www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Recent Developments and Emerging Trends in ETS: An In-Depth Overview

[1] Harika Lakshmi Bachineni, [2] Dr. S. Ramarao,

[1,] PG Student, Department of CSE, MAM Women's Engineering College, Kesanupalli, Narasaraopet, Palnadu District.

[2] Professor, Department of CSE, MAM Women's Engineering College, Kesanupalli, Narasaraopet, Palnadu District.

ABSTRACT

The fast growth of textual data across industries has made text summarization a must-have for effective data management. There have been a lot of books and review articles written to summarize the many approaches that have been suggested to improve summarization procedures in the last several decades. In particular, when it comes to detailed architectural frameworks, the present status of the subject, assessment procedures, and unsolved problems, existing evaluations often fall short of providing a thorough retrospective of recent advances. This work fills that need by providing an in-depth evaluation of the extraction methods, including all of their advantages, disadvantages, and underlying processes. Our comprehensive, multi-tiered architectural framework is intended to help researchers with their attempts by advancing and developing summarization models. Text preprocessing, feature extraction, sentence scoring, base model use, sentence selection, output summary, and post-processing are the primary components of the text summarizing system. Also, this analysis of 145 studies classifies methods of summary according to their respective domains, with an eye on the particular difficulties of and solutions for news, scientific, and social media summaries. Statistics, fuzzy logic, rules, optimization, graphs, clustering, machine learning, and deep learning are all part of this set of approaches. We provide a comprehensive review of the most popular datasets and metrics used in performance assessment, focusing on ROUGE-1, ROUGE-2, ROUGE-L, and ROUGE-S, and we highlight the importance of evaluation metrics and benchmark datasets. By outlining potential avenues for further study and current obstacles, this review paper serves as a great tool for enhancing text summarizing methods within the fields of machine learning and natural language processing. Notable issues include enhancing assessment metrics, refining stopping criteria, dealing with multi-format and multilingual data, adding more documents to the summarizing process, and dealing with multi-modal user input.

INDEX TERMS

Evaluation metrics, generic architecture, datasets, domain-specific summarization, survey, text summarization, and transformer-based models.

I. INTRODUCTION

Data is growing at an unprecedented rate in the present day, mostly due to human activity on websites, blogs, social media, and news platforms [1]. But this rise has created a problem: When individuals look for data, They often uncover an overwhelming amount of data, which makes it difficult to get useful outcomes. Condensing the findings to provide individuals with the data they need to resolve this issue. On the other hand, manually summing the massive amounts of data is still another following problem [2].

FIGURE 1. Model architecture of an automatic text summarizer.

The problem of human summarizing has motivated researchers to develop automated text summary methods. These methods generate summaries by extracting relevant information from the text using keywords, all the while keeping the original meaning of the sentences intact [3]. In Figure 1 we can see the ATS system's fundamental architecture.
A number of methods are included into ATS. One way to classify the source papers is as a text summary of a single document or many documents [4]. Extractive and abstractive summarization are the two main types of summary output [5]. In Natural Language Processing (NLP), extractive summarization is a key component. The main purpose of it is to summarize long texts into short ones [6]. A sentence's significance is conditional on its linguistic and statistical properties [7].

When it comes to natural language processing, ATS is up there with the biggest dogs. Since 1958, researchers have been actively engaged in the field of ATS [8]. A more effective ATS system is still a goal of the research community. Improvements to the ATS system have continued to propel the industry forward [9]. Kirmani et al. [10] presented a number of extraction techniques and important statistical features in 2018. Dutta et al. [11] devoted their attention to researching extractive strategies used for microblog summarization in detail. With the use of fuzzy logic approaches, Kumar et al. [12] present a thorough evaluation of extractive ATS systems. A 2019 survey by Mosa et al. [13] investigated how ATS may benefit from Swarm Intelligence Optimization techniques. Applying Deep Learning (DL) methods to Extractive Text Summarization (ETS) was the subject of a research conducted by Suleiman et al. [14]. Bhattacharya et al.'s 2019 study [15] zeroed in on a specific subset of summarization, namely that which is employed in legal texts. A number of Machine Learning (ML) models were used to create an ensemble method for ETS by Singh et al. in 2020 [16]. An extracted summary was generated in 2020 by Gupta et al. [17] using the Elmo embedding approach. The summary was derived using a Modified PageRank approach in a research conducted by Elbarougy et al. [18]. By integrating Gensim's Word2Vec technique with K-means clustering, Haider et al. [19] presented a strategy for ETS in 2020. A method for ETS that gives many algorithm options for a given job was proposed by Jugran et al. [20]. A document classifier based on Deep Learning Modifier Neural Networks (DLMNNs) was suggested by Muthu et al. in 2021 [21]. For ETS, Aljevic et al. came up with a novel graph-based method [22]. Using a basic model based on reinforcement learning, Yadav et al. began their study in 2022 [23]. Srivastava et al. [24] suggested a method for unsupervised ETS that uses topic modeling and clustering to reduce topic bias. Clustering, evolutionary algorithms, and fuzzy logic are all parts of the hybrid strategy that Verma et al. suggested for ETS [25]. According to this research [26], Umair et al. provide a neural model known as N-GPETS, which integrates a BERT model with a graph attention model. Gupta et al. presented a sentence-raking-based approach to ETS in 2022 [27].

Joshi et al. introduced the DeepSumm method for ETS in 2023 [28]. An integrated extractive summary was generated by Khassawneh et al. [29] using a textual graph approach. For the purpose of extractive summarization, Thirumoorthy et al. [30] put out a social mimic hybrid optimization approach. For the purpose of extractive summarizing of multi-documents, Ghadimi et al. introduced SGCSumm (Sub-modular Graph Convolutional Summarizer) in 2023 [31]. Through the use of explanatory approaches such as input modification, SHAP, decision trees, and linear regression, Vo et al. (2024) were able to understand the decision-making process of a DL model that had been developed using a meta-learning methodology [32]. One graph-based approach to summarizing that Yadav et al. introduced is the TGETS model [33]. This model takes the average graph weight and adds it to the weights of individual sentences to generate summaries. Several techniques for extractive and abstract summarization are detailed in the review article by Adhika et al. [5] up till 2020. A review of the state of the art in extractive summarization up till 2016 was provided by Moratanch et al. [7]. In their review article, Sharma et al. [34] compare and contrast several approaches of abstractive and extractive summarization. Similarly, Kassas et al. [1] only covers research up to 2020 in their review study, leaving out any groundbreaking methods or advancements achieved beyond that year. To fill this void and cover all the bases, including the latest trends, problems, and advancements in the field of abstract summarization, a timely and all-encompassing study is required. However, our review paper's focus is on approaches and advancements in extractive summarization up to 2024. Our work stands out because to this distinctive feature, which allows us to provide a thorough and current evaluation of ETS. In order to assist scholars understand the trends, difficulties, and advances in the area of abstract summarization, our study provides them with useful information on current research and breakthroughs in the field [35]. The following are some of the contributions of this manuscript: a thorough investigation of extractive summarizing

• The categorization of methods for extractive text summary. • The history and present status of extractive text summarization.
• Uses of extractive text summarization. • A framework for efficiently extracting summary text.Here you will find information about: • Evaluation criteria used to measure summary quality; • Current benchmark data sets for extractive text summarizing research; and • Open research concerns and difficulties related to extractive text summarization.

The remaining portions of the paper are organized as follows: The methodology for the survey and a discussion of the chosen articles are presented in Section II. In Section III, we'll look at how extractive text summarization has developed and where the field is now. In Section IV, we see how several methods for extracting text summarization are organized. The results of using the system for extractive text summarization are detailed in Section V. An architecture for text summarizing that is both generic and layered is shown in Section VI. Part VII details the benchmark data sets used for text summarizing and elucidates the criteria used to evaluate the process. In Section VIII, we'll talk about the problems and questions that still need answering in extractive text summarization. Section IX concludes the work by reviewing and expanding upon the key ideas and arguments offered throughout. See Figure 2 for a visual representation of the paper's structure.

II. SURVEY METHODOLOGY AND SELECTED PAPERS

Even though previous comprehensive review articles provide useful background on extractive text summarization, they don't go into great detail about recent advances in the area. In order to provide a complete and current picture, our method in this research involves reviewing all the publications published between 2017 and 2024. Although not all articles published before 2017 are included, only the most notable ones that are pertinent to the issue are chosen. Using terms like "Extractive text summarization," "extractive summarization techniques," and "automatic text summarization," we were able to locate almost all of the publications via our searches on Google Scholar,1 semantic scholar2, and dblp3. In addition, we looked into the reference lists of chosen publications to broaden our search. We thoroughly checked each article after downloading to make sure it was related to the subject. Part A. Selected Articles Approximately 145 articles have been identified as meeting our unique criteria throughout the paper selection process. Thus, in order to shed light on the most recent cutting-edge methods, this research will examine these 145 papers that are pertinent to extractive text summarization. The detailed process for selecting articles is shown in Figure 3.

III. THE EVOLUTION AND CURRENT STATE OF EXTRACTIVE TEXT SUMMARIZATION

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

The evolution of ETS field may be traced back to early algorithms. These first-generation algorithms debuted at different times, and each one has its own unique strategy for the purpose of summary. Using a frequency-based algorithm, Luhn et al. presented their method for text summary that finds important sentences by counting their occurrences and giving them structural and content-based ratings [8]. In addition to laying the groundwork for a frequency-based algorithm, this research offered an extraction-based method for selecting key phrases and assembling them into summaries.

The rule-based algorithm was a novel approach introduced by H.P. Edmundson in the late 1960s [36]. This method took structural indications, pragmatic terms, title and headers, and important words and phrases into account. When it came to improving the quality of the extracts, these three suggested components were more important than the frequency component. Mani et al. [37] presented a new approach to multi-document text summarizing in the late 90s using a graph-based algorithm. Using a graphical representation of the text, this method compared and contrasted two linked publications. After that, in the late 90s, Aone et al. presented yet another NLP-based summary system. Information technology, corpus-driven statistical NLP, and internet resources were among the powerful natural language processing technologies used by this system [38]. By incorporating characteristics from various state-of-the-art methods, it aimed to overcome the shortcomings of conventional knowledge-based, frequency-based, and discourse-based summarizing approaches. In the late 90s, Mani et al. [39] presented a method to generate both generic and user-specific summaries using ML from corpora. This method was effective and laid forth clear guidelines. These rules combined user-specific keywords with precise geographical data from the general rules to target personalized interests. Because it does away with the need for human tagging, this approach has seen extensive adoption. Y Gong et al. [40] introduced two methods for text summarization in the early 2000s: rating the phrases in the documents and choosing them. To begin, we ranked the significance of phrases using information retrieval algorithms. On the other hand, semantic latent analysis was used by the alternative technique to identify the crucial phrases for the summary. Selecting unique and very important phrases is the goal of both approaches. This allows for the creation of a summary that is more comprehensive and has less or no repetition of the document's content. A novel method for text summarization based on neural networks was presented by K Khaikhah in the mid-2000s [41]. In this method, a Neural Network (NN) is trained to determine which phrases should be included in the summary. Then, the NN is trained to recognize and integrate the key elements from the phrases that make up the summary. In the end, this tweaked NN filters out unnecessary details to provide brief summaries.
The hybrid approach to ETS was introduced by Arman et al. in the middle of the 2000s [42]. This fresh method integrates Genetic

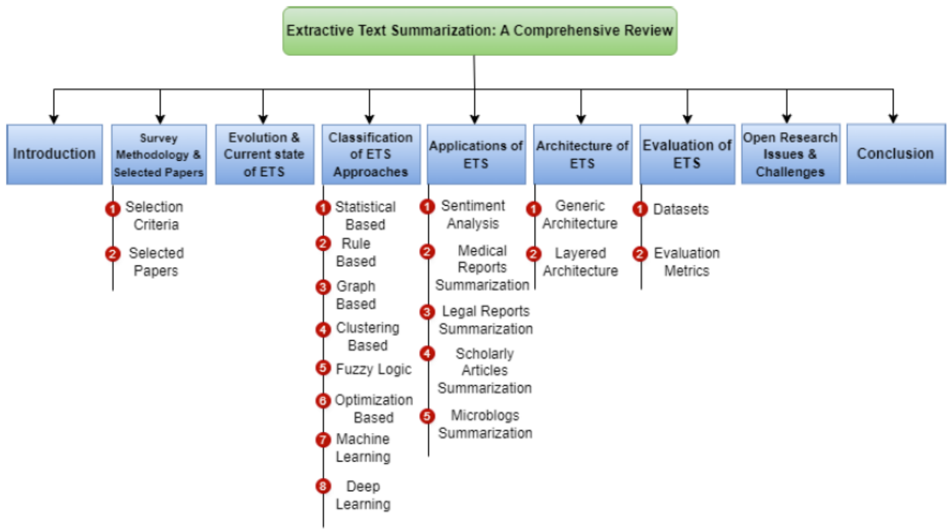

FIGURE 2. Paper Structure: Main content of the paper.

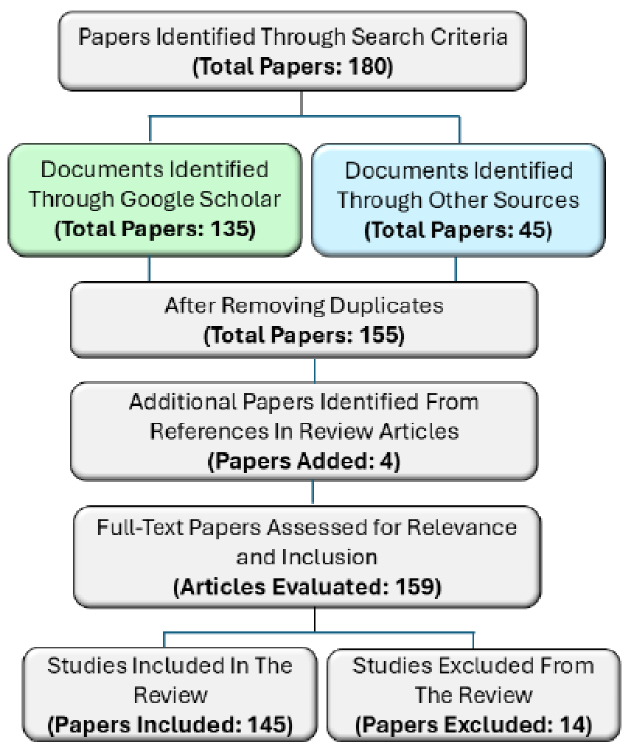

FIGURE 3. The flowchart depicts the comprehensive selection process, from the initial identification of articles to final inclusion for detailed analysis.

Gestural Programming (GP) and Algorithms (GA) to enhance rule sets and membership functions in fuzzy systems. This method relies on GA to manage the string part and the GP handles the structural aspect. A summary-creation approach based on reinforcement learning was presented by Ryang et al. in the mid-2010s [43]. When using the particular scoring formula may be improved based on the summary's individual feature representation. The results demonstrate that ATS issues may be effectively addressed using the reinforcement learning method. Egonmwan introduced a transformer-based approach to ETS in the latter half of the 2010s [44]. The framework applies a sequence-to-sequence model after encoding the source text using a transformer. Based on their research, it is clear that combining the sequence-to-sequence model with the transformer yields a more accurate encoded vector representation.

Late in the 2010s, Liu et al. showed how BERT could be used for text summarization and provided a framework that could be used for both extractive and abstract models [45]. They unveiled a document-level encoder that uses BERT to capture the semantics of a text and generate sentence representations. Using many inter-sentence Transformer layers, they constructed an extractive model. The development of ETS methods over the years is seen in Figure 4.

IV. CLASSIFICATION OF EXTRACTIVE TEXT SUMMARIZATION APPROACHES

A number of approaches are available for ETS. Figure 5 shows an example of one of these strategies. There is also a quick rundown of these strategies here.

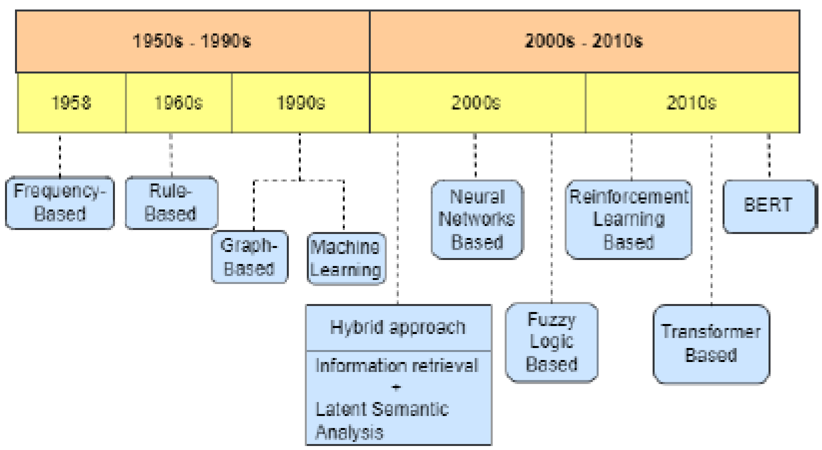

FIGURE 4. Evolution of extractive text summarization techniques.

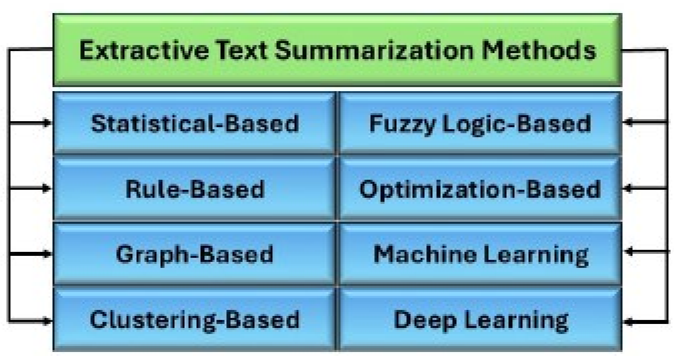

FIGURE 5. Different methods of extractive text summarization.

Table A. Methods Based on Statistics These methods extract meaningful phrases from texts by using statistical models. The document determines the weighted frequency of each word, and the sentences are ranked using these weights. The final summary is then formed by selecting the sentences with the greatest ratings [46]. For ETS, the TF-IDF algorithm is used [47]. The ranking of sentences is determined by calculating the TF-IDF score of each word. This TF-IDF score is used to determine the total sentence score. The result is that the summary contains all the sentences that scored higher than a certain threshold. In addition to linguistic and statistical aspects, the TF-IDF algorithm may also extract semantic information from the input material [48]. To generate extractive summaries from a collection of texts, a sentence ranking method is used [27]. B. Methods Based on Rules Algorithms that rely on rules and guidelines to extract the most relevant and crucial information from a text document are known as rule-based algorithms for extractive summarization. Various predefined criteria, such as the relative importance of keywords, the location of the phrase inside the title, the length of the sentence, the degree to which sentences are similar to each other, numerical value, and proper nouns, are used by these algorithms to rank and score sentences [49]. To reduce dimensionality, a fuzzy rule-based approach analyses associated characteristics; this improves expert systems for automated text assessment and enables effective text summarization [50].

Section C: Graphical Methods The ETS makes use of graphs that are based on sentences to depict texts. When it comes to ETS, the CoRank model combines a graph-based approach with word-to-sentence correlations [51]. Extractive summaries are generated by a system using a submodularity framework and a greedy approach [52]. A lexical connection is provided, which allows for the creation of a graph-representation for the purpose of extracting theme-conveying words. These words are then used to identify significant sentences within the text document [53]. Graph Convolutional Networks and Sentiment Relation Graphs are used in the neural summarization of many texts. To extract word embeddings, greedy algorithms are employed, and Recurrent Neural Networks (RNNs) are included [54]. In an unsupervised approach, the similarity or dissimilarity of words is used to define the edges in a weighted graph, which represents sentences as vertices [55]. A semantic graph is constructed by extracting triples such as subject-object-verb. A classifier is trained using PSO to create a sub-graph and then a summary is generated [56]. Alternatively, you may use the following methods: maximum marginal relevance, semantic distance measurement, PAS extraction, PageRank algorithm, and selecting top-n sentences [57]. Using Fuzzy logic for both single and numerous texts, an unsupervised graph-based technique may be used [58]. To find the themes within a corpus, a new method employs topic models based on semantic word clustering and hypergraph traversals [59]. Sparse graph partitioning with weighted edges is an additional technique that makes use of a modified TextRank algorithm [60]. One method, which makes use of topic modeling and semantic measuring, takes into account not only the degree to which sentences are similar but also their link to the text as a whole [61]. The Multiplex-Graph Convolutional Network (Multi-GCN) method is the foundation of the Multiplex-Graph Summarization model (Multi-GraS) [62], which models many sentence- and word-related interactions concurrently. A new unsupervised graph-based technique called "EdgeSumm" combines approaches based on statistics, graphs, centrality, and semantics [63]. To maintain semantic consistency, another graph-based approach employs a text processing tool known as "KUSH" and maximal independent sets [64]. By analyzing distances and similarities, an unsupervised graph-based method may choose significant statements [65]. A system exists that takes text and turns it into a sentence graph. It then uses selectivity measures to identify significant nodes and uses Cosine similarity, Jaccard, and Mihalcea's metrics to decide the edges [22]. Using three graph channels—one for each node's location and centrality, one for each node's textual attributes with Bipartite graphs integrating words and sentences—MuchSUM, a multiple channels graph convolutional network, incorporates several important characteristics [66]. A suggested WSD (word sense disambiguation) strategy finds relevant sentences by determining the actual meaning of words, and a feature-based approach computes semantic similarity; TGETS is a query-based summarizing method that uses this metric.

similarity ratings, with redundant data removed to provide a concise and useful summary [33]. D. Methods Based on Clustering Extractive text summarizing approaches based on clustering include grouping related phrases into

clusters and then selecting important sentences from each cluster to generate the summary. In this paper, we provide an extractive summarization model for a single document that can accomplish the following: extract relevant features, score sentences using similarity measures, group sentences into clusters, and then combine the best phrases from each cluster to create a summary [67]. Another method of document clustering exists, which uses topic modeling and semantic analysis to provide an extracted summary of the text [68]. Using Gensim word2vec for effective feature extraction of semantic subjects, there is a sentence-based clustering approach called K-means that is developed for single-document summarization [19]. Using Fuzzy clustering techniques and LDA algorithms to cluster the documents by subject and then extract essential lines is another strategy for summarizing several Arabic papers [69].
E. Methods Based on Fuzzy Logic Fuzzy-Logic, which is analogous to human thinking, is used by these methods to categorize sentence feature values as either zero or one [12]. For ETS, there is a fuzzy-logic method that takes a number of things into account to choose the best phrases to include in a summary [70]. Fuzzy logic is used in a feature-based statistical method to deal with feature weight uncertainty and imprecision, and cosine similarity is used to remove duplication [71]. A model that optimizes feature weights using a meta-heuristic function is suggested for use in fuzzy logic-based models; the summary is then constructed by taking the dot product of feature scores and these weights [72]. F. Methods Based on Optimization The summarization issue is recast as an optimization problem using these methods. Optimal ensemble techniques aim to improve summary accuracy and quality by combining the best features of many summarizing models, building on the concept of voting classifiers [73]. For ETS with several documents, we provide a multi-objective method that relies on decomposition. The MOABC/D employs an asynchronous parallel implementation that takes use of a multi-core architecture [74]. In order to improve redundancy, relevance, and coverage within a predefined length limit, MTSQIGA employs a modified quantum-inspired evolutionary algorithm to extract key words from many documents. This optimization process is based on a binary optimization problem [75]. To choose the best set of sentences to extract from text, a new extractive text summarizer uses a discrete differential evolution algorithm [76]. To optimize feature weights, a new extractive multi-document summarization approach uses Dolphin Swarm Optimization, and for content coverage and non-redundancy, it uses Modified Normalised Google Distance and Word Mover Distance [77]. For the purpose of summarizing several documents, a firefly algorithm employs a fit function that takes into account criteria such as readability, coherence, and subject relationship [78]. G. Methods Based on Machine Learning Using a dataset of texts combined with their corresponding human-created summaries, these methods convert the summarizing problem into a supervised sentence-level classification challenge. The system then utilizes these examples as training to classify every phrase as a summary or not. One way to successfully extract important features from supervised text summarization is to utilize a NN that has been trained on 10 features, such word vector embedding [79].

An alternative model that generates summaries using discriminative, resilient, and minimalistic characteristics employs a supervised technique [80]. After seeing that many ML models have failed to produce satisfactory results, the authors of this study set out to compare and contrast several different methods, including Logistic Regression, Decision Tree, Neural Network, Random Forest, SVMmodels, XGBoost, and Naive Bayes, before suggesting an ensemble approach that improves accuracy [16]. The issue of extractive summarization employing binary classification and linguistic characteristics has been addressed by using two additive models alongside interactions, namely GAMI-NET and Explainable Boosting Machine [81]. H. Methods Based on Deep Learning To find the most important phrases, these methods employ models like RNNs, Transformers, and GNNs. These features can be improved using a Restricted Boltzmann Machine and an unsupervised deep auto-encoder for feature learning. Three steps make up the process: extracting features, improving them, and finally, generating a summary [82]. We use an attention-based extractor using side information and a hierarchical document encoder [83]. The data-driven neural network learns n-gram features and sentence-level information using a combination of memory networks and multilayered bidirectional long-short-term networks [84]. Different deep learning methods have been proposed for ETS, such as 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 auto-encoders with defined sentence feature vectors [89]. Further methods include a topic-aware T-BERTSum model that uses BERT for extractive summarization [91], a Deep Learning-Modifier-Neural Network classifier that uses entropy values [21], and Elmo Embeddings, which encodes text into contextual vectors for sentence ranking [17]. Extraction of informative sentences pertaining to the issue at hand is achieved by combining BERT, sentence keyword extraction, and LSA topic modeling on a text.

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.

article's subject matter [92]. In an unsupervised method, pre-trained sentence vectors, deep document representations based on positional encoding and self-attention are used for phrase significance score 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]; Deep- Summ 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 a The SGCSumm technique utilizes a GCN for feature learning and performs many transformations to normalize, non-negatively, submodularly, and non-reducing monotonely the texts represented by the BERT pre-trained language model [31]. Combining BERT with an attention mechanism based on graph neural networks (GAT) is a new method called TGA4ExSum [100]. To extract the summary, a BERT-based summarization method processes the BERT outputs via RNN, GRU, and LSTM, and then employs several pathways for feature learning. [101] By combining Multiple Objective Differential Evolution, a weighted aggregate technique, and an enhanced ATS repair mechanism, the new methodology known as MODE/D-WS improves ATS [102]. By combining Generative Pre-trained Transformer (GPT)-4, semantic clustering, and layered positional encoding, the

DeepExtract system is able to optimize coherence, novelty, and relevance while producing contextually exact summaries [103]. To reduce repetition and improve understandability, the DCDSum architecture applies contrastive learning to convert the problem into sentence reranking [104]. A comparison of several summarizing approaches and the algorithms employed by each is shown in Table 1. By outlining the benefits and drawbacks of each method, it gives a thorough evaluation of their efficacy. Reference articles that go into depth about these methods and their uses are cited in the table as well.

V. APPLICATIONS OF EXTRACTIVE TEXT SUMMARIZATION

Analytics and text mining make use of ETS for tasks including as answering questions, extracting data, and retrieving information. Li et al. [105] offer a multi-modal extractive summarizing system that can create textual summaries from asynchronous inputs such as pictures, audio, video, and text. Domains like sentiment analysis [106], succinct summarizing of academic publications [99], and synthesis of material from microblogs [107] are noteworthy examples of where extractive summarization is useful. Aside from email and news, summarization algorithms find utility in many other sectors, such as biomedical [108] and legal papers [109]. Here are a few examples of how the summarizing system may be used: • The study of people's feelings, thoughts, and assessments of things is known as sentiment analysis. This uses extractive text in a major way. summary, whereby summary methods are used to make models more understandable so that we can become better at classifying things. Using fuzzy c-means clustering, this research proposes a method to summarize user evaluations [106]. It automatically generates summaries of reviews of electronic products by comparing sets of phrases for semantic and content similarities, taking into account the review's substance as well as the author's reliability. By utilizing attention weights within cascaded transformers or integrating sentence-level probabilities for document-level classification, transformer-based methods have been applied to sentiment analysis to extract important sentences as summaries, making the model more interpretable [110]. An approach that combines natural language processing (NLP) with long short-term memory (LSTM) has been suggested for sentiment analysis and review summarization. To get a full picture of what customers think, our model uses pre-processing, LSTM-based sentiment classification, and hybrid feature extraction that takes into account both review-related and aspect-related variables. [111] Timely prevention of infectious illnesses may be aided by extracting critical data from medical records and discussions and sharing it with patients and clinicians. The goal of training deep learning and transformer models using PubMed clinical text data was to enhance the efficiency of medical literature information retrieval [112]. Quick access to research evidence is made possible by a lightweight extractive summarization method that uses basic features and medical word embeddings. It achieves performance that is similar to state-of-the-art models [113]. For the purpose of automatically compiling medical papers from biological data, there is a summarizing strategy that merges Deep Dense LSTM with CNN [108]. Through the use of phrase selection and subject modeling, it is able to transform badly punctuated transcriptions of discussions into intelligible summaries.

The goal of the government-news summarizing method is to extract important facts from legitimate government publications. In this day of overwhelming amounts of data, it aids readers in quickly comprehending the news. Here we provide a model for text summarizing that makes use of several aspects. Extracting features from sentences is the first step in the TF-IDF approach using word vector embeddings from the BERT model [114]. Next, sentences are evaluated according to location, similarity, and keywords. The summary is then formed from the highest-ranked sentences [109]. In terms of extractive summarization of news datasets, BERT and TextRank have both been used; however, tests have shown that TextRank performs better than BERT on the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) score, recall, and F-measure, while BERT shows better precision. This highlights their respective strengths in summarization tasks [115]. Utilizing WordNet, sophisticated scoring algorithms, and pronoun resolution, lexical chain-based approaches have been used to summarize news items. These approaches effectively detect important areas of text and provide coherent and efficient summaries that are customized to the layout of news material [116].

To ensure that readers understand the key points and significance of the study without reading the whole work, scholarly publications should be succinctly summarized. A new approach uses BERT, which has been pre-trained on huge, self-supervised datasets, to produce embeddings at the phrase level. After the embeddings are processed using BiGRU, important information may be extracted and sequential relationships can be captured [99]. In order to extract useful information from Wikipedia articles, text summarizing systems use statistical techniques, feature engineering, and sophisticated preprocessing. These systems solve the problems of previous algorithms by using fuzzy logic, cosine similarity, and LDA modeling to provide accurate and efficient summaries that preserve significant material [117]. A greedy extractive method using Variable Neighborhood Search (VNS) for scientific article summarization obtains competitive ROUGE scores on PubMed and arXiv datasets [118]. This method provides a less resource-intensive substitute for complicated neural networks by selecting words with high TF-IDF values.

Twitter and Facebook alone get millions of shares per day. Countless tweets are made during crises, and Twitter delivers critical real-time information. Consequently, microblog summarization has become more important in recent times. The provided article [107] provides an example of a system that may summarize tweets. To improve the quality of the summaries and reduce duplication, the MFMMR-BertSum model has been used for sentence-level extractive summarization using BERT with an additional classification layer. On the CNN/DailyMail dataset, this method performs quite well, particularly when it comes to summarizing texts from social media [98]. The LCSTS dataset, a high-quality collection of Chinese short texts from 'Sina Weibo,' has been successfully used to train sophisticated models that include BERT, sequence-to-sequence approaches, and reinforcement learning for social media summarization. The number 119. The models' remarkable gains in ROUGE scores demonstrate how well they summarize material from social media.

VI. ARCHITECTURE FOR EXTRACTIVE TEXT SUMMARIZATION

Section A: ETS's Generic Architecture The structure of an ETS system is shown in Figure 6. It has several stages, which will be covered later on:First, text preprocessing. Cleaning and preparing the input document(s) for further analysis, including sentence processes such as tokenization, bag-of-words, stemming, segmentation, and elimination of stop words [46]. The process of separating text into individual sentences is known as sentence segmentation. This may aid with word boundary recognition, which in turn helps facilitate phrase processing. When a full stop or punctuation mark is present, a sentence tokenizer will cause segmentation to occur [120]. Tokenization is the process of breaking down phrases into smaller pieces, such as words, punctuation, and characters [120]. Tokenizing the following statement into its component parts would achieve the same goal: "NLP is reshaping search engines." Put an end to terms The most frequent words in a language that don't really imply anything are called stop words. Examples of these words include "a," "is," and "the" [121]. The number of unique terms in the dictionary may be greatly reduced by removing these words.

Words may be reduced to their essential form by removing prefixes and suffixes using stemming, a powerful preprocessing technique [122]. Shortening the words "ran," "runner," and "running" to just "run" is what stemming is all about. Words to Keep: To train a classifier, the bag-of-words method first extracts characteristics from the text. The characteristics are organized based on the frequency of word usage in the text [120]. 2) Feature Extraction: Feature extraction is the process of finding important sentences by extracting them from text [48]. Methods that are considered traditional sometimes include intricate ranking or scoring systems. Deep learning methods have been included into recent strategies to enhance feature extraction. Global Vectors for Word Representation (GloVe) is one tool that may be used to calculate word feature vectors and sentence similarity. To take it a step further, feature extraction is boosted by using the Bi-Gated Recurrent Unit in conjunction with attention mechanisms and sliding windows. This method overcomes the shortcomings of traditional B-GRU techniques, which often fail to catch important details, and captures more nuanced textual elements [123]. Thirdly, sentence scoring: use feature vectors produced from local information to assess sentences in order to detect their relevance [31]. In addition, sophisticated scoring methods may further clarify the relevance of individual sentences by including inter-sentence links via textual entailment [124].

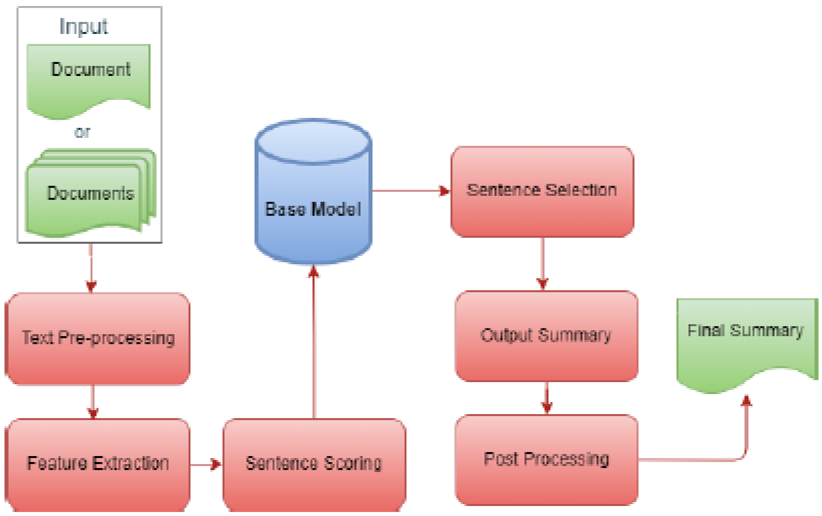

FIGURE 6. The Framework of an extractive text summarization system.

Fourthly, make use of a basic model to understand the text and structure, and then use that knowledge to verify or rethink the score. [27]. One approach is to use a Restricted Boltzmann Machine (RBM) to improve accuracy while keeping important information intact by abstracting and refining the characteristics [82]. 5) Output summary and sentence selection: After that, you'll want to generate the main summary using the phrases that were deemed most important based on their ratings [49]. Integer linear programming and greedy algorithms are two examples of selection approaches that may be used to include the highest-scoring phrases while also removing redundancy and preserving coherence [124]. 6. Processing After: This summary will be logical, grammatically correct, and formatted according to your specifications once post-processing ensures it. At this stage, we refine the output summary by eliminating unnecessary words to make it more readable and coherent [63]. B. ETS's Layered Architecture To illustrate the usage of the generic architecture mentioned in figure 6, Figure 7 shows a layered architecture that represents the detailed processes of a summarization system that employs an ML approach. The first step is text preparation, which may include a variety of approaches depending on the nature of the job. Here, we're using sentence segmentation, stemming, tokenization, and the elimination of stopwords [16]. The next step in feature extraction is to identify important characteristics based on the needs of the system. Feature vectors are generated in this case by selecting TF-ISF, location, length, proper nouns, numerical, and sentence-sentence similarity characteristics [81]. After that, we use these vectors to rank the phrases according to how important they are for the summary.

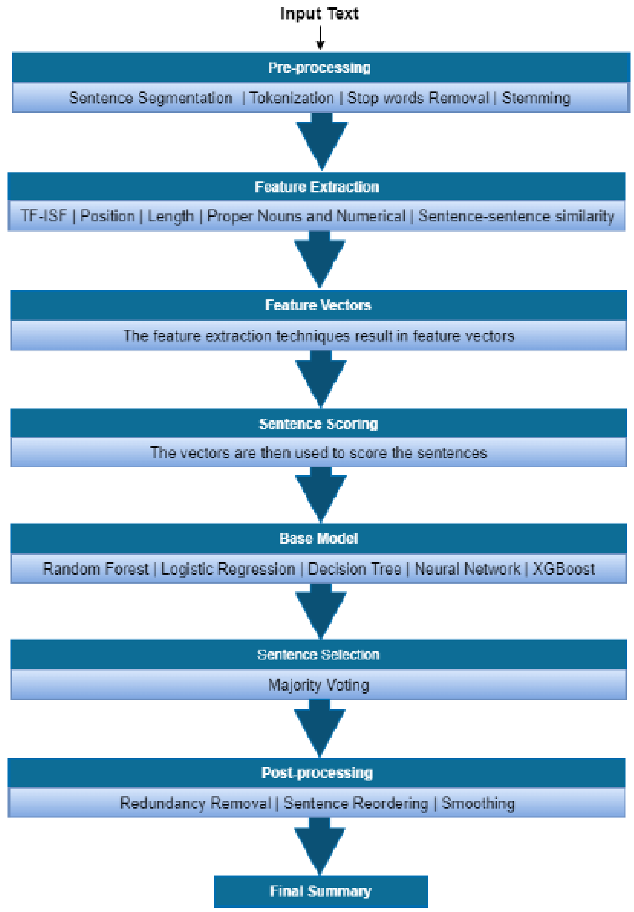

FIGURE 7. Layered architecture of Machine-Learning based summarization system.

A basic model is then given the scored phrases. Here, we use an ensemble method, which involves merging many models to enhance the reliability and precision of the summary [16]. The next phase is sentence selection, when a voting system is used to choose the most relevant sentences. Lastly, a summary that is both cohesive and brief is produced by subjecting the chosen phrases to postprocessing, which includes methods like sentence ordering and redundancy elimination.

VII. EVALUATION OF EXTRACTIVE TEXT SUMMARIZATION APPROACHES

Key resources, such as standard datasets and assessment standards, used to assess the ETS systems are detailed in this section. Section A. Datasets A thorough review of the several corpora used for summarizing tasks is provided by El-Kassas et al. [1]. The most popular benchmarking datasets used to evaluate ETS systems are summarized in Table 2. Among them are: 1) Data Under Construction (DUC): Text summarization researchers often use datasets provided by the National Institute of Standards and Technology (NIST). Presented during the DUC conferences between 2001 and 2007, these datasets include papers accompanied by three kind of summaries: those that were hand-crafted, those that were produced automatically for baselines, and those that were made by participants for the challenges. The area of extractive summarization systems makes extensive use of these datasets [48], [49] [51], [53] [54].

Application forms may be found on the DUC website and must be completed in order to get access to DUC datasets. 1.

2) The TAC Datasets: In 2008, the DUC summarizing track became a member of the TAC. Applying extractive summarization techniques to these datasets is common practice [72], [125]. Application forms may be found on the TAC website and must be filled out in order to get access to these datasets. 2. Thirdly, the CNN/Daily Mail dataset has been used to test the ETS system on many occasions [27, 61, 65, 81]. Since it contains texts, summaries, and highlights [27], [128], the CNN-corpus Dataset [127] is ideal for summarizing a single document. By contacting the authors, researchers may acquire the whole annotated corpus for free. The fifth dataset is the PubMed dataset, which contains articles from the PubMed database that pertain to medical and biological research [129]. Extractive summarization problems are its usual domain [81, [95], [99], [130]. 6) The arXiv Dataset [129]: This repository has scholarly articles spanning several disciplines, including computer science, mathematics, physics, and many more. For tasks involving the summarizing of a single document, it works well [94], [99]. The seventh resource is the BBC News Dataset [131], which includes stories from many different fields of journalism, including as sports, business, technology, politics, and entertainment. Summarization of both individual and multiple documents is possible with this dataset [19], [27], [57], [70]. 8) Opinosis Dataset [132]: Contained in 51 files, each of which is devoted to a certain product feature and includes evaluations written by customers about that function. There are 51 different subjects covered, with each file including around 100 sentences and 5 hand-written gold summaries for each topic. Extractive summarization systems are evaluated using it [61], [133]. 9) EASC Dataset [134]: All articles written in Arabic and their extracted summaries were created by humans. EASC makes use of copyrighted material. With almost two million texts and summaries, the LCSTS Dataset [135] is the tenth option. A portion of the LCSTS dataset comes from SinaWeibo, a microblogging site in China. A number of features shared by all datasets are detailed in Table 2. the name of the dataset, the language used to write the data, the data domain, the size of the dataset (the total number of documents), and the support for multi-document and single-document summarization. The document count for multi-document datasets is "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

displaying sixty clusters, each containing about ten papers. B. Criteria for Assessment Two main approaches may be used to assess the summaries: 1) Performance in particular tasks, including reading comprehension or information retrieval, is used by extrinsic approaches to evaluate the quality of summaries. 2) intrinsic approaches rely on human review to determine the summary's quality; these methods pay special attention to the summary's covering of material and its coherence. There are two ways that summaries may be assessed: automatically and manually. 1) Evaluation by Hand Summaries are assessed by human judges using many quality criteria. These measures include readability, conciseness, grammaticality, structure, topic coverage, nonredundancy, and referential clarity [136]. Because it involves reading both the original papers and the summaries, manual review and analysis takes a lot of time [7]. To evaluate summary systems using automated metrics, the lightweight pyramids method use the semi-automatic standard gold summary [137]. 2) Evaluation via Computer Both more conventional measures, like ROUGE, and more contemporary ones, such BERTscore, will be covered in this section. a: Rough Measure When it comes to automatically assessing produced summaries, ROUGE is by far the most used tool. The method relies on comparing several human-written summaries with machine-generated ones [136]. The number of n-grams and other overlapping units between the reference texts and the candidate summaries is measured [138]. Depending on the variation used, ROUGE may additionally account for accuracy and F-measure, but recall is its primary focus. • ROUGE-1 compares the candidate summary to the reference summary by looking at the unigrams between the two. • The overlap of bigrams is assessed using ROUGE-2. • ROUGE-L prioritizes candidate and reference summaries with the longest matching sequences. When comparing the two summaries, ROUGE-S looks at the ratio of skip-bigram overlap. b: Bleu indicator Although it is most often used for machine translation, the Bilingual Evaluation Understudy (BLEU) score may also be used to measure the accuracy of n-gram summarizations. To ensure accuracy, it compares the reference text with the produced summary based on the number of n-grams (n-word sequences) that overlap [139]. To address the problem of too brief summaries, it measures the degree to which the produced summary corresponds to the reference summary in terms of the number of n-grams. The number 140 name: G-EVAL The quality of summaries is measured by this sophisticated assessment criteria, which focuses on the relevance to context and semantics. The produced summary's quality, relevance, coherence, and fluency are captured by Generative Evaluation (G-Eval). A framework that uses GPT-4 and chain-of-thought reasoning to evaluate NLG outputs, including summarization, has just been added to G-Eval [141]. G-Eval is now a trustworthy instrument for assessing the contextual alignment and semantic accuracy of extractive summaries, thanks to this method, which greatly enhances the alignment with human assessments.

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.

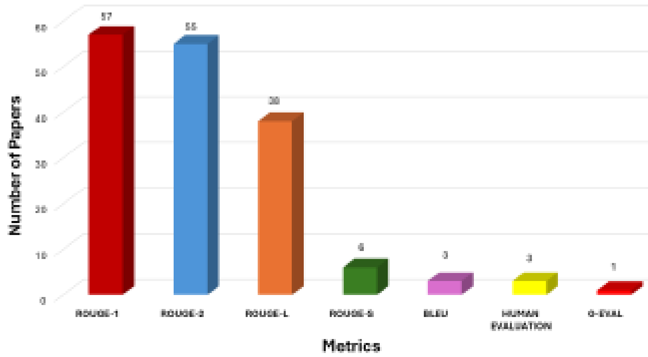

FIGURE 8. Frequency of evaluation metrics across reviewed studies.

d: Betreffosse Modern text generating tasks, including summarization, are evaluated using BERTScore, which makes use of contextual embeddings from pre-trained language models [142]. The BScore compares the reference and candidate summaries based on the degree to which their contextual BERT token embeddings coincide [142]. When comparing the retrieved sentences' semantic meaning to the reference summary, rather than lexical overlap, it performs very well [143]. It provides a more comprehensive assessment of summary quality by comparing tokens in the embedding space based on their cosine similarity. Figure 8 shows that the research field makes use of a variety of assessment measures, as shown by the analysis of the evaluated publications. With its appearance in 57 articles, ROUGE-1 clearly has the greatest traction as a valid and practical outcome evaluation tool. The significance of ROUGE-2 is highlighted in several studies, as shown by its 55 occurrences. The relevance of ROUGE-L and ROUGE-S in assessments is shown by their 38 and 6 applications, respectively. Three studies use BLEU, one uses manual human assessment, and one uses G-Eval. Evaluation metrics are compared in Table 3:

VIII. OPEN RESEARCH ISSUES AND CHALLENGES

In this part, we discuss the problems and difficulties in ETS and point forth possible areas for further study. Using an ETS system isn't without its share of difficulties, such as: Section A: Scaling Summarization for Difficult Tasks Website content, online reviews, and news pieces are just a few examples of the specialized uses for existing technologies. Complex applications, including book summaries and long texts, now need more labor [144]. Multiformat and multilingual data extraction may be improved. Summarizing information from a variety of semi-structured and textual sources (e.g., databases and websites) in a way that is suitable for individual users in terms of language, size, and format is the main difficulty. More studies on multimedia, multi-document, and multilingual summarizing are needed since the massive amount of data is accessible in many forms and languages [1]. C. Problems with Condensing Multiple Documents Reordering sentences, removing repetition, resolving conflicts between references, and considering temporal elements are among problems that could arise when trying to summarize many documents at once [9]. A specific problem might be the possibility of wrong references, such as when a proper noun is used in one phrase and then referred to by a pronoun in the next. The produced summary can be inaccurate if these references are not handled properly.Section D: Improving the Semantic The features includeThe detection of complex language traits and statistical characteristics of sentences and words that allow for the semantic extraction of the influential sentences from the original text [9]. Part E: Going Beyond Textual Inputs in Summarization Inputs and outputs based on text are the main emphasis of most summarization systems. New summarizing techniques are required that can process non-textual inputs such as audios, films, and meeting minutes and generate outputs in those forms [1]. Chapter F: Improving Detection Criteria When humans summarize materials, they do so iteratively, assessing how much to continue or stop once the first summary is generated. An enhanced mechanism for deciding when to end the summarizing process is desperately needed [145]. G. Developing Assessment Criteria Evaluating summaries, whether by hand or by machine, is fraught with difficulty. Since machine-generated summaries may be just as good as human-created ones, if not better, it can be difficult to tell which is the best [7]. Therefore, there is a need for fresh approaches and solutions for summarizing automatically.

IX. CONCLUSION

With an emphasis on ETS's consistency, reduced duplication, and wealth of information, this evaluation covers all of the methods used in ETS. Despite the fact that ETS research has progressed much over the years, there is still a lot that needs to be investigated since text summarization is seen as a challenging endeavor. The discipline has broadened its focus beyond scientific papers to include a wider variety of text kinds, necessitating new ways to text summarization, due to the increasing amount of text created on various social networks and news sites. When it comes to creating good summaries, important factors such as length, semantics, similarity, frequency, and keywords are essential. Other methods, such those based on fuzzy logic or machine learning, often rely on statistical methodologies for assistance. Multilingual summarization, complicated applications, and multidocument/multimedia processing should be the focus of future studies.

The fact that the summarized text has to be more accurate and widely accepted is another important consideration. As a result, it is crucial to find sophisticated statistical and linguistic elements for semantic extraction, as well as ways to find the best conclusions and enhance automated assessment. The following important findings are underlined in this study: A comprehensive examination and assessment of various strategies and techniques used in ETS is presented in this paper. 2) There's still a lot of room for improvement in summarization research, despite its longevity. Summarizing blogs, ads, news items, and emails has surpassed summarizing scientific studies as the primary focus. 3) When creating a useful summary, it's crucial to consider aspects like keywords, similarity, frequency, semantics, sentence location, and sentence length. 4) You may mix statistical techniques with fuzzy-based or ML approaches. 5) It's fairly uncommon to combine statistical methods with other approaches, such finding frequencies, evaluating similarities, and identifying keywords. 6) The need for study on complicated applications and multidocument, multimedia, and multilingual summarizing is one of the problems and future prospects. 7) It is crucial to identify sophisticated statistical and linguistic characteristics for efficient semantic extraction. 8) We also need new ways for automated assessment and more sophisticated ways to figure out the best way to end summaries. By outlining important topics and possible future advancements in this field, this review article hopes to provide the groundwork for future research.

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.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

[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.