
DUC 2006	English	News	50 x 25	Multi
DUC 2007	English	News	25 x 10	Multi
TAC 2008	English	News	48 x 20	Multi
TAC 2009	English	News	44 x 20	Multi
TAC 2010	English	News	44 x 20	Multi
TAC 2011	English	News	44 x 20	Multi
CNN/Daily Mail	English	News	312,084	Single
CNN-corpus	English	News	3,000	Single
PubMed	English	Science	278,000	Single
arXiv	English	Science	194,000	Single
BBC News	English	News	2,225	Single
Opinosis	English	Reviews	51 x 100	Multi
EASC	Arabic	News, Wikipedia	153	Single
LCSTS	Chinese	Blogs	2,400,591	Single

which shows 60 clusters with approximately 10 documents each.

B. EVALUATION METRICS

There are two primary methods for evaluating the summaries: 1) extrinsic methods assess summary quality based on performance in specific tasks, such as reading comprehension or information retrieval. 2) intrinsic methods use human evaluation to measure the quality of a summary which focuses on content coverage and coherence. There exist two approaches for the evaluation of summaries: automatic and manual.

1) MANUAL EVALUATION

Human judges manually evaluate the summaries based on a number of quality metrics such as readability, non-redundancy, conciseness, referential clarity, grammaticality, content coverage, structure, and coherence [136]. Manual analysis and evaluation are highly time-consuming and require people to read both the source documents and the summaries [7]. The lightweight pyramids technique is semi-automatic which uses standard gold summary to assess summarization systems alongside automated metrics [137].

2) AUTOMATIC EVALUATION

In this subsection, we will discuss some of the traditional metrics such as ROUGE, and some modern metrics such as BERTScore.

a: ROUGE METRIC

ROUGE is the most widely utilized tool for automatically evaluating generated summaries. It works by comparing machine-generated summaries with multiple summaries written by humans [136]. It measures the amount of overlapping units, like n-grams between the reference texts and the candidate summaries [138]. ROUGE primarily emphasizes recall, but it can also account for F-measure and precision, depending on the variant employed.

- ROUGE-1 evaluates the unigrams between a reference summary and a candidate summary.
- ROUGE-2 evaluates the overlap of bigrams.
- ROUGE-L focuses on the longest matching sequences between the candidate and reference summaries.
- ROUGE-S assesses the skip-bigram overlap ratio between the two summaries.

b: BLEU METRIC

Bilingual Evaluation Understudy (BLEU) score is commonly applied for machine translation but it can be employed in measuring (n-gram) precision for summarization tasks. It assesses the overlapped n-grams (n-word sequences) between the reference text and the generated summary, emphasizing precision [139]. It evaluates the proportion of n-grams in the generated summary that match those in the reference summary, including a brevity penalty to overcome the issue of excessively short summaries [140]

c: G-EVAL

It is an advanced evaluation metric developed to measure the quality of summaries, emphasizing contextual and semantic relevance. Generative Evaluation (G-Eval) captures the quality, relevance, coherence, and fluency of the generated summary. Recently, G-Eval has been improved with a framework that utilizes GPT-4 and chain-of-thought reasoning to assess natural language generation (NLG) outputs, including summarization [141]. This technique considerably improves the alignment with human evaluations, establishing G-Eval as a reliable tool for evaluating the semantic precision and contextual alignment of extractive summaries.

TABLE 3. Comparison of evaluation metrics for extractive text summarization.

Metric	Explanation	Application	Advantages	Disadvantages
ROUGE (1, 2, L)	Evaluates the overlapping between n-grams (e.g. unigrams, bigrams, etc.) of system-generated summaries and source summaries.	Standard metric for extractive text summarization. It is suitable for factual and brief summaries	Easily computable and captures fundamental n-gram overlaps	Neglects semantic meaning
BLEU	Evaluates the overlapping between n-grams of system-generated summaries and source summaries. It emphasizes precision (i.e., the percentage of the system summary included in the reference).	Frequently employed in machine translation, and also suitable for summarization.	It is based on precision and identifies incomplete sentence matches.	Imposes penalties for excessive content generation. It fails to manage paraphrasing effectively.
G-Eval	Assesses the standard of summaries using human-like judgments. Considering factors including fluency, coherence, informativeness, and relevance.	Assesses summaries from a human-centric, comprehensive viewpoint.	Includes several quality measurements: fluency, coherence, and informativeness. It outperforms n-grams-based metrics.	Complexity related to implementation and probable subjectivity in evaluation.
BERTScore	Employs BERT embeddings to assess the semantic similarities between system-generated summaries and reference summaries.	suitable for extractive text summarization as well as abstractive text summarization	Gathers contextual meaning and assesses summaries using a deep understanding.	Computationally expensive owing to the requirement for pre-trained models.

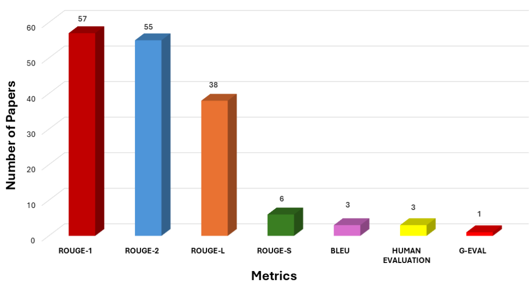


FIGURE 8. Frequency of evaluation metrics across reviewed studies.

d: BERTSCORE

BERTScore is a contemporary evaluation metric for text generation tasks, such as summarization, that utilizes contextual embeddings from pre-trained language models [142]. BScore measures the slight overlap of contextual BERT token embeddings between the reference and candidate summary [142]. It excels in evaluating the relevance of extracted sentences to the reference summary based on their semantic meaning rather than lexical overlap [143]. It offers a more thorough evaluation of summary quality by comparing tokens through cosine similarity in the embedding space.

The analysis of the reviewed papers demonstrates a diverse use of evaluation metrics within the research field, as given in Figure 8. ROUGE-1 is the most frequently used, appearing in 57 papers, indicating that it is widely accepted and useful in evaluating outcomes. ROUGE-2 is utilized 55 times, demonstrating its importance in many papers. ROUGE-L and ROUGE-S are applied 38 and 6 times, respectively, indicating their significance in evaluations. BLEU and manual human evaluation are employed in 3 papers each, with G-Eval appearing in one. Table 3 presents a comparison of evaluation metrics:

VIII. OPEN RESEARCH ISSUES AND CHALLENGES

In this section, we address the issues and challenges in the field of ETS, highlighting potential domains for further research in the future.

There are various challenges with using an ETS system, including:

A. SCALING SUMMARIZATION FOR COMPLEX TASKS

Existing systems often serve specific applications, such as blog content, online reviews, and news articles. Now more work is needed for more complex applications, such as summarizing books, and lengthy texts [144].

B. IMPROVING MULTIFORMAT AND MULTILINGUAL DATA EXTRACTION

The key challenge involves effectively summarizing content from various semi-structured and textual sources, such as web pages and databases, in the appropriate language, size, and format for specific users. The large volume of data is available in multiple languages and formats, so there is a need for increased research on multimedia, multi-document, and multilingual summarization [1].

C. CHALLENGES IN SUMMARIZING MULTIPLE DOCUMENTS

Another challenging task is multi-document summarization with several issues, including sentence reordering, redundancy, co-reference resolution, and temporal aspects [9]. One particular issue is the potential for incorrect references, where one sentence might include a proper noun, while the next sentence uses a pronoun referring to it. If these references are not managed correctly, it can lead to inaccuracies in the generated summary.

D. ENHANCING THE EXTRACTION OF SEMANTIC FEATURES

It is necessary to identify advanced linguistic features and statistical features for sentences as well as words that enable the semantic extraction of the significant sentences within the source document [9].

E. EXPANDING SUMMARIZATION BEYOND TEXTUAL INPUTS

Most summarization systems focus on text-based inputs and outputs. There is a need for new summarization tools that can handle inputs like audios, videos, and meetings and produce outputs in non-textual formats [1].

F. REFINING STOPPING CRITERIA

Humans use an iterative approach to summarize documents, evaluating the extent to which continue or stop after generating an initial summary. There is a significant need to develop an advanced method for determining when to conclude the summarization process [145].

G. ADVANCING EVALUATION METRICS

There are many challenges in evaluating summaries, whether they be done manually or automatically. Identifying an ideal or correct summary is complex, as machine-generated summaries can be high-quality but they can still differ from human-created ones [7]. So, new solutions and methods for automatically evaluating summaries are needed.

IX. CONCLUSION

This review discusses the various techniques involved in ETS, highlighting its coherence, less redundancy, and rich informational content. While ETS research has evolved significantly over the years, much remains to be explored as text summarization is considered to be difficult task. With more and more text being generated on different social networks and news sites, the field has expanded beyond scientific articles to encompass a broader range of text types, demanding new approaches in text summarization. Key features like keywords, similarity, frequency, semantics, sentence position, and length play crucial roles in generating effective summaries. Statistical methods often support other techniques, such as fuzzy-based or ML approaches. Future research should focus on complex applications, multi-document, multimedia, and multilingual summarization. Another key aspect is to consider that the accuracy and acceptance of the summarized text need to be improved. Identifying advanced linguistic and statistical features for semantic extraction is thus an essential feat, along with developing methods to determine optimal conclusions and improve automatic evaluation. In this study we have highlighted the following significant insights:

1) This manuscript provides a thorough review and analysis of diverse methodologies and techniques employed in ETS.
2) Although summarization research has been around for a while, more work remains to be done. The emphasis has changed from summarizing scientific articles to encompassing blogs, advertisements, news articles, and emails.
3) Keywords, similarity, frequency, semantics, sentence position, and sentence length are all important features for generating a practical summary.
4) It is possible to combine fuzzy-based or ML approaches with statistical methods.
5) Statistical approaches are often applied togther with other techniques, such as keyword identification, similarity assessment, and frequency determination.
6) The challenges and future directions include the need for research on complex applications and multi-document, multimedia, and multilingual summarization.
7) Identifying advanced linguistic and statistical features for effective semantic extraction is essential.
8) Developing advanced methods to determine the optimal conclusion of summarization and new approaches for automatic evaluation are also necessary.

This review paper is aimed at establishing the basis for further study of key issues and potential developments in this area.

LIST OF ABBREVIATIONS

ATS	Automatic Text Summarization.1-3,7
BERT	Bidirectional Encoder Representations from Transformers.2,5,7-9,11,12
BLEU	Bilingual Evaluation Understudy.11
DL	Deep Learning.2,6,7
DLMNN	Deep Learning Modifier Neural Network.2,8
ETS	Extractive Text Summarization.2,3,5-7,9,10,12,13
G-Eval	Generative Evaluation.11
GA	Genetic Algorithms.3
GP	Genetic Programming.3
GPT	Generative Pre-trained Transformer.7,11
LSTM	Long Short-Term Memory.7,8
ML	Machine Learning.2,3,6,9,13
NLP	Natural Language Processing.2,3,7
NN	Neural Network.3,6
RNNs	Recurrent Neural Networks.5,6
ROUGE	Recall-Oriented Understudy for Gisting Evaluation.8,9,11

REFERENCES

[1] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Syst. Appl.*, vol. 165, Mar. 2021, Art. no. 113679.
[2] G. C. V. Vilca and M. A. S. Cabezudo, "A study of abstractive summarization using semantic representations and discourse level information," in *Proc. 20th Int. Conf. Text, Speech, Dialogue (TSD)*, Prague, Czech Republic, Jan. 2017, pp. 482–490.
[3] M. Azmi-Murad and T. P. Martin, "Similarity-based estimation for document summarization using fuzzy sets," *Int. J. Comput. Sci. Secur.*, vol. 7, pp. 1–12, Jan. 2007.

[4] J.-G. Yao, X. Wan, and J. Xiao, "Recent advances in document summarization," *Knowl. Inf. Syst.*, vol. 53, no. 2, pp. 297–336, Nov. 2017.

[5] A. P. Widyassari, S. Rustad, G. F. Shidik, E. Noersasongko, A. Syukur, A. Affand, and D. R. I. M. Setiadi, "Review of automatic text summarization techniques & methods," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1029–1046, 2022.

[6] L. Abualigah, M. Q. Bashabsheh, H. Alabool, and M. Shehab, "Text summarization: A brief review," *Recent Adv. NLP, Case Arabic Lang.*, vol. 2019, pp. 1–15, Nov. 2019.

[7] N. Moratanch and S. Chitrakala, "A survey on extractive text summarization," in *Proc. Int. Conf. Comput., Commun. Signal Process. (ICCCSP)*, Jan. 2017, pp. 1–6.

[8] H. P. Luhn, "The automatic creation of literature abstracts," *IBM J. Res. Develop.*, vol. 2, no. 2, pp. 159–165, Apr. 1958.

[9] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: A survey," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 1–66, Jan. 2017.

[10] M. Kirmani, N. Hakak, M. Mohd, and M. Mohd, "Hybrid text summarization: A survey," in *Soft Computing: Theories Applications: Proceedings of SoCTA*. Cham, Switzerland: Springer, 2018, pp. 63–73.

[11] S. Dutta, V. Chandra, K. Mehra, S. Ghatak, A. K. Das, and S. Ghosh, "Summarizing microblogs during emergency events: A comparison of extractive summarization algorithms," in *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS*, vol. 2. Cham, Switzerland: Springer, 2019, pp. 859–872.

[12] A. Kumar and A. Sharma, "Systematic literature review of fuzzy logic based text summarization," *Iranian J. Fuzzy Syst.*, vol. 16, no. 5, pp. 45–59, Oct. 2019.

[13] M. A. Mosa, A. S. Anwar, and A. Hamouda, "A survey of multiple types of text summarization with their satellite contents based on swarm intelligence optimization algorithms," *Knowl.-Based Syst.*, vol. 163, pp. 518–532, Jan. 2019.

[14] D. Suleiman and A. A. Awajan, "Deep learning based extractive text summarization: Approaches, datasets and evaluation measures," in *Proc. 6th Int. Conf. Social Netw. Anal., Manage. Secur. (SNAMS)*, Oct. 2019, pp. 204–210.

[15] P. Bhattacharya, K. Hiware, S. Rajgaria, N. Pochhi, K. Ghosh, and S. Ghosh, "A comparative study of summarization algorithms applied to legal case judgments," in *Proc. Adv. Inf. Retr., 41st Eur. Conf. IR Res. (ECIR)*, Cologne, Germany. Cham, Switzerland: Springer, Jan. 2019, pp. 413–428.

[16] P. Singh, P. Chhikara, and J. Singh, "An ensemble approach for extractive text summarization," in *Proc. Int. Conf. Emerg. Trends Inf. Technol. Eng. (ic-ETITE)*, Feb. 2020, pp. 1–7.

[17] H. Gupta and M. Patel, "Study of extractive text summarizer using the elmo embedding," in *Proc. 4th Int. Conf. I-SMAC (IoT Social, Mobile, Analytics Cloud) (I-SMAC)*, Oct. 2020, pp. 829–834.

[18] R. Elbarougy, G. Behery, and A. El Khatib, "Extractive Arabic text summarization using modified pagerank algorithm," *Egyptian Informat. J.*, vol. 21, no. 2, pp. 73–81, Jul. 2020.

[19] M. M. Haider, Md. A. Hossin, H. R. Mahi, and H. Arif, "Automatic text summarization using gensim Word2 Vec and K-means clustering algorithm," in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Jun. 2020, pp. 283–286.

[20] S. Jugran, A. Kumar, B. S. Tyagi, and V. Anand, "Extractive automatic text summarization using SpaCy in Python & NLP," in *Proc. Int. Conf. Advance Comput. Innov. Technol. Eng. (ICACITE)*, Mar. 2021, pp. 582–585.

[21] B. Muthu, S. Cb, P. M. Kumar, S. N. Kadry, C.-H. Hsu, O. Sanjuan, and R. G. Crespo, "A framework for extractive text summarization based on deep learning modified neural network classifier," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 3, pp. 1–20, May 2021.

[22] D. Aljevic, L. Todorovski, and S. Martincic-Ipšic, "Extractive text summarization based on selectivity ranking," in *Proc. Int. Conf. Innov. Intell. Syst. Appl. (INISTA)*, Aug. 2021, pp. 1–6.

[23] A. K. Yadav, A. Singh, M. Dhiman, Vineet, R. Kaundal, A. Verma, and D. Yadav, "Extractive text summarization using deep learning approach," *Int. J. Inf. Technol.*, vol. 14, no. 5, pp. 2407–2415, Aug. 2022.

[24] R. Srivastava, P. Singh, K. P. S. Rana, and V. Kumar, "A topic modeled unsupervised approach to single document extractive text summarization," *Knowl.-Based Syst.*, vol. 246, Jun. 2022, Art. no. 108636.

[25] P. Verma, A. Verma, and S. Pal, "An approach for extractive text summarization using fuzzy evolutionary and clustering algorithms," *Appl. Soft Comput.*, vol. 120, May 2022, Art. no. 108670.

[26] 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," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–14, Nov. 2022.

[27] P. Gupta, S. Nigam, and R. Singh, "A ranking based language model for automatic extractive text summarization," in *Proc. 1st Int. Conf. Artif. Intell. Trends Pattern Recognit. (ICAITPR)*, Mar. 2022, pp. 1–5.

[28] A. Joshi, E. Fidalgo, E. Alegre, and L. Fernández-Robles, "DeepSumm: Exploiting topic models and sequence to sequence networks for extractive text summarization," *Expert Syst. Appl.*, vol. 211, Jan. 2023, Art. no. 118442.

[29] Y. A. Al-Khassawneh and E. S. Hanandeh, "Extractive Arabic text summarization-graph-based approach," *Electronics*, vol. 12, no. 2, p. 437, Jan. 2023.

[30] K. Thirumoorthy and J. J. J. Britto, "A hybrid approach for text summarization using social mimic optimization algorithm," *Iranian J. Sci. Technol., Trans. Electr. Eng.*, vol. 47, no. 2, pp. 677–693, Jun. 2023.

[31] A. Ghadimi and H. Beigy, "SGCSumm: An extractive multi-document summarization method based on pre-trained language model, submodularity, and graph convolutional neural networks," *Expert Syst. Appl.*, vol. 215, Apr. 2023, Art. no. 119308.

[32] S.-N. Vo, T.-T. Vo, and B. Le, "Interpretable extractive text summarization with meta-learning and BI-LSTM: A study of meta learning and explainability techniques," *Expert Syst. Appl.*, vol. 245, Jul. 2024, Art. no. 123045.

[33] A. K. Yadav, Ranvijay, R. S. Yadav, and A. K. Maurya, "Graph-based extractive text summarization based on single document," *Multimedia Tools Appl.*, vol. 83, no. 7, pp. 18987–19013, Jul. 2023.

[34] G. Sharma and D. Sharma, "Automatic text summarization methods: A comprehensive review," *Social Netw. Comput. Sci.*, vol. 4, no. 1, p. 33, Oct. 2022.

[35] N. Rahman and B. Borah, "Query-based extractive text summarization using sense-oriented semantic relatedness measure," *Arabian J. Sci. Eng.*, vol. 49, no. 3, pp. 3751–3792, Mar. 2024.

[36] H. P. Edmundson, "New methods in automatic extracting," *J. ACM*, vol. 16, no. 2, pp. 264–285, Apr. 1969.

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

[38] C. Aone, M. E. Okurowski, J. Gorlinsky, and B. Larsen, "A scalable summarization system using robust NLP," in *Proc. Intell. Scalable Text Summarization*, Jul. 1997. [Online]. Available: https://www.aclweb.org/anthology/W97-0711

[39] I. Mani and E. Bloedorn, "Machine learning of generic and user-focused summarization," in *Proc. AAAI/IAAI*, Jan. 1998, pp. 821–826.

[40] Y. Gong and X. Liu, "Generic text summarization using relevance measure and latent semantic analysis," in *Proc. 24th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Sep. 2001, pp. 19–25.

[41] K. Kaikhah, "Automatic text summarization with neural networks," in *Proc. 2nd Int. IEEE Conf. Intell. Syst.*, Oct. 2004, pp. 40–44.

[42] A. Kiani-B and M. R. Akbarzadeh-T, "Automatic text summarization using hybrid fuzzy GA-GP," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jun. 2006, pp. 977–983.

[43] S. Ryang and T. Abekawa, "Framework of automatic text summarization using reinforcement learning," in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn.*, Jul. 2012, pp. 256–265.

[44] E. Egonmwan and Y. Chali, "Transformer-based model for single documents neural summarization," in *Proc. 3rd Workshop Neural Gener. Transl.*, 2019, pp. 70–79.

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

[46] J. N. Madhuri and R. Ganesh Kumar, "Extractive text summarization using sentence ranking," in *Proc. Int. Conf. Data Sci. Commun. (IconDSC)*, Mar. 2019, pp. 1–3.

[47] K. U. Manjari, S. Rousha, D. Sumanth, and J. S. Devi, "Extractive text summarization from Web pages using selenium and TF-IDF algorithm," in *Proc. 4th Int. Conf. Trends Electron. Informat. (ICOEI)(4)*, Jun. 2020, pp. 648–652.

[48] R. Rani and D. K. Lobiyal, "A weighted word embedding based approach for extractive text summarization," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115867.

[49] S. S. Naik and M. N. Gaonkar, "Extractive text summarization by feature-based sentence extraction using rule-based concept," in *Proc. 2nd IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, May 2017, pp. 1364–1368.
[50] F. B. Goularte, S. M. Nassar, R. Fileto, and H. Saggion, "A text summarization method based on fuzzy rules and applicable to automated assessment," *Expert Syst. Appl.*, vol. 115, pp. 264–275, Jan. 2019.
[51] C. Fang, D. Mu, Z. Deng, and Z. Wu, "Word-sentence co-ranking for automatic extractive text summarization," *Expert Syst. Appl.*, vol. 72, pp. 189–195, Apr. 2017.
[52] A. Tixier, P. Meladianos, and M. Vazirgiannis, "Combining graph degeneracy and submodularity for unsupervised extractive summarization," in *Proc. Workshop New Frontiers Summarization*, 2017, pp. 48–58.
[53] V. Ravinuthala, S. Chinnam, and A. University, "A keyword extraction approach for single document extractive summarization based on topic centrality," *Int. J. Intell. Eng. Syst.*, vol. 10, no. 5, pp. 153–161, Oct. 2017.
[54] M. Yasunaga, R. Zhang, K. Meelu, A. Pareek, K. Srinivasan, and D. Radev, "Graph-based neural multi-document summarization," 2017, *arXiv:1706.06681*.
[55] G. D. L. P. Sarracén and P. Rosso, "Automatic text summarization based on betweenness centrality," in *Proc. ACM Int. Conf. Ser.*, 2018, pp. 1–4, doi: 10.1145.
[56] V. Dalal and L. Malik, "Semantic graph based automatic text summarization for Hindi documents using particle swarm optimization," in *Information and Communication Technology for Intelligent Systems (ICTIS 2017)*. Cham, Switzerland: Springer, 2018, pp. 284–289.
[57] S. Ullah and A. B. M. A. A. Islam, "A framework for extractive text summarization using semantic graph based approach," in *Proc. 6th Int. Conf. Netw., Syst. Secur.*, Dec. 2019, pp. 48–56.
[58] M. R. Alfarra, A. M. Alfarra, and J. M. Alattar, "Graph-based fuzzy logic for extractive text summarization (GFLES)," in *Proc. Int. Conf. Promising Electron. Technol. (ICPET)*, Oct. 2019, pp. 96–101.
[59] H. Van Lierde and T. W. S. Chow, "Query-oriented text summarization based on hypergraph transversals," *Inf. Process. Manage.*, vol. 56, no. 4, pp. 1317–1338, Jul. 2019.
[60] C. Mallick, A. K. Das, M. Dutta, A. K. Das, and A. Sarkar, "Graph-based text summarization using modified TextRank," in *Proc. Soft Comput. Data Anal., Proc. Int. Conf. (SCDA)*. Cham, Switzerland: Springer, Aug. 2018, pp. 137–146.
[61] R. C. Belwal, S. Rai, and A. Gupta, "A new graph-based extractive text summarization using keywords or topic modeling," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 10, pp. 8975–8990, Oct. 2021.
[62] B. Jing, Z. You, T. Yang, W. Fan, and H. Tong, "Multiplex graph neural network for extractive text summarization," 2021, *arXiv:2108.12870*.
[63] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "EdgeSumm: Graph-based framework for automatic text summarization," *Inf. Process. Manage.*, vol. 57, no. 6, Nov. 2020, Art. no. 102264.
[64] T. Uçkan and A. Karcı, "Extractive multi-document text summarization based on graph independent sets," *Egyptian Informat. J.*, vol. 21, no. 3, pp. 145–157, Sep. 2020.
[65] J. Liu, D. J. D. Hughes, and Y. Yang, "Unsupervised extractive text summarization with distance-augmented sentence graphs," in *Proc. 44th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2021, pp. 2313–2317.
[66] Q. Mao, H. Zhu, J. Liu, C. Ji, H. Peng, J. Li, L. Wang, and Z. Wang, "MuchSUM: Multi-channel graph neural network for extractive summarization," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2022, pp. 2617–2622.
[67] A. Sharaff, M. Jain, and G. Modugula, "Feature based cluster ranking approach for single document summarization," *Int. J. Inf. Technol.*, vol. 14, no. 4, pp. 2057–2065, Jun. 2022.
[68] R. C. Belwal, S. Rai, and A. Gupta, "Extractive text summarization using clustering-based topic modeling," *Soft Comput.*, vol. 27, no. 7, pp. 3965–3982, Apr. 2023.
[69] A. T. Al-Taani and S. H. Al-Sayadi, "Extractive text summarization of Arabic multi-document using fuzzy C-means and latent Dirichlet allocation," *Int. J. Syst. Assurance Eng. Manage.*, vol. 15, no. 2, pp. 713–726, Feb. 2024.
[70] A. Sharaff, A. S. Khaire, and D. Sharma, "Analysing fuzzy based approach for extractive text summarization," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS)*, May 2019, pp. 906–910.
[71] D. Patel, S. Shah, and H. Chhinkaniwala, "Fuzzy logic based multi document summarization with improved sentence scoring and redundancy removal technique," *Expert Syst. Appl.*, vol. 134, pp. 167–177, Nov. 2019.
[72] M. Tomer, M. Kumar, A. Hashmi, B. Sharma, and U. Tomer, "Enhancing metaheuristic based extractive text summarization with fuzzy logic," *Neural Comput. Appl.*, vol. 35, no. 13, pp. 9711–9723, May 2023.
[73] A. Bharadwaj, A. Srinivasan, A. Kasi, and B. Das, "Extending the performance of extractive text summarization by ensemble techniques," in *Proc. 11th Int. Conf. Adv. Comput. (ICoAC)*, Dec. 2019, pp. 282–288.
[74] J. M. Sanchez-Gomez, M. A. Vega-Rodríguez, and C. J. Pérez, "A decomposition-based multi-objective optimization approach for extractive multi-document text summarization," *Appl. Soft Comput.*, vol. 91, Jun. 2020, Art. no. 106231.
[75] M. Mojrian and S. A. Mirroshandel, "A novel extractive multi-document text summarization system using quantum-inspired genetic algorithm: MTSQIGA," *Expert Syst. Appl.*, vol. 171, Jun. 2021, Art. no. 114555.
[76] K. Chettah and A. Draa, "A discrete differential evolution algorithm for extractive text summarization," in *Proc. Int. Conf. Innov. Intell. Syst. Appl. (INISTA)*, Aug. 2021, pp. 1–6.
[77] A. K. Srivastava, D. Pandey, and A. Agarwal, "Extractive multi-document text summarization using dolphin swarm optimization approach," *Multimedia Tools Appl.*, vol. 80, no. 7, pp. 11273–11290, Mar. 2021.
[78] M. Tomer and M. Kumar, "Multi-document extractive text summarization based on firefly algorithm," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6057–6065, Sep. 2022.
[79] A. Jain, D. Bhatia, and M. K. Thakur, "Extractive text summarization using word vector embedding," in *Proc. Int. Conf. Mach. Learn. Data Sci. (MLDS)*, Dec. 2017, pp. 51–55.
[80] D. Krishnan, P. Bharathy, Anagha, and M. Venugopalan, "A supervised approach for extractive text summarization using minimal robust features," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS)*, May 2019, pp. 521–527.
[81] V. C. da Silva, J. P. Papa, and K. A. P. da Costa, "Extractive text summarization using generalized additive models with interactions for sentence selection," 2022, *arXiv:2212.10707*.
[82] S. Verma and V. Nidhi, "Extractive summarization using deep learning," 2017, *arXiv:1708.04439*.
[83] S. Narayan, N. Papasarantopoulos, S. B. Cohen, and M. Lapata, "Neural extractive summarization with side information," 2017, *arXiv:1704.04530*.
[84] A. K. Singh, M. Gupta, and V. Varma, "Hybrid MemNet for extractive summarization," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 2303–2306.
[85] A. Sinha, A. Yadav, and A. Gahlot, "Extractive text summarization using neural networks," 2018, *arXiv:1802.10137*.
[86] C. Feng, F. Cai, H. Chen, and M. de Rijke, "Attentive encoder-based extractive text summarization," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 1499–1502.
[87] A. Joshi, E. Fidalgo, E. Alegre, and L. Fernández-Robles, "SummCoder: An unsupervised framework for extractive text summarization based on deep auto-encoders," *Expert Syst. Appl.*, vol. 129, pp. 200–215, Sep. 2019.
[88] J. Xu and G. Durrett, "Neural extractive text summarization with syntactic compression," 2019, *arXiv:1902.00863*.
[89] A. Rezaei, S. Dami, and P. Daneshjoo, "Multi-document extractive text summarization via deep learning approach," in *Proc. 5th Conf. Knowl. Based Eng. Innov. (KBEI)*, Feb. 2019, pp. 680–685.
[90] Y. Diao, H. Lin, L. Yang, X. Fan, Y. Chu, D. Wu, D. Zhang, and K. Xu, "CRHASum: Extractive text summarization with contextualized-representation hierarchical-attention summarization network," *Neural Comput. Appl.*, vol. 32, no. 15, pp. 11491–11503, Aug. 2020.
[91] T. Ma, Q. Pan, H. Rong, Y. Qian, Y. Tian, and N. Al-Nabhan, "T-BERTSum: Topic-aware text summarization based on BERT," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 3, pp. 879–890, Jun. 2022.
[92] H. Gupta and M. Patel, "Method of text summarization using lsa and sentence based topic modelling with bert," in *Proc. Int. Conf. Artif. Intell. Smart Syst. (ICAIS)*, Mar. 2021, pp. 511–517.
[93] M. Jang and P. Kang, "Learning-free unsupervised extractive summarization model," *IEEE Access*, vol. 9, pp. 14358–14368, 2021.
[94] Q. Ruan, M. Ostendorff, and G. Rehm, "HiStruct+: Improving extractive text summarization with hierarchical structure information," 2022, *arXiv:2203.09629*.
[95] Q. Xie, J. A. Bishop, P. Tiwari, and S. Ananiadou, "Pre-trained language models with domain knowledge for biomedical extractive summarization," *Knowl.-Based Syst.*, vol. 252, Sep. 2022, Art. no. 109460.

[96] T. S. Barros, C. E. S. Pires, and D. C. Nascimento, "Leveraging BERT for extractive text summarization on federal police documents," *Knowl. Inf. Syst.*, vol. 65, no. 11, pp. 4873–4903, Nov. 2023.
[97] S. R. Chowdhury and K. Sarkar, "A new method for extractive text summarization using neural networks," *Social Netw. Comput. Sci.*, vol. 4, no. 4, p. 384, May 2023.
[98] J. Fan, X. Tian, C. Lv, S. Zhang, Y. Wang, and J. Zhang, "Extractive social media text summarization based on MFMMR-BertSum," *Array*, vol. 20, Dec. 2023, Art. no. 100322.
[99] S. Bano, S. Khalid, N. M. Tairan, H. Shah, and H. A. Khattak, "Summarization of scholarly articles using BERT and BiGRU: Deep learning-based extractive approach," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 9, Oct. 2023, Art. no. 101739.
[100] T. Vo, "An approach of syntactical text graph representation learning for extractive summarization," *Int. J. Intell. Robot. Appl.*, vol. 7, no. 1, pp. 190–204, Mar. 2023.
[101] J. Huang, "A novel extractive summarization method based on multi-path feature learning," in *Proc. 5th Int. Conf. Artif. Intell. Electromechanical Autom. (AIEA)*, Jun. 2024, pp. 448–452.
[102] M. H. H. Wahab, N. A. W. A. Hamid, S. Subramaniam, R. Latip, and M. Othman, "Decomposition–based multi-objective differential evolution for extractive multi-document automatic text summarization," *Appl. Soft Comput.*, vol. 151, Oct. 2024, Art. no. 110994.
[103] A. Onan and H. A. Alhumyani, "DeepExtract: Semantic-driven extractive text summarization framework using LLMs and hierarchical positional encoding," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 36, no. 8, Oct. 2024, Art. no. 102178.
[104] J. Zhang, L. Lu, L. Zhang, Y. Chen, and W. Liu, "DCDSum: An interpretable extractive summarization framework based on contrastive learning method," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108148.
[105] H. Li, J. Zhu, C. Ma, J. Zhang, and C. Zong, "Multi-modal summarization for asynchronous collection of text, image, audio and video," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 1092–1102.
[106] R. K. Roul and K. Arora, "A nifty review to text summarization-based recommendation system for electronic products," *Soft Comput.*, vol. 23, no. 24, pp. 13183–13204, Dec. 2019.
[107] M. H. Sarkhoosh, S. Gautam, C. Midoglu, S. S. Sabet, and P. Halvorsen, "Multimodal AI-based summarization and storytelling for soccer on social media," in *Proc. ACM Multimedia Syst. Conf. ZZZ*, Apr. 2024, pp. 485–491.
[108] P. P. S. Bedi, M. Bala, and K. Sharma, "Extractive text summarization for biomedical transcripts using deep dense LSTM-CNN framework," *Expert Syst.*, vol. 41, no. 7, p. 13490, Jul. 2024.
[109] Y. Yang, Y. Tan, J. Min, and Z. Huang, "Automatic text summarization for government news reports based on multiple features," *J. Supercomput.*, vol. 80, no. 3, pp. 3212–3228, Feb. 2024.
[110] L. Bacco, A. Cimino, F. Dell'Orletta, and M. Merone, "Explainable sentiment analysis: A hierarchical transformer-based extractive summarization approach," *Electronics*, vol. 10, no. 18, p. 2195, Sep. 2021.
[111] G. Kaur and A. Sharma, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis," *J. Big Data*, vol. 10, no. 1, p. 5, Jan. 2023.
[112] N. Datta, G. C. Shekar, K. S. Teja, P. G. S. Abhinay, and M. K. S. Prasad, "Extractive text summarization of clinical text using deep learning models," in *Proc. 2nd Int. Conf. Emerg. Trends Inf. Technol. Eng. (ICETITE)*, Feb. 2024, pp. 1–6.
[113] A. Sarker, Y.-C. Yang, M. A. Al-Garadi, and A. Abbas, "A light-weight text summarization system for fast access to medical evidence," *Frontiers Digit. Health*, vol. 2, Dec. 2020, Art. no. 585559.
[114] S. Bano and S. Khalid, "BERT-based extractive text summarization of scholarly articles: A novel architecture," in *Proc. Int. Conf. Artif. Intell. Things (ICAIoT)*, Dec. 2022, pp. 1–5.
[115] S. R. K. Harinatha, B. T. Tasara, and N. N. Qomariyah, "Evaluating extractive summarization techniques on news articles," in *Proc. Int. Seminar Intell. Technol. Its Appl. (ISITIA)*, Jul. 2021, pp. 88–94.
[116] P. Sethi, S. Sonawane, S. Khanwalker, and R. B. Keskar, "Automatic text summarization of news articles," in *Proc. Int. Conf. Big Data, IoT Data Sci. (BID)*, Dec. 2017, pp. 23–29.
[117] W. Waseemullah, Z. Fatima, S. Zardari, M. Fahim, M. A. Siddiqui, A. A. A. Ibrahim, K. Nisar, and L. F. Naz, "A novel approach for semantic extractive text summarization," *Appl. Sci.*, vol. 12, no. 9, p. 4479, Apr. 2022.
[118] I. Akhmetov, A. Gelbukh, and R. Mussabayev, "Greedy optimization method for extractive summarization of scientific articles," *IEEE Access*, vol. 9, pp. 168141–168153, 2021.
[119] Z. Kerui, H. Haichao, and L. Yuxia, "Automatic text summarization on social media," in *Proc. 4th Int. Symp. Comput. Sci. Intell. Control*, Nov. 2020, pp. 1–5.
[120] A. Tabassum and R. R. Patil, "A survey on text pre-processing & feature extraction techniques in natural language processing," *Int. Res. J. Eng. Technol.*, vol. 7, no. 6, pp. 4864–4867, 2020.
[121] K. Juluru, H.-H. Shih, K. N. K. Murthy, and P. Elnajjar, "Bag-of-Words technique in natural language processing: A primer for radiologists," *RadioGraphics*, vol. 41, no. 5, pp. 1420–1426, Sep. 2021.
[122] N. Babanejad, A. Agrawal, A. An, and M. Papagelis, "A comprehensive analysis of preprocessing for word representation learning in affective tasks," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 5799–5810.
[123] R. RoselinKiruba, S. Sowmyayani, S. Anitha, J. Kavitha, R. Preethi, and C. S. Jothi, "Text summarization based on feature extraction using GloVe and B-GRU," in *Proc. 2nd Int. Conf. Sustain. Comput. Smart Syst. (ICSCSS)*, Jul. 2024, pp. 517–522.
[124] D. H. Long, M.-T. Nguyen, N. X. Bach, L.-M. Nguyen, and T. M. Phuong, "An entailment-based scoring method for content selection in document summarization," in *Proc. 9th Int. Symp. Inf. Commun. Technol. - SoICT*, 2018, pp. 122–129.
[125] Á. Hernández-Castañeda, R. A. García-Hernández, Y. Ledeneva, and C. E. Millán-Hernández, "Extractive automatic text summarization based on lexical-semantic keywords," *IEEE Access*, vol. 8, pp. 49896–49907, 2020.
[126] K. M. Hermann, T. Kočiský, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend," in *Proc. Adv. Neural Inf. Process. Syst.*, Jun. 2015, pp. 1–11.
[127] R. D. Lins, H. Oliveira, L. Cabral, J. Batista, B. Tenorio, R. Ferreira, R. Lima, G. de França Pereira e Silva, and S. J. Simske, "The CNN-corpus: A large textual corpus for single-document extractive summarization," in *Proc. ACM Symp. Document Eng.*, Sep. 2019, pp. 1–10.
[128] K. Shetty and J. S. Kallimani, "Automatic extractive text summarization using K-means clustering," in *Proc. Int. Conf. Electr., Electron., Commun., Comput., Optim. Techn. (ICEECCOT)*, Dec. 2017, pp. 1–9.
[129] A. Cohan, F. Dernoncourt, D. Soon Kim, T. Bui, S. Kim, W. Chang, and N. Goharian, "A discourse-aware attention model for abstractive summarization of long documents," 2018, *arXiv:1804.05685*.
[130] H. Zhang, X. Liu, and J. Zhang, "Extractive summarization via ChatGPT for faithful summary generation," 2023, *arXiv:2304.04193*.
[131] J. Wang, H. Zhang, C. Zhang, W. Yang, L. Shao, and J. Wang, "Generating an overview report over many documents," 2019, *arXiv:1908.06216*.
[132] N. H. Tien, N. M. Le, and N. M. Le, "Opinions summarization: Aspect similarity recognition relaxes the constraint of predefined aspects," in *Proc. Int. Conf. Recent Adv. Natural Lang. Process. (RANLP)*, 2019, pp. 487–496.
[133] R. C. Belwal, S. Rai, and A. Gupta, "Text summarization using topic-based vector space model and semantic measure," *Inf. Process. Manage.*, vol. 58, no. 3, May 2021, Art. no. 102536.
[134] A. Alshanqiti, A. Namoun, A. Alsughayyir, A. M. Mashraqi, A. R. Gilal, and S. S. Albouq, "Leveraging DistilBERT for summarizing Arabic text: An extractive dual-stage approach," *IEEE Access*, vol. 9, pp. 135594–135607, 2021.
[135] Z. Li, Z. Peng, S. Tang, C. Zhang, and H. Ma, "Text summarization method based on double attention pointer network," *IEEE Access*, vol. 8, pp. 11279–11288, 2020.
[136] E. Lloret, L. Plaza, and A. Aker, "The challenging task of summary evaluation: An overview," *Lang. Resour. Eval.*, vol. 52, no. 1, pp. 101–148, Mar. 2018.
[137] O. Shapira, D. Gabay, Y. Gao, H. Ronen, R. Pasunuru, M. Bansal, Y. Amsterdamer, and I. Dagan, "Crowdsourcing lightweight pyramids for manual summary evaluation," 2019, *arXiv:1904.05929*.
[138] S. Wang, X. Zhao, B. Li, B. Ge, and D. Tang, "Integrating extractive and abstractive models for long text summarization," in *Proc. IEEE Int. Congr. Big Data (BigData Congr.)*, Jun. 2017, pp. 305–312.
[139] A. Ibrahim, M. Alfonse, and M. Aref, "A systematic review on text summarization of medical research articles," *Int. J. Intell. Comput. Inf. Sci.*, vol. 23, no. 2, pp. 50–61, Jun. 2023.

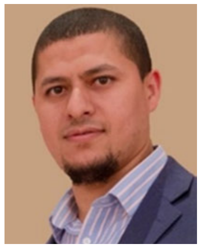
[140] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics*, Philadelphia, PA, USA, 2002, pp. 311–318.
[141] Y. Liu, D. Iter, Y. Xu, S. Wang, R. Xu, and C. Zhu, "G-Eval: NLG evaluation using GPT-4 with better human alignment," 2023, *arXiv:2303.16634*.
[142] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating text generation with BERT," 2019, *arXiv:1904.09675*.
[143] M. Hanna and O. Bojar, "A fine-grained analysis of bertscore," in *Proc. 6th Conf. Mach. Transl.*, 2021, pp. 507–517.
[144] Z. Wu, L. Lei, G. Li, H. Huang, C. Zheng, E. Chen, and G. Xu, "A topic modeling based approach to novel document automatic summarization," *Expert Syst. Appl.*, vol. 84, pp. 12–23, Oct. 2017.
[145] R. Belkebir and A. Guessoum, "TALAA-ATSF: A global operation-based Arabic text summarization framework," *Intell. Natural Lang. Process., Trends Appl.*, vol. 2017, pp. 435–459, Nov. 2017.

MARYAM AZAM is currently pursuing the M.S. degree in data science with the National University of Sciences and Technology (NUST), with a focus on large language models. She is also a Researcher and a Writer with expertise in natural language processing (NLP) and machine learning. She has completed several projects in machine learning and deep learning. As a writer, she shares insights on data analytics through articles designed to help beginners navigate the complexities of data analysis. Looking forward, she plans to further her research in artificial intelligence and contribute to advancements in NLP technology.

SHAH KHALID received the M.S. degree from the University of Peshawar, Pakistan, and the Ph.D. degree from Jiangsu University, China. He is currently an Assistant Professor with the School of Electrical Engineering and Computer Science (SEECS), National University of Science and Technology (NUST), Islamabad, Pakistan. He has been involved in numerous research projects in Pakistan and abroad. His research interests include information retrieval, web search engines, scholarly retrieval systems, recommender systems, knowledge graphs, social web, real-time sentiment analysis, web engineering, text summarization, federated searches, and digital libraries. For more information, please visit his website at https://sites.google.com/view/shahkhalid.

SULAIMAN ALMUTAIRI received the bachelor's degree in computer science from Bluefield State University, USA, the master's degree in information technology from RMU, USA, and the Ph.D. degree in biomedical informatics from Rutgers University, NJ, USA. He has been a Tenure with Qassim University, since 2005. He is currently a Distinguished Data Governance Consultant and the Director of quality scholarship services with the Ministry of Education. He also holds the position of an Assistant Professor of health informatics with Qassim University. He has held various pivotal roles, including the Vice Dean of planning, development, and quality. He has showcased his expertise as a Data Governance Consultant with prominent tech firms in Saudi Arabia, such as Elm and SITE. Additionally, he served as a Cultural Mission Diplomat representing Saudi Arabia in Canada and contributed globally as a Data Consultant with Development Gateway and UNICEF. His research interests include data governance, bio-informatics, information and policy, and software development.

HASAN ALI KHATTAK received the Ph.D. degree in electrical and computer engineering degree from Politecnico di Bari, Bari, Italy, in April 2015. He has been an Associate Professor with Sohar University, Oman, and an Associate Professor with the School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan, since October 2020. Along with publishing in good research venues and completing successfully funded national and international projects, he also serves as a reviewer in reputed venues. His perspective research areas are in the application of machine learning and data sciences for improving and enhancing quality of life in smart urban spaces especially in the healthcare and transportation domain through knowledge management, predictive analysis, and visualization. His current research interests include future internet architectures, such as the Web of Things and leveraging data sciences and social engineering for future smart cities.



ABDALLAH NAMOUN (Member, IEEE) received the bachelor's degree in computer science, in 2004, and the Ph.D. degree in informatics from The University of Manchester, U.K., in 2009. He is currently a Full Professor of intelligent interactive systems and the Director of the AI Center, Islamic University of Madinah. He has authored more than 100 peer-reviewed publications in research areas spanning intelligent systems, machine learning, human–computer interaction, smart cities, and multi-agent systems. He has extensive experience in leading complex research projects (worth more than 23 million Euros) with several distinguished SMEs, such as SAP, BT, and ATOS. His recent research interests include integrating state-of-the-art artificial intelligence models in the development of interactive systems and smart spaces.

AMJAD ALI received the M.S. degree in computer science from the University of Peshawar, Pakistan, in 2007, and the Ph.D. degree in engineering (computer science) from the Department of Informatics, Gyeongsang National University, South Korea, in 2016. He is currently an Assistant Professor with the Department of Computer and Software Technologies, University of Swat, Khyber Pakhtunkhwa, Pakistan. He has been actively involved in academia and research, contributing to the advancement of his field. His research interest includes real-time systems. He serves as a reviewer for several journals and conferences.

HAFIZ SYED MUHAMMAD BILAL received the M.S. degree from the National University of Sciences and Technology (NUST), Pakistan, and the Ph.D. degree in computer science and engineering from Kyung Hee University. With over 15 years of experience in teaching and research, he has made significant contributions to the computing field, authoring nearly 19 journal publications and over 20 conference papers. Currently, he is an Assistant Professor with the Department of Computing, School of Electrical Engineering and Computer Science (SEECS), NUST. His research interests include health informatics, artificial intelligence, and data science. He continues to play a vital role in advancing academic and scientific excellence in his domain.