[1]Vandana Jagtap

[2]Pallavi Parlewar

[3]Sheetal Dhande

[4]Anagha Langhe

[5]Harsh Choudhary

[6]Ashutosh Mishra

Advancements in Text Summarization Through Machine Learning: A Comprehensive Survey and Analysis

Journal of Electrical Systems

Abstract: - Due to the extensive amount of text available for any given task, for instance, a research project, it has become a need to have the gist of these documents in a succinct format. In this review paper, we discussed various methods used for single and multi-document summarization. It explores extractive, abstractive, and hybrid methods, along with the role of deep learning models like RNNs, CNNs, and transformers. The survey examines datasets, evaluation metrics, recent advancements, and future scopes in this field. A comparative analysis of methodologies and approaches is also presented.

Keywords: RNNs, CNNs, Transformers, single and multi-document summarization, extractive, abstractive, and hybrid summarization techniques.

I. INTRODUCTION:

The capacity to properly filter knowledge has become crucial at a time when there is an excessive volume of textual information available. Massive volumes of text are produced daily across many different sectors, from news. stories and academic researchto social media posts and corporate reports, as a result of a lot more information available due to the digital age. As a result, people and organizations have to work hard to go through a huge amount of text to find useful information for making smart choices and staying informed.

An effective way to address this issue is by using text summarization, which is a combination of natural language processing (NLP) and artificial intelligence (AI) technologies. It provides a methodical process for filtering extensive texts into concise, clear, and informative summaries that convey the main points of the original information. Text summarization makes it much easier to deal with a lot of information by automatically condensing content. It helps people and organisations quickly access the information they need.

Text summarization is important for more reasons than just practicality; it has wide-ranging effects on a variety of real-world applications. Automated news summary in the media industry helps readers quickly understand the main points of breaking articles. Condensed literature reviews let academic scholars explore the body of knowledge more quickly. Legal practitioners use condensed case materials to speed up the review process in legal practice. Medical professionals use summarization in the healthcare industry to effectively extract important data from patient records. The applications go beyond conventional areas and include a wide range of features including chatbots, content extraction, and recommendation algorithms for online content.

Despite text summarization's universal applicability and usefulness, the subject is ever-growing to meet new obstacles brought on by various text formats, languages, and sources. This extensive survey aims to explore the complex world of single and multi-document text summarization techniques, dive deep into the complexities of deep learning, assess the effectiveness of various evaluation metrics, and shed light on the open research problems that beckon both academics and industry professionals.

Our aim in this review is to help readers understand the basics, current trends, and future developments in text summarization. As we explore this diverse field, it becomes clear that the need for condensing knowledge is essential in the information age.

[1,4,5,6]School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Pune.

[2]Department of Electronics and Communication Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

[3]Department of Computer Science, Sipna College of Enginnering and Technology, Amravati, India

II. RELATED WORK:

In recent years, several notable survey papers and research studies have delved into the field of text summarization using machine learning techniques. In a recent survey [1] by Congbo Ma et al., they focused on deep learning techniques for multi-document summarization (MDS), introducing a novel taxonomy to organize existing techniques. Their survey discusses the similarities and differences between single-document summarization (SDS) and MDS while emphasizing the need for creating benchmarked datasets and improving objective functions and evaluation metrics. They also suggest exploring adversarial attack and defense techniques and introducing multimodality in the deep learning approach for MDS.

Sheng-hua Zhong, Yan Liu, and Bin Li presented a research paper on "Query-oriented unsupervised multi-document summarization via deep learning" in Expert Systems with Applications [2], which leverages deep learning for extractive multi-document summarization without the need for labelled training data. They also propose potential improvements, such as incorporating compression decisions and integrating the model with effective classifiers like SVM.

Yang Liu, Ivan Titov, and Mirella Lapata [3] introduced a unique approach to single document summarization, treating it as a tree induction problem, ultimately leading to the SUMO (Structured Summarization Model).

Shengli Song et al. (2018) presented "Abstractive Text Summarization using LSTM-CNN based deep learning," [4] which employs LSTM-CNN based ATS framework (ATSDL) for extractive summarization.

In addition to these techniques, Min-Yuh Day and Chao-Yu Chen [5] introduced three AI models for generating candidate titles, leveraging various keyword indicators and features extracted from abstracts, while Afsaneh Rezaei et al. explored autoencoder and deep belief network methods [6].

Nikita Munot and Sharvari S. Govilkar [7] conducted a comparative study of text summarization methods, categorizing them into extractive and abstractive methods. Nikhil S. Shirwandkar and Dr. Samidha Kulkarni presented an extractive summarization approach using a combination of Restricted Boltzmann Machine and Fuzzy Logic for feature extraction [8].

Heena A. Chopade and Dr. Meera Narvekar [9] introduced a hybrid system combining the Restricted Boltzmann Machine (RBM) and fuzzy logic for feature extraction. Mehdi Allahyari et al. explored techniques for evaluating summarization methods, emphasizing the impact of context on summarization [10].

G. MalarSelvi and Dr. A. Panadian in [11] discussed various automatic text summarization techniques, including statistical, graph-based, machine learning, deep learning, and semantic approaches. Siyao Li et al. in [12] proposed Deep Reinforcement Learning with Distributional Semantic Rewards (DSR) for abstractive summarization, while Riva Malik et al. (2023) introduced Maximal gSpan, an extension of the gSpan frequent subgraph mining algorithm for Multi-Document Summarization [13]. Their approach aims to reduce redundancy in summaries. Additionally, Ishitva A. et.al. [14] explored various techniques in text summarization, including extractive, abstractive, reinforcement learning, and hybrid methods.

Aishwarya Padmakumar et al. introduced an unsupervised text summarization method using sentence embeddings, clustering, and sentence selection [15]. Xiaohan Fang et al. [16] presented an improved Pointer-generator network and a TextCNN model for text summarization and classification. Prabhudas Janjanam and CH Pradeep Reddy comprehensively studied various text summarization techniques, including text representation, graph-based methods, semantic approaches, and optimization methods [17].

Adding to these, Hamzah Siddiqui et.al. in [18], discussed two extractive methods: TF-IDF and TextRank. Juan Ramirez-Orta and Evangelos Milios introduced an unsupervised document summarization approach using the methods of SBERT sentence embeddings and graph centrality [19].

In [9], Shruti Aggarwal et al. discuss devices that machine learning techniques to enhance rice production. The paper provides an overview and analysis of numerous papers published in the past eight years, covering various methodologies related to the identification of crop diseases, seedling health, and grain quality.

The authors of [10] use a color-aware two-branch DCNN for efficient plant disease classification which outperforms the baseline model. Using the Plant Village dataset, the two-branch architecture achieves 98.2% accuracy and a multiclass F1 score of 0.981. However, the paper lacks comparisons with transfer learning methods and exploration of alternative color spaces.

In [11] Andrea G. presents a continual development methodology for unbounded intelligent systems in machine learning. The approach involves generating dynamic multitask models through sequential extensions and generalizations. This when applied to image classification using the Plant Village dataset, their μ2Net+ model outperforms ResNet, DenseNet, and NASNet in accuracy, F1 score, and AUC-ROC metrics. The paper suggests potential future work to enhance system capabilities across multiple modalities.

An adaptive minimal ensemble method with EfficientNet CNNs is proposed by Bruno A. et. al. in [12]. They augment the PlantVillage dataset and achieve state-of-the-art accuracy at 99.1%. However, there is an absence of comparisons with other methods and a lack of exploration of alternative CNN architecture along with the impact of different data augmentation techniques.

Schwarz Schuler et. al. in [13] use a lightweight DCNN for plant leaf disease classification using a modified Inception V3 architecture with two branches for L and AB channels. After testing on the PlantVillage dataset, the proposed method achieves 99.06% accuracy, outperforming many state-of-the-art methods. While the paper lacks explicit discussion of research gaps, it excels in evaluating metrics like precision, recall, F1-score, and AUC-ROC.

These various techniques and studies help us to understand how to summarize text using machine learning. The way we summarize text is always changing, from using deep learning and neural networks to using graph-based techniques. The aim is to make short, helpful, and clear summaries of the text. In the future, scientists may try to make text summaries better and shorter and find new ways to do this important language task.

III. BACKGROUND FOR TEXT SUMMARIZATION:

The system offers a promising solution for the management of plant diseases in agriculture, leading to more sustainable environment-friendly farming practices.Text summary is the process of extracting the main ideas and information from long text paragraphs and reducing it to a shorter, more contextual version. In order to manage the ever-growing amount of textual material that is accessible on the internet and other digital sources this has become more important now than ever. Efficient content consumption has become necessary due to the vast abundance of information on the internet. It is impractical to read every document in its entirety since users frequently experience information overload [20]. This problem is solved by text summarizing, which offers brief summaries that enable readers to rapidly determine a document's significance before reading the full document [21]. Text summarization has been made possible in large part by machine learning, especially deep learning [22]. Neural networks, including CNNs, RNNs, and Transformer models, have been used by researchers [23] to train models that comprehend the semantics and context of a document. These models are trained to extract and produce key sentences that summarize the important ideas in the material.

Although there are other techniques for text summary, extractive and abstractive summarization are the most popular ones. In order to create the summary, extractive summarizing techniques choose and put together sentences from the original text [22]. Abstractive summarization methods aim to generate summaries that are not directly extracted from the source text but synthesized by rephrasing and restructuring the content [24]. Hybrid approaches that combine extraction and abstraction have also gained attention [20]. Additionally, techniques such as reinforcement learning, pointer-generator networks, and pre-trained language models like BERT and GPT-3 have shown promise in improving the quality of summaries. Text summarization is a vital NLP task, driven by the need to efficiently process vast amounts of textual data. Machine learning, especially deep learning, has revolutionized text summarization by enabling the development of models that can generate coherent and concise summaries.

Fig. (1) shows the classification of the most used state-of-the-art text summarization techniques in the form of a tree diagram.

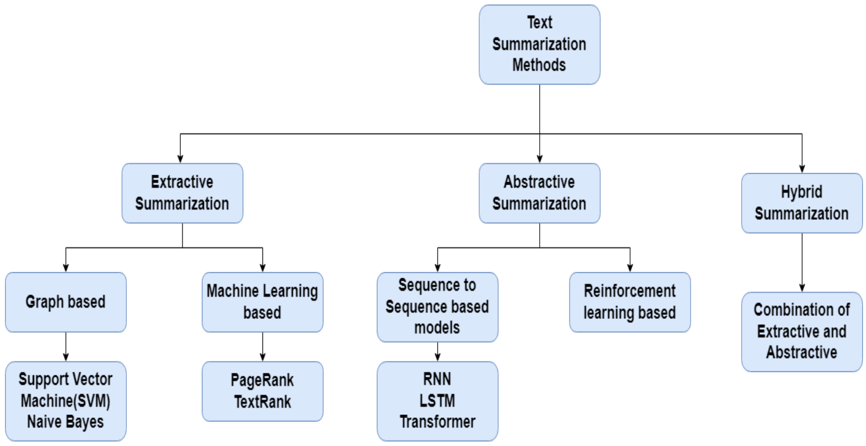

Figure. 1.Classification of text summarization methods

Researchers continue to explore various techniques and approaches to enhance the quality and adaptability of text summarization systems in response to evolving information needs.

Some of the state-of-the-art methods have been discussed in the following sections along with the various evaluation metrics used for the task of machine learning text summarization.

IV. SINGLE DOCUMENT SUMMARIZATION:

Single-document summarization is the task of automatic generation of shorter versions of a document while retaining its most important information [25]. It has received much attention in the NLP community for its potential for various information access applications. Some examples of single document summarization include the tools which when fed textual content provide recommendations, answer questions, etc. It aims at summarizing a single document. Many summarization paradigms have been identified over the years, but two of them have consistently grabbed attention, *extractive* and *abstractive* summarization.

1.1 Extractive Summarization

Extractive Summarization aims to create a summary by selecting and extracting sentences or phrases directly from the source text or document [6]. This approach assumes that the most informative content is already present in the original text. Some of the key aspects of extractive summarization include -

Sentence scoring: In this approach, sentences are scored based on various criteria that include the importance of terms, sentence features, and many more features. The most common scoring methods are TF-IDF, TextRank, and PageRank.

Content selection and Redundancy Handling: Sentences with the highest scores are chosen to be included in the summary. Techniques such as Maximum Marginal Relevance (MMR) are often employed to improve diversity and reduce redundancy.

Unsupervised extractive document summarization is one of the many strategies that have been used to date. This method's strategy aims to identify the most significant sentences inside a document without the need for a labeled corpus. The Graph-Based Unsupervised Summarization Approach [26], which is predicated on the idea of including just the most significant sentences while omitting the sentences with identical meanings in the summary, is one of the most useful strategies in this method. In order to provide a summary, it makes use of graph-based approaches to determine and pick the most crucial sentences or nodes from a document. Although it can also be used for multi-document summarizing, this method is mostly utilized for single-document summarization. This is a thorough description of the GUSUM model.

Sentence-to-sentence Similarity Graph: The implementation of the model starts with the representation of the content as a graph. Each sentence in the document is represented by a node in the graph. The edges between the nodes are weighted based on the similarity between the sentences. The similarity metrics that can be used are cosine similarity, Jaccard similarity, or some more advanced measures such as Word2Vec or BERT embeddings.

The stronger the similarity between the edges, the higher the similarity between the two sentences represented by the nodes.

Centrality Score Calculation: After the generation of the similarity graph, the GUSUM model calculates centrality scores for each node in the graph. Some of the major centrality measures that are been used are degree centrality and closeness centrality which are used to assess the importance of a sentence within the entire document. The central idea is that the sentences that are more central in the graph are likely to be important and representative of the document's content.

Sentence Ranking: The sentences are ranked based on their centrality scores. The sentences with higher centrality scores are considered more important and are more likely to be included in the final summary.

Selection of Top Sentences: The GUSUM model selects the most relevant sentences to construct the final summary. The number of sentences that are selected can be controlled based on the desired summary length or compression ratio. After all these steps are followed, the sentences that are most relevant are finally selected for the summary construction and are combined to form the extractive summary. The order of the sentences can be preserved as they appeared in the original document, or they can be re-arranged to create a more coherent summary.

Other than the GUSUM model, there are a lot of other notable models that have been used for extractive summarization such as the BERTSUM neural network model that fine-tunes the BERT model for extractive summarization. Also, the Latent Semantic Analysis Method which uses dimensionality reduction techniques to capture latent semantic relationships between words and sentences.

1.2 Abstractive Summarization:

Abstractive summarization is a technique to create coherent summaries of single documents by paraphrasing and rephrasing the source text [12]. This approach provides a more human-like summary hence it may contain new sentences that are not present in the original text. Thus, this technique is very effective for condensing complex or lengthy documents. One of the key techniques that are used in abstractive summarization is the Encoder-Decoder Architecture.

Encoder-Decoder Architecture: It is a framework used in various NLP tasks [14]. Most of the abstractive summarization models are built on encoder-decoder architectures. It consists of two main components: the encoder and the decoder. This architecture is frequently used for sequence-to-sequence tasks, where one sequence of data (i.e. text from the source document) is transformed into another sequence i.e. the summary document.

The primary role of the encoder in the encoder-decoder architecture is to process the input text and create a fixed-length representation which is referred to as the "context vector", which captures the essential information of the source text. Prominent deep learning models like recurrent neural networks(RNNs), convolutional neural networks(CNNs), or transformers like BERT or GPT can serve as the encoders in the architecture. The final representation of the encoder serves as the context vector and this is passed to the decoder. The decoder then takes the context vector and generates the output sequence, often by taking one token at a time. The decoder can also be implemented using RNNs, transformers, or other neural network architectures like encoder. Therefore, the primary tasks of the decoder are generating tokens, capturing the context of the source text or document, and passing this to the output layer which is responsible for predicting the next token in the sequence.

V. MULTI-DOCUMENT SUMMARIZATION:

Multi-document summarization is an NLP task [28] that looks to condense information from multiple documents into a coherent and concise summary. Multi-document summarization aims to integrate the information from multiple documents on the same topic unlike single document summarization, which focuses on condensing a single document or text into a summary. Thus this technique can be used for tasks such as information retrieval, document clustering, news aggregation, and content summarization. Some of the key aspects of multi-document summarization are:

1. Cross-Document Summarization:

The fundamental step of multi-document summarization is the extraction and synthesis of information of information from different source documents. This is required so that the model can identify the relevant content, remove redundancy and lastly create a unified content that represents the key points across multiple texts.

2. Content Fusion:

Multi-document summarization models must consider the different strategies for content fusion to create coherent summaries. This step includes selecting sentences or passages from source documents, identifying similar sentences merging them and extracting key entities and events that span multiple texts.

3.Diversity and Coverage:

An effective multi-document summarization should ensure diversity and comprehensive coverage on the topic. This means capturing a wide range of perspectives and aspects related to the topic while avoiding over-representation of specific information.

One of the prominent techniques used for multi-document summarization is the Frequent subgraph mining [13], which is a data mining technique used for discovering recurring patterns or subgraphs within a collection of graphs or networks. It can be applied to find common patterns across a set of documents being referred for multi-document summarization that aids in the summarization process. We'll look in more detail at how this technique works.

Graph Representation:

Documents are referred to as graphs in multi-document summarization, where nodes represent elements (e.g., sentences, paragraphs, or sections) and the edges represent the relationships between these elements. These relationships can be based on various criteria such as semantic similarity, co-occurrence of terms, etc.

Constructing the Document Graph and Frequent Subgraph Mining:

All the documents in the collection are transformed into a graph, and the graphs from all the documents combine into a single large graph. This final graph represents the entire document collection, with the nodes corresponding to elements from all the documents and edges that reflect the relationships between them.

Frequent Subgraph mining algorithms are then applied to analyse the graph. It identifies recurring subgraphs (subsets of nodes and edges) that appear frequently within this final combined graph which represent the structural patterns that are shared among multiple documents. Frequent subgraph mining algorithms such as gSpan or FSG (Frequent SubGraphs) search for the subgraphs that meet certain support criteria. This support criterion is a threshold that determines how frequently a subgraph must appear across the document collection to be considered as "frequent". Finally, the discovered frequent subgraphs can be used as building blocks for generating a multi-document summary. The subgraphs themselves or elements within them (i.e., sentences) can be extracted and included in the summary (extractive summarization) or the subgraphs can be used to guide the generation of abstractive summaries (abstractive summarization) by filling in the summary content with the content generated by natural language generation techniques.

VI. EVALUATION METRICS:

In the context of text summarization using machine learning, the assessment of summary quality is an important factor. Evaluation metrics play a critical role in objectively measuring the performance of various summarization techniques. These metrics enable researchers and practitioners to quantitatively compare the effectiveness of different algorithms, providing a foundation for assessing the progress and innovation in the field of automatic text summarization [29,39].

1.3 ROGUE:

The ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [3][10][12][22] metric is a set of evaluation measures commonly used in natural language processing and text summarization to check the quality of generated text. ROUGE scores are widely employed for tasks such as machine translation, text summarization, and machine-generated content generation.

ROUGE-1[2][3][4][5][6][15][16] measures the overlap of unigrams (single words) between the generated and reference text. It assesses how many words in the generated text are also present in the reference text and this provides a simple and intuitive way to assess the lexical similarity between the generated and reference text. ROUGE-2[2][3][4][5][6][15][16] calculates the overlap of bigrams (pairs of consecutive words) between the generated and reference text i.e. it measures the similarity in word pairs which helps in capturing some degree of phrase-level similarity and is more informative than ROUGE-1 for assessing the quality of longer phrases in the text.ROUGE-3 and ROUGE-4 extend the concept to trigrams (ROUGE-3) and four-grams (ROUGE-4), respectively. They assess the quality of longer word sequences in the generated text. These scores are particularly useful for capturing the quality of longer phrases and sentence structures. ROUGE-L [3][5][6][16] measures the longest common subsequence (LCS) between the generated and reference text. It evaluates the structural similarity in terms of word ordering which is valuable for assessing the fluency and coherence of generated text. ROUGE-N [19] (N-gram Co-occurrence) is a generic formulation that encompasses all the above N-gram-based ROUGE scores (ROUGE-1, ROUGE-2, etc.). It allows for the evaluation of text similarity at various N-gram levels providing the flexibility required to choose the appropriate N-gram level for a specific evaluation task. ROUGE-W (Weighted N-gram Overlap) introduces a weighted version of N-gram overlap, where word frequencies in the reference text are considered. It assigns higher importance to content-bearing words which addresses issues related to stop words that might artificially inflate N-gram overlap scores and provides a more nuanced evaluation of content overlap.

ROUGE scores offer several key benefits in the evaluation of generated text. They provide an objective and quantitative means to assess text quality, ensuring systematic and replicable evaluations. Also, their granularity allows for versatile evaluation ranging from individual words to longer phrases, making them adaptable to various tasks. Additionally, ROUGE scores enable effective comparisons between different models and systems, which aids in the identification of superior methods. Their interpretability is invaluable for pinpointing areas of similarity between generated and reference text, hence assisting in error analysis and model enhancement.

1.4 HowNet:

HowNet [30]is a comprehensive lexical knowledge base that was developed specifically for the Chinese language. It aims to capture the semantic relationships between words and concepts, providing a structured representation of the meaning of Chinese words. HowNet is a useful tool for evaluating the semantic quality of text summarization since it is structured as a lexical network, in which words are connected according to their semantic characteristics.

HowNet works by supplementing words and phrases with semantic annotations that include important information such as concepts, qualities, and relationships, allowing for more in-depth comprehension of their meanings. This knowledge base not only semantically annotates individual terms, but it also creates a hierarchical structure (like a semantic tree) in which concepts are interconnected via connections such as "is-a," "part-of," and "attribute-of." This hierarchical design depicts the relationships between multiple words and concepts, allowing for a more sophisticated interpretation of semantics. HowNet then determines semantic similarity using this framework by analysing the semantic properties of words and phrases present by taking into account their hierarchy. This similarity is then used to determine how close a generated summary's content corresponds with the content of the reference document.

HowNet's unique approach to semantic evaluation has various advantages over traditional metrics such as ROUGE. HowNet provides a more comprehensive assessment by prioritizing semantic similarity over exact word overlap, which is useful for evaluating abstractive summarization models that may express concepts differently.HowNet's language independence qualifies it to overcome language-specific problems, making it adaptable to a wide range of symbol-based languages. Its ability to recognize synonyms and paraphrases increases robustness, and mitigates limitations associated with traditional metrics that rely solely on exact word matches. It fosters a deeper content understanding, extending beyond surface-level lexical overlap to evaluate coherence and appropriateness in summarized content. Additionally, its customization capability permits adaptation to domain-specific terminology and concepts, further enhancing its relevance and applicability across various summarization tasks. The use of HowNet as an evaluation metric for Chinese text summarization [30] offers a semantic perspective on summarization quality. It provides a more nuanced and meaningful evaluation compared to traditional metrics, especially in the context of abstractive summarization. However, it's worth noting that while HowNet excels in capturing semantic similarity, it may not fully capture aspects like fluency and coherence, which are also important dimensions of summarization quality. Therefore, it is often advisable to complement HowNet with other evaluation metrics for a comprehensive assessment of text summarization systems.

1.5 F1 Score:

F1 Score [3][8][6][11] [42][43] is a widely used evaluation metric, particularly in the context of text summarization and natural language processing tasks. It measures the balance between precision and recall, providing a single score that reflects the trade-off between these two aspects. Precision meticulously quantifies the accuracy of the content chosen or generated for inclusion in the summary, offering insights into the proportion of correctly identified information. Recall, on the other hand, assesses the summary's comprehensiveness by calculating its capacity to extract pertinent information from the reference text. These two metrics work together to form the base for calculating an F1 Score. The F1 Score is the harmonic mean of precision and recall that delivers a well-rounded rating of summary quality. This is mathematically presented in the F1 Score formula: F1 Score = 2 * (Precision * Recall) / (Precision + Recall).

The F1 Score's strength comes from its ability to condense the evaluation process into a single score that provides a clear picture of overall quality. This simplicity, combined with its basic calculating approach, which does not require complex algorithms or linguistic resources, facilitates both, implementation and understanding. The F1 Score is a versatile and necessary instrument for summarising system evaluation since it analyses the quality of selected sentences or phrases in extractive summarization and the fluency and correctness of generated content in abstractive summarization. Furthermore, it offers valuable insights into the balance struck between content inclusion and accuracy, helping in the identification of areas where a summary may favor one aspect over the other.

1.6 BERTScore:

BERTScore [31][26] is a more recent evaluation metric used for assessing the quality of machine-generated text, mainly in comparison to reference text. BERTScore harnesses the power of BERT (Bidirectional Encoder Representations from Transformers), a cutting-edge language model which is known for its contextual understanding of text. By utilizing BERT, BERTScore obtains contextual embeddings for words and phrases within both the machine-generated and reference text. These embeddings capture not only the lexical meaning of individual terms but also their context within sentences and paragraphs, resulting in a more profound understanding of semantics and context. Following that, BERTScore uses cosine similarity to quantify the degree of similarity between the created and reference text BERT embeddings. This similarity evaluation works at both the word and phrase levels, allowing for a fine-grained assessment of how closely the generated text matches the reference text in terms of meaning. Finally, BERTScore aggregates the individual cosine similarity scores at the phrase level, combining them using an appropriate aggregation approach, such as mean or weighted mean [31]. This sentence-level aggregation guarantees that the entire text has a consistent and unique BERTScore, providing a comprehensive measure of semantic alignment and contextual fidelity between machine-generated and reference information.

BERTScore provides a holistic method of text evaluation, encompassing various useful benefits. It excels, first and foremost, at capturing semantic similarity, transcending surface-level lexical overlap, and efficiently evaluating the preservation of meaning and context in the generated text—a critical attribute for tasks such as abstractive summarization and machine translation. This is achieved through BERT embeddings, which are sensitive to contextual nuances, rendering BERTScore robust to variations in word order, sentence structure, and paraphrasing. The model's ability to recognize synonyms and alternative phrasings contributes to its versatility. Users can employ BERTScore with different BERT models, affording the flexibility to select the most suitable model for their specific evaluation requirements, and enabling applicability across diverse languages and domains. BERTScore further offers the advantage of sentence-level evaluation, allowing for a granular assessment of text quality at both phrase and sentence levels, pinpointing areas that may require improvement. Its state-of-the-art performance, evident in competitive rankings across various natural language processing tasks, underscores its effectiveness as an evaluation metric. Notably, BERTScore streamlines the evaluation process by obviating the need for reference text preprocessing, as it directly compares raw text, simplifying the assessment of machine-generated content. However, its large computational complexity results in slower evaluation times compared to traditional metrics. Additionally, BERTScore's sensitivity to the choice of the BERT model can introduce variability in scores. Handling rare or out-of-vocabulary words can pose challenges, and its performance is closely linked to the quality of pre-trained BERT models.

1.7 G-Score:

The G-score [11], also known as the G-measure or G-score, is a natural language processing and text summarization assessment metric used to assess the quality of machine-generated content. It is well and suited to tasks involving various aspects of evaluation, such as content relevance, fluency, and informativeness.

The G-Score comprises a number of evaluation components or criteria, each of which is designed to examine different parts of the output text. Content relevance, coherence, fluency, and informativeness are examples of these components. Different components have different weights assigned to them based on how important they are in the evaluation process. For example, in some summarising situations, subject relevance could matter more than fluency. Subsequently, every element of the test is scored separately. Consider a scenario in which fluency is assessed using a metric based on language models, but content relevance is determined using a metric similar to ROUGE. In the end, the component scores are combined into a single composite score using a weighted aggregation method. Although the G Score calculation varies, it usually involves adding up the scores of each component after multiplying them by their respective weights.

The G-Score is a dynamic and comprehensive method that considers numerous important aspects when evaluating the quality of text. Its thorough examination takes into account a variety of text quality factors, including informativeness, coherence, fluency, and relevance of the content. For tasks that call for a thorough review, this makes it a suitable substitute. Because of the G Score's adaptability, evaluations can be tailored to meet the unique objectives of different natural language processing tasks by changing the weights and creating custom criteria. Its adaptability makes it suitable for a wide range of applications. Moreover, the G Score can be customized to the needs and preferences of the user, guaranteeing that the assessment is in line with practical requirements.

While the G Score has numerous advantages, it also necessitates careful design and adjustment to meet individual tasks and areas. The evaluation components and their weights should be chosen in accordance with the evaluation objectives.

1.8 BLEU:

A popular metric for evaluating the quality of machine-generated text, especially in machine translation, is called BLEU (Bilingual Evaluation Understudy) [32]. It compares the number of n-grams in the generated text with the reference text. By counting matching n-grams, it calculates the precision for different n-gram types (such as unigrams and bigrams). In order to avoid translations that are both too brief and too accurate, BLEU has a shortness penalty that encourages longer summaries. The metric uses a cumulative method to create an overall BLEU score by computing precision scores for various n-gram lengths [32]. This method guarantees a fair assessment that takes into account both summary length and precision, which makes it an invaluable tool for assessing the completeness and accuracy of text generated by machines, especially when it comes to machine translation.

The natural language processing community has widely adopted and recognized BLEU due to its rich set of features. First off, it offers a quantifiable and objective way to quantify text quality, which makes it an essential tool for contrasting and evaluating machine translation systems. Since evaluating translation accuracy is crucial in research and development, this objectivity is especially crucial. Apart from that, BLEU is well-known for its simple computation process, which compares n-grams and applies a brevity penalty. Because of its simplicity practitioners may easily adopt and comprehend it, assuring accessibility.BLEU has gained widespread acceptance as a standard and frequently used metric in the field of machine translation due to its objectivity, simplicity, and efficacy. It provides a dependable and consistent framework for the creation and assessment of translation systems.

BLEU is a useful metric, however, it has limits when it comes to assessing machine-generated texts. Its emphasis on n-grams, which means it ignores word order and semantic subtleties, is one significant disadvantage. This may result in cases when BLEU gives translations that are grammatically accurate but semantically inaccurate high marks, which could lead to inaccurate assessments. Furthermore, BLEU may be sensitive to little phrasing discrepancies between the generated text and the reference, which makes it difficult to offer insightful feedback for small translation modifications.

1.9 SacreBLEU:

An extension to BLEU called SacreBLEU [33] is designed to deal with the lack of reference data in machine translation tasks, especially in languages or domains with limited resources. This is accomplished by first using a small, well-researched reference dataset, which is then purposefully made larger by rearranging and paraphrasing it to produce a larger reference dataset. Similar to BLEU, SacreBLEU assesses the generated text's n-gram precision against this enhanced reference dataset, imposes a shortness penalty, and generates an overall score. This methodology guarantees reliable assessment in situations where authentic reference translations are limited, enabling precise evaluation of machine translation quality.

SacreBLEU efficiently tackles the problem of sparse reference data, making it a useful tool in machine translation evaluation. This is achieved by intentionally increasing the size of the reference dataset [33], which guarantees accurate assessment even in situations where there are few real reference translations. SacreBLEU also preserves compatibility with the standard BLEU metric, providing uniform assessment procedures in a range of machine translation contexts, independent of the availability of data. Its dual benefits of data expansion and alignment with BLEU increase its adaptability and usefulness in resolving issues related to data shortage. Although its reliance on artificially expanded reference datasets may prove highly beneficial it is a double-edged sword in itself. It can introduce biases or inaccuracies, as the generated references are not genuine translations. This can affect the reliability of the evaluation, particularly if the paraphrased references do not adequately capture the desired diversity of translations. Furthermore, SacreBLEU, like BLEU, still relies on n-grams and is subject to the same challenges related to word order and semantics.

VII. COMPARATIVE ANALYSIS:

We have evaluated multiple machine learning pipelines for text summarization using the CNN/Daily Mail Dataset in this paper. The summarization pipelines used for analysis in this paper are GPT-2, T-5, BART, and Pegasus.

1.10 CNN/Daily Mail Dataset:

One popular benchmark for text summarising is the CNN/Daily Mail dataset. There are roughly 312,000 training pairings, 13,000 validation pairs, and 11,000 test pairs in all. A human-generated summary of a news article and the original article make up each training pair. The abstraction of the summaries—that is, their non-repetitive usage of specific terms or phrases from the original articles—makes the dataset challenging. Rather, they concisely and informatively summarise the articles' key ideas.

1.11 Hugging Face Pipelines:

For natural language processing (NLP) applications, such as text summarization, the Hugging Face pipeline offers a strong and adaptable toolset. In addition to offering a number of pre-processing and post-processing options to optimize the NLP workflow, it offers a single interface to access a large number of pre-trained NLP models, such as BART, Pegasus, and T5[34, 40].

By abstracting away, the complexities of model loading, tokenization, and model inference, the Hugging Face pipeline makes the use of pre-trained NLP models easier. This frees academics and practitioners from getting mired in the technical details of the underlying models and enables them to concentrate on the particular task at hand, such as text summarization.

1.12 GPT-2:

Text summarization is just one of the many jobs that may be performed with the pre-trained GPT-2 Hugging Face text-generating model. The GPT-2 architecture, a sizable language model trained on an enormous dataset of text and code, serves as the foundation for this model.

OpenAI created the GPT-2 family of big language models. The models can produce logical and lifelike text because they were trained on a sizable dataset of text and code. GPT-2 can perform a wide range of tasks, such as question answering, translation, and text summarization [35].

There are three sizes for the GPT-2: small, medium, and big. There are 117 million parameters in the small model, 345 million in the medium model, and 1.5 billion in the large model. Though generally more sophisticated, the larger versions also demand more processing power.

1.13 T-5:

Strong and well-trained in text-to-text, the T5 Hugging Face pipeline is a Transformer model that is well-suited for a variety of natural language processing applications, including text summarization. It is based on Google AI's T5 (Text-To-Text Transfer Transformer) architecture, which has shown impressive powers in natural language understanding and generation.

Similar to GPT-2, T5 is available in small, medium, base, large, and 3B sizes. A distinct trade-off between computing resources and performance is available for each size [36]. Though they might not be as accurate as larger models, the smaller models are more effective nonetheless. Larger models, on the other hand, can generate summaries that are more detailed and subtle but also demand more processing power.

1.14 BART:

Another potent text-to-text pre-trained Transformer model created especially for text summarization applications is the BART (Bidirectional Autoregressive Transformer) Hugging Face pipeline. Through the use of Facebook AI Research's BART architecture, which combines the advantages of autoregressive and encoder-decoder techniques, it is able to capture both local and global context while producing summaries.

BART is available in base, large, and distilled sizes, just like other Transformer models. A distinct trade-off between performance and processing resources is provided by each size. Larger models might be required for challenging or complex summarising tasks, but for most other activities, the base model works well as a starting point [37]. Environments with limited resources can benefit from the distilled model, which is a more compact and effective version of the base model.

1.15 Pegasus:

A flexible and strong text-to-text pre-trained Transformer model, the Pegasus Hugging Face pipeline performs exceptionally well across a range of natural language processing applications, including text summarization. It is based on Google AI's Pegasus architecture, which has shown impressive powers in natural language creation and understanding.

Pegasus, in contrast to other Transformer models, focuses exclusively on abstractive summarising, in which the summary produced goes beyond mere paraphrasing of the original text and presents the key points in a clear and informative manner. Pegasus accomplishes this by using a pre-training technique called gap-sentence, in which key sentences are hidden and then pieced together from the remaining text [38]. This method makes the model more capable of producing abstractive summaries by encouraging it to learn how to recognize and produce important information.

Pegasus is currently available in two sizes: large and multi-news. The large model performs better on a wider range of summarization tasks because it was trained on a massive dataset of text and code. The multi-news model is a good option for applications involving news because it is specifically designed to summarise news pieces.

1.16 Implementation:

We have utilised the CNN/Daily Mail Dataset, version 3, for the analysis in this study. The models were evaluated using the first 1000 articles and the accessible highlights from the dataset's test split.

We first created a baseline model that employs a basic summarization technique by selecting the first 3 sentences from the article.

We have used the medium model for GPT-2, the small model for T-5, the large model for the BART pipeline, and the pegasus-cnn_dailymail model for the Pegasus pipeline in order to evaluate their performance for text summarization.

The sample dataset we have used for the purpose of testing these models can be accessed from [41]. The highlights in the dataset are –

Mentally ill inmates in Miami are housed on the "forgotten floor"

Judge Steven Leifman says most are there as a result of "avoidable felonies"

While CNN tours facility, patient shouts: "I am the son of the president"

Leifman says the system is unjust and he's fighting for change.

The performance of different models is as follows –

1. Baseline:

Editor's note: In our Behind the Scenes series, CNN correspondents share their experiences in covering news and analyse the stories behind the events.

Here, Soledad O'Brien takes users inside a jail where many of the inmates are mentally ill. An inmate housed on the "forgotten floor," where many mentally ill inmates are housed in Miami before trial.

MIAMI, Florida (CNN) -- The ninth floor of the Miami-Dade pretrial detention facility is dubbed the "forgotten floor."

---

2. GPT-2:

The Eighth Amendment of the U.S. Constitution protects citizens from unreasonable search & seizure.

This is what Leifman said in a 2012 hearing before Florida's Florida Supreme Court when questioned by Judge Patricia Smith to address the problem of mentally ill people being arrested by police.

That hearing occurred three months after a local judge in Miami ordered the state of Florida to stop using police officers as stand-by for all mentally ill people, something the U.S. Supreme Court ruled as unconstitutional a month before.

"I think I've made it clear that the Fourth Amendment prohibits police officers from using non-law enforcement officers," Leifman said.

"And I believe in common law -- and I'm proud of this fact -- in common law, it is unlawful for police officers to serve as stand-bys for a mentally ill person."

He noted that the Fourth Amendment was intended to stop a person suffering involuntary manslaughter when his or her own death threatened to occur as a result of their inability to protect themselves.

"It was meant to stop a person coming into the presence of the police and having someone else be there in the presence of them," he said, "so I don't think that's very helpful in addressing mentally ill people who have been arrested."

Posted by Jacob at 4:33 PM

---

3. T-5:

inmates with the most severe mental illnesses are incarcerated until they're ready to appear in court. most often, they face drug charges or charges of assaulting an officer. nmentally ill people become more paranoid, delusional, and less likely to follow dir.

5. BART:

Miami-Dade pretrial detention facility is dubbed the "forgotten floor" Here, inmates with the most severe mental illnesses are incarcerated.

Most often, they face drug charges or charges of assaulting an officer.

Judge Steven Leifman says the arrests often result from confrontations with police.

6. Pegasus:

Mentally ill inmates are housed on the "forgotten floor" of a Miami jail.

Judge Steven Leifman says the charges are usually "avoidable felonies"<n>He says the arrests often result from confrontations with police.

Mentally ill people often won't do what they're told when police arrive on the scene.

For the evaluation of these models, we used the ROGUE-1, ROGUE-2, ROGUE-L, and the ROGUESUM evaluation metrics. The evaluation scores of these models are presented in table (1).

Table 1.ROGUE-1, ROGUE-2, ROGUE-L and ROGUE-LSUM scores for the models

Model	ROGUE-1	ROGUR-2	ROGUE-L	ROGUE-LSUM
Baseline	0.365079365079365	0.14516129032258066	0.20634920634920634	0.2857142857142857
GPT-2	0.16666666666666666	0.043795620437956206	0.11594202898550726	0.15217391304347824
T-5	0.1758241758241758	0.0	0.13186813186813187	0.15384615384615383
BART	0.3655913978494624	0.13186813186813184	0.2150537634408602	0.3225806451612903

VIII. CHALLENGES:

With the developments in artificial intelligence and machine learning, it has become easier to summarize texts from different sources and languages. However, these machine learning-based techniques employed for the purpose of summarization face a few challenges as well. Some of the commonly occurring problems include –

A. **Data Scarcity:** Ensuring that summarization models can adapt to new domains and generalize well to unseen data is crucial for real-world applications. Lack of sufficient or high-quality data for training or testing the summarization models, especially for low-resource languages or domains reduces this possibility.

B. **Linguistic Diversity:** Difficulty in capturing nuances, variations, and styles of natural language, such as idioms, sarcasm, or humour.

C. **Domain Specificity:** Difficulty in adapting the summarization models to different domains or genres of text, such as legal, scientific, or literary texts.

D. **Scalability:** It is a performance challenge to build models that can process large-scale datasets and generate high-quality summaries in real time. Scalability is crucial for practical applications.

E. **Summary Diversity:** Generating summaries that are accurate, coherent, diverse, and engaging can be challenging.

IX. FUTURE SCOPES:

The future of machine learning in text summarization holds significant promise and is likely to see several exciting developments:

A. **Multimodal Summarization Integration:** The integration of machine learning models to summarize not only textual content but also multimedia data, including images, audio, and videos, is on the horizon. This extension into multimodal summarization will enable a more comprehensive understanding and summarization of diverse content types.

B. **Domain-Specific Summarization Tailoring:** The customization of machine learning models for specific domains, such as medical, legal, or scientific fields, is projected to result in more accurate and domain-specialized summarization, catering to the unique needs of professionals in these areas.

C. **Real-Time Summarization Capabilities:** The development of machine learning algorithms with the capability for real-time summarization will enable the summarization of live events, news, and social media content as it unfolds, providing up-to-the-minute insights.

D. **Customizable Summarization Solutions:** Machine learning models are expected to offer increased customization options, allowing users to tailor summaries to their preferences, aligning more closely with individual requirements and specific use cases.

E. **Advancements in Multilingual Summarization:** The advancement of multilingual models will facilitate the summarization of content in multiple languages, effectively bridging language barriers and expanding global access to information.

F. **Evolving Evaluation Metrics:** The development of more sophisticated evaluation metrics beyond conventional measures like ROUGE and BLEU is anticipated. These metrics will focus on assessing fluency, coherence, and informativeness more effectively, providing a more comprehensive assessment of text quality.

G. **Ethical Considerations and Fairness:** As machine learning models gain prominence in summarization, heightened attention to ethical concerns, including bias mitigation and fairness, will be necessary to ensure the ethical use of these technologies.

H. **Enhanced Explainability:** Efforts to enhance the interpretability and explainability of machine-generated summaries will be crucial, particularly in high-stakes domains such as legal and medical contexts, where transparency is paramount.

I. **Human-AI Collaboration:** Future developments are expected to encourage increased collaboration between humans and AI in the summarization process, with humans offering guidance and oversight to ensure the quality, relevance, and ethical standards of summaries.

J. **Summarization for Low-Resource Languages:** The application of machine learning will play a pivotal role in making summarization accessible for low-resource languages, preserving linguistic diversity, and enabling broader global dissemination of information.

K. **Privacy-Preserving Summarization:** Summarization models that respect privacy by not revealing sensitive information in summaries will become more prominent, particularly in fields like healthcare and finance where data privacy is paramount.

X. CONCLUSION:

Due to the rapid overload of information every day, there is a critical requirement for the summarization of the information available. This study offers a thorough survey of the wide range of text summarization methods, including hybrid, extractive, and abstractive methods. It also delves deeply into the intricate web of evaluation metrics that are used to assess the quality of summaries generated, emphasizing how challenging it is to gauge how effective summarising algorithms are.

The capacity of machine learning to acquire knowledge from large volumes of textual data has led to the advancement of increasingly complex and contextually-aware summarization methods. The search to extract important information from text has undergone a remarkable transformation, moving from conventional methods to deep learning-based models. These developments have made it possible to provide summaries that are more coherent and fluent in addition to being informative.

This study also makes clear how crucial it is to carefully consider data collecting and analysis. Through the analysis of several measures, ranging from ROUGE to BERTScore, we were able to understand several facets of high-quality summary analysis. To guarantee significant and trustworthy findings, researchers should carefully select study designs that are congruent with the particular goals of their data-gathering endeavour.

REFERENCES

[1] Ma, C., Zhang, W.E., Guo, M., Wang, H., Sheng, Q.Z.: Multi-document Summarization via Deep Learning Techniques: A Survey. ACM Comput. Surv. 55, 5, 37 pages (2023). https://doi.org/10.1145/3529754

[2] Sable, N.P., Rathod, V.U. (2023). Rethinking Blockchain and Machine Learning for Resource-Constrained WSN. In: Neustein, A., Mahalle, P.N., Joshi, P., Shinde, G.R. (eds) AI, IoT, Big Data and Cloud Computing for Industry 4.0. Signals and Communication Technology. Springer, Cham. https://doi.org/10.1007/978-3-031-29713-7_17.

[3] Liu, Y., Zhong, S.H., Li, W.: Query-Oriented Multi-Document Summarization via Unsupervised Deep Learning. AAAI, 26(1), 1699-1705 (2021).

[4] Liu, Y., Lapata, M.: Hierarchical Transformers for Multi-Document Summarization. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 5070–5081, Florence, Italy (2019). Association for Computational Linguistics.

[5] Song, S., Huang, H., Ruan, T.: Abstractive text summarization using LSTM-CNN based deep learning. Multimed Tools Appl 78, 857–875 (2019). https://doi.org/10.1007/s11042-018-5749-3

[6] Nilesh P. Sable, Vijay U. Rathod, Parikshit N. Mahalle, Jayashri Bagade, Rajesh Phursule ; Internet of Things-based Smart Sensing Mechanism for Mining Applications, Industry 4.0 Convergence with AI, IoT, Big Data and Cloud Computing: Fundamentals, Challenges and Applications IoT and Big Data Analytics (2023) 4: 132. https://doi.org/10.2174/9789815179187123040012.

[7] Day, M.Y., Chen, C.Y.: Artificial Intelligence for Automatic Text Summarization. In: 2018 IEEE International Conference on Information Reuse and Integration (IRI), 478-484, Salt Lake City, UT, USA (2018). https://doi.org/10.1109/IRI.2018.00076

[8] V. U. Rathod and S. V. Gumaste, “Role of Deep Learning in Mobile Ad-hoc Networks”, IJRITCC, vol. 10, no. 2s, pp. 237–246, Dec. 2022.

[9] Rezaei, A., Dami, S., Daneshjoo, P.: Multi-Document Extractive Text Summarization via Deep Learning Approach. In: 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), 680-685, Tehran, Iran (2019). https://doi.org/10.1109/KBEI.2019.8735084

[10] N. P. Sable, V. U. Rathod, P. N. Mahalle, and P. N. Railkar, “An Efficient and Reliable Data Transmission Service using Network Coding Algorithms in Peer-to-Peer Network”, IJRITCC, vol. 10, no. 1s, pp. 144–154, Dec. 2022.

[11] Munot, N., Govilkar, S.: Comparative Study of Text Summarization Methods. International Journal of Computer Applications 102, 33-37 (2014). https://doi: 10.5120/17870-8810.

[12] N. P. Sable, R. Sonkamble, V. U. Rathod, S. Shirke, J. Y. Deshmukh, and G. T. Chavan, “Web3 Chain Authentication and Authorization Security Standard (CAA)”, IJRITCC, vol. 11, no. 5, pp. 70–76, May 2023.

[13] Shirwandkar, N.S., Kulkarni, S.: Extractive Text Summarization Using Deep Learning. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 1-5, Pune, India (2018). https://doi.org/10.1109/ICCUBEA.2018.8697465

[14] Chopade, H.A., Narvekar, M.: Hybrid Auto Text Summarization Using Deep Neural Network and Fuzzy Logic System. In: 2017 International Conference on Inventive Computing and Informatics (ICICI), 52-56, Coimbatore, India (2017). https://doi.org/10.1109/ICICI.2017.8365192

[15] Vijay U. Rathod* & Shyamrao V. Gumaste, “Effect Of Deep Channel To Improve Performance On Mobile Ad-Hoc Networks”, J. Optoelectron. Laser, vol. 41, no. 7, pp. 754–756, Jul. 2022.

[16] Allahyari, M., Pouriyeh, S., et al.: Text Summarization Techniques: A Brief Survey. International Journal of Advanced Computer Science and Applications, 8(10) (2017).

[17] Rathod, V.U. and Gumaste, S.V., 2022. Role of Neural Network in Mobile Ad Hoc Networks for Mobility Prediction. International Journal of Communication Networks and Information Security, 14(1s), pp.153-166.

[18] MalarSelvi, G., Pandian, A.: Analysis of Different Approaches for Automatic Text Summarization. In: 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), 812-816, Erode, India (2022). https://doi.org/10.1109/ICCMC53470.2022.9753732

[19] Li, S., Lei, D., Qin, P., Wang, W.Y.: Deep Reinforcement Learning with Distributional Semantic Rewards for Abstractive Summarization. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 6038-6044, Hong Kong, China (2019). Association for Computational Linguistics.

[20] Y. Mali, Vijay U. Rathod "A Comparative Analysis of Machine Learning Models for Soil Health Prediction and Crop Selection", International Journal of Intelligent Systems and Applications in Engineering (IJISAE), vol. 11, no. 10s, pp. 811–828, Aug. 2023.

[21] Malik, R., et al.: Maximal gSpan: Multi-Document Summarization through Frequent Subgraph Mining. In: Proceedings of the 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM).

[22] Walunjkar, G., Birari, D., Desai, R., Gholap, Y., Gadekar, D. P., & Dandavate, A. (2024). Energy Efficient Routing in Group and Disaster Mobility: Comparative Analysis. International Journal of Intelligent Systems and Applications in Engineering, 12(4s), 85-94.

[23] Awasthi, I., et al.: Natural Language Processing (NLP) based Text Summarization - A Survey. In: Proceedings of the Sixth International Conference on Inventive Computation Technologies [ICICT 2021], IEEE Xplore Part Number: CFP21F70-ART; ISBN: 978-1-7281-8501-9

[24] Padmakumar, A., Saran, A.: Unsupervised Text Summarization Using Sentence Embeddings (2016).

[25] Fang, X., Xiao, T., Liu, S.: Research on Text Summarization Generation and Classification Method Based on Deep Learning. In: 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 757-762, Dalian, China (2022). https://doi.org/10.1109/ICAICA54878.2022.9844483

[26] Janjanam, P., Reddy, C.P.: Text Summarization: An Essential Study. In: 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 1-6, Chennai, India (2019). https://doi.org/10.1109/ICCIDS.2019.8862030

[27] Siddiqui, H., Siddiqui, S., Rawat, M., Maan, A., Dhiman, S., Asad, M.: Text Summarization using Extractive Techniques. In: 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 28-31, Greater Noida, India (2021). https://doi.org/10.1109/ICAC3N53548.2021.9725501

[28] N. P. Sable, V. U. Rathod, M. D. Salunke, H. B. Jadhav, R. S. Tambe, and S. R. Kothavle, "Enhancing Routing Performance in Software-Defined Wireless Sensor Networks through Reinforcement Learning", International Journal of Intelligent Systems and Applications in Engineering (IJISAE), vol. 11, no. 10s, pp. 73–83, Aug. 2023.

[29] Ramirez-Orta, J., Milios, E.E.: Unsupervised Document Summarization Using Pre-trained Sentence Embedding's and Graph Centrality. In: Proceedings of the Second Workshop on Scholarly Document Processing (2021)

[30] Ma, C., et al.: Abstractive Multi-Document Summarization via Phrase Selection and Merging. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 4810-4819 (2019).

[31] Vijay U. Rathod, Yogesh Mali, Nilesh Sable, Deepika Ajalkar, M. Bharathi, and N. Padmaja," A Network-Centred Optimization Technique for Operative Target Selection", Journal of Electrical Systems (JEs), vol. 19, no. 2s, pp. 87–96, 2023.

[32] Fabbri, A.R., et al.: Multi-news: A Large-scale Multi-document Summarization Dataset and Abstractive Hierarchical Model. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 1075-1085 (2019).

[33] Narayan, S., Papasarantopoulos, N., Lapata, M., Cohen, S.B.: Neural Extractive Summarization with Side Information. ArXiv abs/1704.04530 (2017).

[34] Vaswani, A., et al.: Attention Is All You Need. In: Advances in Neural Information Processing Systems, 30 (2017).

[35] Zhang, Y., et al.: Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 4055-4065 (2019).

[36] N. P. Sable, V. U. Rathod, R. Sable and G. R. Shinde, "The Secure E-Wallet Powered by Blockchain and Distributed Ledger Technology," 2022 IEEE Pune Section International Conference (PuneCon), Pune, India, 2022, pp. 1-5, doi: 10.1109/PuneCon55413.2022.10014893.

[37] Liu, Y., et al.: Single Document Summarization as Tree Induction. In: Proceedings of NAACL-HLT 2019, 1745–1755, Minneapolis, Minnesota (2019).

[38] Gokhan, T., et al.: GUSUM: Graph-Based Unsupervised Summarization Using Sentence Features Scoring and Sentence-BERT. In: Text-Graphs 16: Graph-based Methods for Natural Language Processing, 44–53, October 16, 2022.

[39] Villavicencio, P., Watanabe, T.: Text Summarization of Single Documents Based on Syntactic Sequences. In: Intelligent Interactive Multimedia Systems and Services. Smart Innovation, Systems and Technologies, 11 (2011). https://doi.org/10.1007/978-3-642-22158-3_31

[40] . U. Rathod and S. V. Gumaste, "Role of Routing Protocol in Mobile Ad-Hoc Network for Performance of Mobility Models," 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, pp. 1-6, doi: 10.1109/I2CT57861.2023.10126390.

[41] V. U. Rathod and S. V. Gumaste, "An Effect on Mobile Ad-Hoc Networks for Load Balancing Through Adaptive Congestion Routing," 2023 International Conference on Integration of Computational Intelligent System (ICICIS), Pune, India, 2023, pp. 1-5, doi: 10.1109/ICICIS56802.2023.10430257.

[42] Rahul, et al.: NLP based Machine Learning Approaches for Text Summarization. In: Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020), IEEE Xplore Part Number: CFP20K25-ART; ISBN: 978-1-7281-4889-2

[43] Mali, Y., Rathod, V. U., Kulkarni, M. M., Mokal, P., Patil, S., Dhamdhere, V., & Birari, D. R. (2023). A Comparative Analysis of Machine Learning Models for Soil Health Prediction and Crop Selection. International Journal of Intelligent Systems and Applications in Engineering, 11(10s), 811-828.

[44] Lee, S., et al.: A Survey on Evaluation Metrics for Machine Translation. Mathematics, 11(4), 1006 (2023). https://doi.org/10.3390/math11041006

[45] Wang, C., Long, L.: HowNet-based Evaluation for Chinese Text Summarization. In: 2008 International Conference on Natural Language Processing and Knowledge Engineering (2008). https://doi.org/10.1109/nlpke.2008.4906789

[46] Devlin, J., et al.: BERT: Bidirectional Encoder Representations from Transformers. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171-4186 (2019).

[47] Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: A Method for Automatic Evaluation of Machine Translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02), 311–318 (2002). https://doi.org/10.3115/1073083.1073135

[48] Birari, D., Hiran, D., & Narawade, V. (2022). Constraint and Descriptor Based Image Retrieval through Sketches with Data Retrieval using Reversible Data Hiding. International Journal of Intelligent Systems and Applications in Engineering, 10(1s), 409-417

[49] Post, M.: A Call for Clarity in Reporting BLEU Scores. In: Proceedings of the Third Conference on Machine Translation: Research Papers, 186–191 (2018).

[50] Wolf, T., et al.: Transformers: State-of-the-Art Natural Language Processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 38–45 (2020).

[51] Radford, A., et al.: Language Models are Unsupervised Multitask Learners. (2019).

[52] Raffel, C., et al.: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. J. Mach. Learn. Res. 21, 1, Article 140 (2020).

[53] Lewis, M., et al.: BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 7871–7880 (2020).

[54] Zhang, J., et al.: PEGASUS: Pre-training with Extracted Gap-Sentences for Abstractive Summarization. In: Proceedings of the 37th International Conference on Machine Learning (ICML'20), Vol. 119. JMLR.org, Article 1051, 11328–11339 (2020).

[55] N. P. Sable, V. U. Rathod, P. N. Mahalle and D. R. Birari, "A Multiple Stage Deep Learning Model for NID in MANETs," 2022 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2022, pp. 1-6, doi: 10.1109/ESCI53509.2022.9758191.

[56] Sorte, B.W., Joshi, P.P., & Jagtap, V.S. (2015). Use of Artificial Intelligence in Software Development Life Cycle: A state of the Art Review.

[57] Jagtap, V. S., & Pawar, K. (2013). Analysis of different approaches to sentence-level sentiment classification. International Journal of Scientific Engineering and Technology, 2(3), 164-170.

[58] V. U. Rathod, N. P. Sable, N. N. Thorat and S. N. Ajani, "Deep Learning Techniques Using Lightweight Cryptography for IoT Based E-Healthcare System," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-5, doi: 10.1109/CONIT59222.2023.10205808.

[59] Pallavi Parlewar,et.al, “Depth Estimation using FCNN”, Biosc.Biotech.Res.Comm. Special Issue Vol 13 No 14 (2020) pp-372-375.

[60] Kumar, S. ., Parlewar, P. ., & Jagtap, V. . (2023). Design of an Ensemble learning Model to Recognize Multilingual Hand Gestures via Interlinked Key Point analysis. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(10s), 103–112. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/3238