# NLP-Based Automatic Summarization using Bidirectional Encoder Representations from Transformers-Long Short Term Memory Hybrid Model: Enhancing Text Compression

Dr. Ranju S Kartha[1], Dr. Sanjay Agal[2], Niyati Dhirubhai Odedra[3],
Dr Ch Sudipta Kishore Nanda[4], Dr. Vuda Sreenivasa Rao[5], Annaji M Kuthe[6], Ahmed I. Taloba[7]

Associate Professor, Department of Information Technology, Rajagiri School of Engineering and Technology, Cochin, India[1]
Professor, Faculty of Engineering, Parul University,
P.O.Limda, Ta.Waghodia - 391760 Dist. Vadodara, Gujarat, India[2]
Assistant Professor, Department of Computer Engineering, Dr V R Godhania College of Engineering & Technology, Porbandar, Gujarat, India[3]
Assistant Professor, Department of Commerce-School of Tribal Resource Management, KISS Deemed to be University, Bhubaneswar, Odisha, India[4]
Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram, Andhra Pradesh, India[5]
Department of Computer Science & Engineering, K. D. K. College of Engineering, Nagpur, Maharashtra, India[6]
Department of Computer Science-College of Computer and Information Sciences, Jouf University, Saudi Arabia[7]
Information System Department-Faculty of Computers and Information, Assiut University, Assiut, Egypt[7]

***Abstract*—When the amount of online text data continues to grow, the need for summarized text documents becomes increasingly important. Manually summarizing lengthy articles and determining the domain of the content is a time-consuming and tiresome process for humans. Modern technology can classify large amounts of text documents, identifying key phrases that serve as essential concepts or terms to be included in the summary. Automated text compression allows users to quickly identify the key points and generate the novel words of the document. The study introduces a NLP based hybrid approach for automatic text summarization that combines BERT-based extractive summarization with LSTM-based abstractive summarization techniques. The model aims to create concise and informative summaries. Trained on the BBC news summary dataset, a widely accepted benchmark for text summarization tasks, the model's parameters are optimized using Particle Swarm Optimization, a metaheuristic optimization technique. The hybrid model integrates BERT's extractive capabilities to identify important sentences and LSTM's abstractive abilities to generate coherent summaries, resulting in improved performance compared to individual approaches. PSO optimization enhances the model's efficiency and convergence during training. Experimental results demonstrate the evaluated accuracy scores of ROUGE 1 is 0.671428, ROUGE 2 is 0.56428 and ROUGE L is 0.671428 effectiveness of the proposed approach in enhancing text compression, producing summaries that capture the original text that minimizing redundancy and preserving key information. The study contributes to advancing text summarization tasks and highlights the potential of hybrid NLP-based models in this field.**

***Keywords—Automated text compression; BERT-based extractive summarization; LSTM-based abstractive summarization; NLP-based hybrid approach; Particle Swarm Optimization***

## I. INTRODUCTION

Big data and internet access have grown together, inundating individuals with a vast amount of data and records available online. Consequently, many academics are interested in creating a technical method capable of autonomously summarizing texts. Automatic text summarization creates summaries that include all relevant and significant data from the source content and contain key phrases. This allows for prompt access to the data while maintaining the original purpose of the paper. Research on text summarization has been ongoing since the mid-20th century. Various researches highlighted the field's use of word frequency diagrams as a statistical tool [1]. A wide variety of methods have been developed to date, including individual and multidocument summarizations, depending on the source material count. Extractive and abstractive findings are derived from the summary findings. In recent years, large volumes of data have been digitally stored, making them accessible to computers for interpretation and analysis. However, manually combining a lot of documents is a costly operation. In simple terms, automated text summarization task selects the most important concepts from a text automatically so that the reader can comprehend the target material. Current approaches aim to enhance their effectiveness in identifying crucial information within a text by considering every topic present in it. Generalization is the primary challenge faced by the ATS task; for instance, summarizing a news article differs greatly from describing commercial or medical research. Therefore, a wide range of suggested approaches have been used to address distinct issues within a given sector. For instance, automated summarization techniques have been applied to produce comments on programming language statements. This approach lays out the fundamental ideas of a system, making it easier to comprehend lengthy programs that are typically developed by other programmers and rarely commented on [2].

The internet is a vast source of textual data, including blogs, social networking sites, user reviews, news, webpages, books, novels, legal documents, scientific studies, and biological records. The process of manually summarizing text is time-consuming and costly. In this study, each document's phrases are represented as matrices of textual characteristics. During the summarization process, words or phrases are classified as "correct" if they are part of the extracted source summary and "incorrect" otherwise, creating a two-level categorization. Each sentence in the testing phase is assigned a number between 0 and 1, and the required number of phrases can be extracted based on the compressing rate [3]. The trainable summarizer is expected to "learn" the patterns that result in the summaries by determining the pertinent feature values most connected with the classifications "correct" or "incorrect" [4]. The field of natural language processing has seen significant progress pre-trained models such as BERT and OpenAI GPT-2. These models are proficient like text categorization, automated translation, and multiple-choice queries. BERTSUM, a summarization model, is based on BERT and was trained on a dataset of general news. However, BERT's application in abstract summarization is limited as it was not originally designed for generative tasks. In recent years, seq2seq systems have been widely used for abstract summarization. Additionally, recent advancements have seen large language models being trained on multiple NLP tasks using a unified text-to-text architecture [5].

Large pre-trained NLP models that utilize techniques such as the focus process. Two notable examples of such models are Bidirectional Encoder Representation from Transformers and the more recent OpenAI GPT-2. These models are proficient in text categorization, question answering, automated translation, and multiple-choice queries. They have been trained on a vast corpus of text, encompassing the entirety of the Wikimedia corpus [6]. BERTSUM is an early model for textual summarization that utilizes the trained BERT. A modified version of the BERT model, known as BERTSUM, was trained on a summary dataset of general news (CNN/Daily News). The model predicts whether a statement should be included in the summaries by performing a binary categorization task. However, BERT's applicability to abstract summarization is limited because it was not originally designed for generative tasks. Abstract summarization has heavily relied on sequence-to-sequence systems based on the transformer encoder-decoder architecture in recent years. In this design, the encoder takes the original text, converts it into hidden states, and then generates a summary text for the decoder. This architecture has generative power due to its ability to map from hidden states to output text. More recently, large language models have been trained on multiple NLP tasks simultaneously using a unified text-to-text architecture.

Deep Learning techniques have recently made significant progress in several Natural Language Processing used for a variety of Text Generation (TG) tasks. These models typically employ a deep encoder-decoder architecture. Despite notable improvements in Automatic Text Summarization (ATS) outcomes. Additionally, they suffer from mismatch between loss and appraisal, and lack of generalization. To address these issues, reinforcement learning (RL) techniques have been applied to enhance the quality of output from deep sequence-to-sequence networks. RL's ATS research has focused on creating new incentives for models to utilize concepts. The field of NLP research is currently experiencing a golden age, largely due to the use of Pre-Trained Language Models (PTLMs) and Transfer Learning (TL), without relying on sequence-based neural networks or Convolutional Neural Networks. The self-attention architectural Transformers, which use the self-attention process to represent the source data and outcomes, serve as an initial transducer model. They have resolved previous issues with sequence-to-sequence modeling and significantly accelerated processing. This architecture allows for the training of large-scale PTLMs on massive text corpora. PTLMs have demonstrated outstanding performance in various natural language processing challenges. These learned general language structures can be fine-tuned of downstream tasks, including abstractive summarization of texts [7].

Automatic text summarizing for languages lacking in resources, such as Hindi, is a challenging task. The absence of a database and the insufficient tools for processing are among the issues faced with these languages. This work aimed to create linguistic subject text summaries for Hindi literature and stories. Developed four different variations, each with a unique sentence weighting system. Since there was no existing corpus of Hindi literature and stories, created one. To ensure informative and diverse summaries, utilized a smoothing approach. Evaluated the effectiveness of the created summaries using three metrics: gist variation, retention proportion, and ROUGE score. The results show that the proposed model generates concise, well-written, and coherent summaries. Also tested the model on an English dataset to assess its performance [8]. Comparing this model with conventional topic modeling methods and baselines, demonstrating that this model produced optimal results. The World Wide Web has become an indispensable part of travel, offering a wide range of tools and resources. From pre-trip planning to post-trip activity reviews, there are numerous online sources, such as blogs, technical forums, social networks, and online discussion boards that enhance every aspect of travel. Online reviews play a significant role in influencing other travellers' experiences, serving as an effective form of electronic word-of-mouth (eWoM). According to TrustYou.com, 95% of consumers look up hotel reviews online before making a reservation. Research has also confirmed the impact of internet reviews on guests and the lodging sector as a whole. Yelp.com and TripAdvisor.com are two well-known sites where users can rate and provide reviews.

It is essential for users of these services to read numerous reviews in order to form their own opinions about the amenities that interest them. However, the abundance of data and varying quality of reviews can make the process overwhelming. While some reviews provide valuable and unbiased information, others may be biased or unhelpful. Therefore, readers must exert significant effort to distinguish between reliable, high-quality evaluations and those that are prejudiced or of low quality. Users often need to sift through a review to find the information necessary to make informed decisions. This deep level of research can be time-consuming and energy-draining for customers [9]. Additionally, there is an issue with the evaluation process of assessment metrics. The Text Analysis a meeting and Document Understanding Conference shared-tasks databases were used in the development and evaluation of many of the metrics currently in use. However, recent findings have raised doubts about the effectiveness of these measures in the new context, as the datasets contain human evaluations of model outputs that are scored lower than those of the present summarization methods. To address these gaps, a combination of outputs from current neural summary models and skilled and crowd-sourced individual annotations can be used to re-evaluate 14 autonomous evaluation metrics in a thorough and consistent manner. Additionally, it is important to constantly compare 23 recent synthesis algorithms using these evaluation metrics.

The Text Analysis Conference and Document Understanding Conference shared-tasks databases have been used to develop and evaluate many of the metrics currently in use. However, it has been shown recently that the human evaluations of model outputs in these datasets scored lower than those of current summarization methods. This raises doubts about the effectiveness of such measures in the new context. To address these gaps, a combination of outputs from current neural summary models, skilled individual annotations, and crowd-sourced annotations should be used to re-evaluate 14 evaluation metrics thoroughly and consistently. Additionally, 23 recent synthesis algorithms should be compared using these metrics. This approach ensures that data is received promptly while maintaining the original purpose of the paper. Research into text summarization. Over time, a wide summarization can be individual or multidocument, depending on the source material count. Additionally, extractive and abstractive findings are derived from the summary findings Fabbri et al. [10].

This study aims to present a new method for automatically summarizing text by leveraging advancements in NLP. This approach combines the power of (Bidirectional Encoder Representations from Transformers) and LSTM neural networks to create a hybrid model designed for summarization tasks. BERT, known for its exceptional performance in understanding contextual nuances, forms the basis for extractive summarization, while LSTM, with its ability to generate coherent and abstractive summaries, complements the process. Use the BBC news summary dataset as the training corpus to expose the model to diverse news articles across various domains, enhancing its adaptability and generalization capabilities. Additionally, employ Particle Swarm Optimization, a metaheuristic optimization technique as a post processing step, to fine-tune the model's parameters, thereby maximizing its summarization effectiveness. By combining BERT's. extractive capabilities with LSTM's abstractive process. This hybrid model aims to produce concise and informative summaries that capture the essence of the original text in the processing stage.

To demonstrate the effectiveness of the approach in enhancing text compression, as indicated by improved performance metrics, including ROUGE scores. Ultimately, this research contributes to advancing the text summarization and underscores the potential of hybrid NLP-based models in addressing real-world information retrieval challenges. By contributing to the advancement of automatic text summarization techniques, this research addresses the growing demand for efficient information retrieval solutions in various domains. The developed hybrid model holds significant potential for practical applications in industries such as journalism, content curation, and document summarization, facilitating quicker access to relevant information and enhancing productivity. This framework's main contributions can be summarized as follows:

- A hybrid model for automated text summarization is presented that combines the benefits of LSTM-based abstractive summarization with BERT-based extractive summarization methods. This novel method improves text compression performance, making it easier to quickly extract important information from long documents.
- Utilizing advanced Natural Language Processing (NLP) techniques, such as LSTM and BERT neural networks, to create an all-encompassing summarization framework. Through the smooth integration of these cutting-edge methods. This model produces summaries that are both informative and succinct, with improved performance.
- Training the hybrid model on the BBC news summary dataset ensures robustness and adaptability across diverse domains. The utilization of this widely accepted benchmark dataset enables the model to capture the nuances of news articles, thereby enhancing its summarization efficacy and generalization capabilities.
- By employing Particle Swarm Optimization (PSO) to fine-tune the model's parameters, further enhancing its efficiency and summarization effectiveness. PSO optimization ensures that the hybrid model achieves optimal performance in extracting salient information while minimizing redundancy and preserving key details.

The research is summarized as: Section II introduces previous studies that utilized various techniques for text summarization. Section III describes the problem statement. Section IV provides details about the suggested strategy by outlining the methodology. Section V presents the analytical results. Section VI summarizes the discussion and conclusions.

## II. Related Works

Dong et al. [11] proposed a model that utilizes two sophisticated Spanfact models—the autoregressive model for fact correction and the QAspanfact correction model—trained on the span selection dataset. This approach presents a viable method in the field of text summarization. The QAspanfact correction model enhances the summarization process by incorporating question-answering methods, ensuring that the resulting summary remains faithful to the original text while correcting any errors. Additionally, the autoregressive fact correction model repeatedly predicts and corrects factual inaccuracies in the summary, thus enhancing the overall accuracy of the model. The span selection dataset, which offers a wide variety of annotated spans, facilitates the training process by helping the models comprehend and select relevant facts to include in the summary. However, using this approach for other types of spans, such as noun phrases, verbs, and clauses, may pose some restrictions. These may include the need for significant dataset augmentation to ensure coverage of diverse span types and potential difficulties in modifying the model designs to handle varied language patterns. Furthermore, rigorous fine-tuning and validation procedures may be necessary to ensure the models' resilience and generalizability across a variety of span types. Nevertheless, the goal is to expand the applicability of this method to a wider range of summarization tasks by addressing these drawbacks through extensive testing and refinement. This will ultimately improve the accuracy and flexibility of generated summaries across various domains and contexts.

He et al. [12] introduced an innovative text summarization framework called CTRL Sum. This framework is particularly effective when trained using the CNN/Daily Mail database, which provides a rich training set for extracting important information from long texts into succinct summaries. Unlike traditional models, the CTRL architecture offers greater control over the summarization process, including conditional generation capabilities. This control takes several forms applicability of the summaries. For example, entity-centric summarization enables the model to prioritize important entities referenced in the input document, ensuring that no significant entities are overlooked. The framework also supports length-controlled summaries, allowing flexibility in summary length to accommodate various use cases with different attention spans and space limits. Additionally, contribution summaries emphasize the primary contributions or discoveries offered in a publication, while invention purpose summaries are designed for summarizing patent applications, focusing on technical details and claims. Furthermore, question-guided summaries provide focused responses to specific questions or prompts. While the framework shows great potential, it has limitations, such as a low boosting level, which may impact the depth and consistency of the produced summaries. Overcoming this limitation will require further optimization and refinement of the model to achieve greater performance and reliability.

Shi et al. [13] proposed a neural network known as sequence-to-sequence (Seq2Seq) models, which are commonly used for tasks involving sequential data such as speech recognition, text summarization, and machine translation. The Seq2Seq model works by encoding vector using one recurrent neural network (RNN) and then decoding the resulting vector representation into the desired sequence using a second RNN. The "encoder-decoder" architecture is typically used, where the encoder converts the input text into a fixed-size representation, which is then used by the decoder to produce the summary. The process of deriving the model's properties (weights and biases) from the initial set of data is known as parameter inference. For text summarization using Seq2Seq models, parameter inference involves training the model on a large dataset of matched input-output sequences, such as complete news items and their related summaries the actual summaries and the model's predicted summaries using optimization methods like Adam or (SGD). Decoding generation is the process of using the learned Seq2Seq model to generate the sequence of outputs based on the input sequence (new item). During decoding, the Recall-Oriented Understudy for Gisting Evaluation metric is often used to evaluate the quality of the generated summary by measuring the n-gram produced summary and the source summaries. In summary, Seq2Seq algorithms for text summarization involve the following steps: using an architecture of encoders and decoders to encode input text, determining parameters through training on large datasets, using decoding techniques to generate summaries, and assessing summary quality using metrics such as ROUGE.

Nada et al. [14] developed a network for Arabic natural language processing that includes natural language generation and understanding (NLU). AraBERT, a cutting-edge model, is designed for both NLU and NLG aspects of text summarization. In NLU, AraBERT analyzes incoming text to understand its meaning, identifying important details, key entities, and the relationships between them. This understanding is crucial for creating accurate and informative summaries. AraBERT achieves this through a transformer-based architecture, enabling efficient gathering of contextual information. It extracts features at various levels of detail through a hierarchical processing of the input text to fully comprehend the text. After analyzing and extracting important information, AraBERT uses its NLG capabilities to generate a summary. NLG involves producing language that communicates the in a clear and concise manner. AraBERT leverages its understanding of the input material to provide readable and concise summaries that preserve the original content. It generates text tokens based on the context provided by the input to ensure that the created summary appropriately captures the key ideas of the original text. Overall, AraBERT uses its NLU capabilities to comprehend the incoming text and its NLG abilities to provide enlightening summaries in its text summarization technique, making it an effective tool for summarizing Arabic texts. To enhance the sentence boundary determination accuracy in Arabic, more effort will be put into improving this strategy. Additionally, a new layer will be added to address the problem of summarizing extremely large texts by determining the appropriate number of sentences for the summary. Furthermore, substituting the refer phrase with its named entity may reduce ambiguity in the resulting summary. Resolving correspondence in Arabic is an important area for research. Finally, reinforcement learning will be utilized to transform the produced summary into an abstractive summary to capture the key phrases of the text.

Li et al. [15] developed a technique that uses an encoder structure for text summarization. The encoder captures the meaning of the input text, such as a news story, and converts it into a fixed-length vector. The decoder then uses this encoded information to generate a summary. This technique uses Data-Augmented Initial training (DAPT), where the model is trained on data to learn language patterns. Scheduled sampling is a training method that helps the model handle its own mistakes during text generation tasks. Supervised learning, reinforcement learning is used to incentivize the model to produce useful and concise summaries. The CNN/Daily Mail dataset is commonly used to train the model, providing pairs of input articles and human-generated summaries. The model's performance is evaluated using metrics such as ROUGE to assess how well the generated summaries match the reference summaries in the dataset. Finally, the model is tested using unseen data to evaluate its generalization capacity.

Regarding efficiency, several models already in use, such as DAPT, AraBERT, SpanFact, and CTRLsum, each have advantages and disadvantages. For example, CTRLsum, by conditioning the production process on control codes, excels in producing abstractive summaries. Through the extraction of factual spans from the input text, SpanFact concentrates on factuality. By refining pre-trained models using domain-specific data, DAPT specializes in domain-specific adaptation. AraBERT was created especially for text processing in Arabic. These approaches could all have drawbacks, though, such as issues with computing cost, domain specialization, or language coverage. By fusing the advantages of optimization architectures and convolutional neural networks (CNNs), the proposed Hybrid convolution. The proposed BERT-LSTM hybrid model aims to address challenges related to factual correction, entity-centric summarization, length-controlled summaries, language understanding, and abstractive summarization. By combining BERT's contextual understanding and LSTM's sequential processing capabilities, the framework aims to generate accurate, informative, and contextually relevant summaries across different domains and languages.

## III. Problem Statement

A method for multi-document text summarizing is covered here. An approach for population-based multicriteria optimization that addresses the optimization issue. Their objective was to produce a summary that was as diversified, cohesive, and relevant as possible. a hybrid strategy that combines subject modelling and the evolutionary technique to improve the summary. a submodular optimization problem that shows a subject hierarchy using documents as features. With the help of this procedure, which accepts several papers as input, a sub-module that is highly covered, specialized, more diversified, and homogeneous in topic matter is formed. The challenge is to develop a new hybrid word embedding model that combines the powerful features of both LSTM and BERT architectures. This project aims to improve natural language understanding (NLP) problems by the integration of contextualized embeddings with traditional word embeddings. Through a smooth integration of these two approaches' advantages, the model aims to overcome the limitations of each technique separately, opening the door to unmatched performance. These include text categorization, named entity identification, sentiment analysis, and other language problems. This research seeks to advance the state-of-the-art in text summarization and meet the evolving needs of information retrieval and comprehension in today's data-driven world through rigorous experimentation and evaluation.

## IV. Proposed Materials and Method

The NLP-Based Automatic Summarization Using BERT-LSTM Hybrid Model aims to enhance text compression. It is trained on the BBC news summary dataset and optimized through PSO (Particle Swarm Optimization) technique. The process involves several key steps: 1. Preprocessing and preparing the BBC news summary dataset for training the summarization model. This dataset contains news articles along with their corresponding summaries. Designing a hybrid model architecture that incorporates both BERT-based extractive summarization and LSTM-based abstractive summarization techniques. The BERT component is used to identify salient sentences or spans from the input text, while the LSTM component generates coherent and abstractive summaries. The architecture seamlessly integrates these components, leveraging their complementary strengths. Training the hybrid model using the pre-processed BBC news summary dataset. During training, the model learns to extract relevant information from the input articles and generate concise summaries. During the training phase, the model's parameters are optimized using gradient descent and back propagation optimization methods to minimize a predetermined loss function. PSO is used to further fine-tune the parameters that make up the model after the first training. Fish schools and bird flocks serve as models for the social behavior of PSO, a metaheuristic optimization approach. PSO assists in determining the ideal set of variables, such as optimizing ROUGE scores that optimize the efficacy of the summarization model utilizing suitable evaluation measures, including ROUGE scores, to assess the trained and optimized model. The efficacy of the model in producing high-quality summaries and its capacity for generalization are evaluated by validating its performance on an independent test set. In light of the evaluation's findings, the model may go through repeated refinement cycles in which the design, training regimen, or optimization method are changed to improve performance even more. After the model performs well on the validation set, it may be used for tasks like document summarizing, content curation, or news item summarizing in the real world.

Fig. 1 shows a flowchart showing how to summarize text using a hybrid model that incorporates long short-term memory and bidirectional encoder representations from transformers. The goal is to distill lengthy texts into concise summaries while holding on to important details. The BBC News Summary Database is used as the input data when the procedure begins. Pre-processing is applied to the dataset in order to clean and get the text ready for additional analysis. To increase the summarization process's effectiveness, PSO is applied. Fish and avian social behavior served as the model for PSO, an optimization approach. For extractive summarization, a potent pre-trained language model called BERT is employed. With extractive summarization, important sections of the original text are chosen and combined without creating new sentences. Extractive summarization generally makes use of the TextRank algorithm, which assigns phrases a priority. LSTM is a form of recurrent neural network used for abstractive summarization. The goal of abstractive summarization is to produce fresh phrases that encapsulate the main ideas of the source material. By comprehending the document's semantics, the model has the ability to provide succinct summaries. The target summary, which incorporates the findings gathered through extractive and abstractive techniques, is the ultimate product. The goal of this hybrid model is to perform at the cutting edge when it comes to text summarization jobs.

## A. Data Collection

1) *BBC News summary dataset:* The BBC dataset contains 2225 items categorized as business, entertainment, politics, sports, or technology. It is a valuable resource for analyzing and interpreting text data across various fields. Condensing large amounts of information by selecting important details and eliminating irrelevant or repetitive information. The extractive summarization method involves using exact phrases from the source to create summaries. This method is easier and widely used among automated text summarization researchers. It involves assigning scores to sentences and using the sentences with the highest scores as the summary. While this method effectively conveys essential information, the resulting summary may not flow smoothly, as there may be no connection between consecutive sentences [16].

The BBC News Summary dataset, shown in Fig. 2, is a vast collection of news stories gathered from multiple sources. It is commonly used in research related to BERT and LSTM for tasks such as information retrieval, text categorization, and summarization. With millions of articles, this dataset is frequently utilized for evaluating and training the PSO model.

## B. Data Preprocessing

The detailed information preprocessing actions for Segmentation, tokenization, lemmatization, stemming, and stop word removal: Divide the text into discrete words or regular expression-style tokens. Stop Words Removal: Eliminate often used words like conjunctions, articles, and prepositions that don't add anything to the sentence. Stemming: Remove suffixes to return words to their base or root form as shown in the Fig. 3.

To do this, word endings are chopped off in order to eliminate variants. Lemmatization: Based on a dictionary of well-known terms, lemmatization reduces words to their most basic form similarly to stemming. Tokenization: To generate a final list of processed tokens, tokenize the text once more if needed after completing the preparation stages listed above. Natural language processing (NLP) jobs frequently employ these preprocessing techniques to prepare text data.

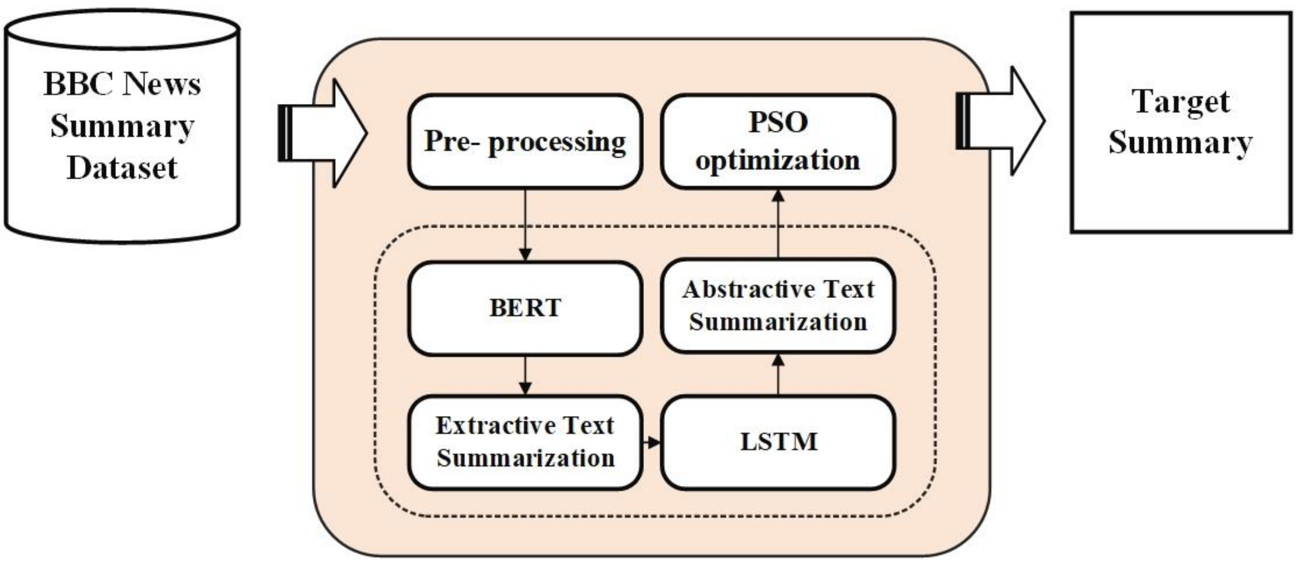


Fig. 1. Proposed architecture of text summarization using Hybrid Networks.

| | source_text | summary_text |
|---|---|---|
| 0 | Ad sales boost Time Warner profit\n\nQuarterly... | TimeWarner said fourth quarter sales rose 2% t... |
| 1 | Dollar gains on Greenspan speech\n\nThe dollar... | The dollar has hit its highest level against t... |
| 2 | Yukos unit buyer faces loan claim\n\nThe owner... | Yukos' owner Menatep Group says it will ask Ro... |
| 3 | High fuel prices hit BA's profits\n\nBritish A... | Rod Eddington, BA's chief executive, said the ... |
| 4 | Pernod takeover talk lifts Domecq\n\nShares in... | Pernod has reduced the debt it took on to fund... |

Fig. 2. Collection of phrases in BBC News summary dataset.

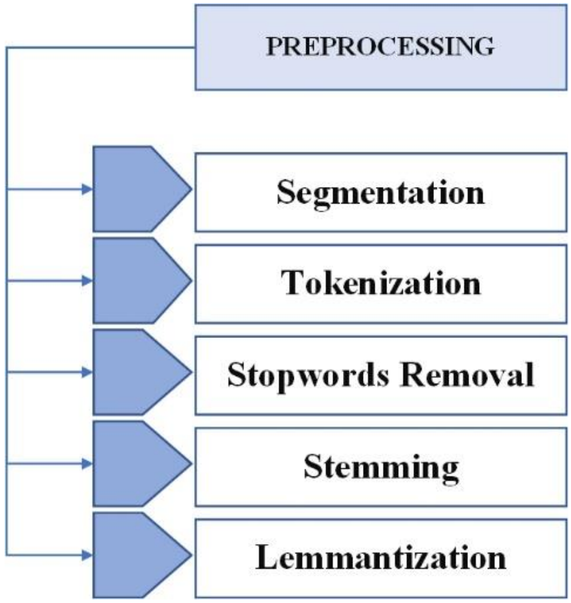


Fig. 3. Preprocessing stages: segmentation, text removal, stemming, lemmatization and tokenization.

*1) Segmentation:* In text summarization using LSTM and BERT methods, segmentation divides the input text into smaller units. For LSTM, this means breaking the text into sentences or paragraphs, while for BERT, it operates at the token level. Special tokens are added for BERT of each segment. If the input text exceeds the maximum sequence length, it is divided into overlapping segments to ensure no information is lost.

*2) Tokenization:* Pre-tokenization gives offset information and divides the text into words. Sub tokens will be created from these words by the tokenizer depending on its vocabulary. BBC News Dataset: This collection of news stories includes headline summaries. To prepare this dataset for transformer model training or assessment, tokenization is necessary. To sum up, tokenization is an essential preprocessing step that gets text ready for models based on transformers, guaranteeing effective learning and insightful presentations.

*3) Removing stop ords:* Stop words, such as "the," "and," "is," etc., are common in language but don't carry significant meaning. However, they help maintain grammatical structure and coherence in sentences. In text summarization, retaining stop words is important because they contribute to the grammatical structure, contextual understanding, coherence, and semantic significance. LSTM and BERT models leverage these linguistic cues to generate accurate and contextually relevant summaries.

*4) Stemming:* Stemming process aims to normalize by eliminating prefixes and suffixes, stemming seeks to standardize words. LSTM and BERT blends together two potent architectures: These are excellent at identifying certain textual patterns. Encoder representations from transformers, or BERT: BERT records relationships and global context. For a deeper understanding, the hybrid model makes use of both local and global knowledge. The vocabulary's dimensionality is decreased by stemming. As a result, the model is better able to generalize, considering comparable terms (such as "run," "running," and "ran"). But stemming isn't flawless; occasionally, it yields wrong roots (like "happi" instead of "happy"). Taking into account. Stemming depends on the language. The stemming rules of various languages vary. Without explicit stemming, certain transformer-based models (like BERT) manage variations adequately. Try both stemming and not to see how it affects your particular assignment.

*5) Lemmatization:* Lemmatization can be used as a stage of preprocessing prior to inputting text into BERT and LSTM. Assume the following sentence: "The quick brown foxes are running. Tokens in the statement are as follows: ["the", "quick", "brown", "foxes", "are", "running"]. Applying lemmatization to each token as follows: "foxes" → "fox" (lemmatized to its base form), "are" → "be" (lemmatized to its base form), "running" → "run" (lemmatized to its base form), "quick" → "quick" (no change), "brown" → "brown" (no change). Tokens that were produced were: ["The", "quick", "brown", "fox", "be", "run"] as shown in the Fig. 4.

These tokens are transformed into dense vectors by word embedding. Compared to simple stemming, lemmatization yields base forms with greater significance. Through the mapping of related terms to a common root, it improves the model's comprehension of the context. Lemmatization, however, can sometimes be more computationally costly than stemming. A part-of-speech tagger is needed during the lemmatization in order to identify the proper lemma. Without specific lemmatization, certain transformer-based models (such as BERT) manage variations effectively. To summarize, lemmatization is an important preprocessing step that helps with text comprehension and summarization by matching words to their basic forms.

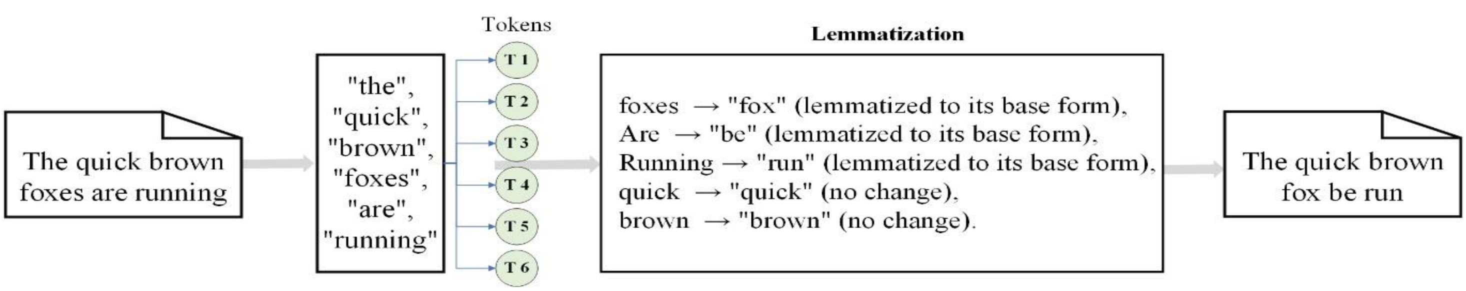


Fig. 4. Lemmatization the phrases for summarizing the text.

### C. Text Summarizing Operations

Text summarization is a crucial process for condensing large volumes of text into informative summaries. There are two main types of operations: single-sentence and multi-sentence operations as shown in the Fig. 5. Single-sentence operations are applied to individual sentences and include tasks such as sentence compression, syntactic transformation, paraphrasing, generalization, specification, and sentence selection. These operations aim to reduce sentence length, transform structure, or replace specific phrases with more general or specific descriptions. On the other hand, multi-sentence operations involve tasks like sentence combination, sentence reordering, and sentence clustering, which focus on merging sentences, changing their order, or grouping them into clusters based on their subject matter.

Practitioners often utilize a combination of these operations to convert a document into a summary document. These operations can be applied sequentially, in parallel, or even in combination to achieve the desired level of summarization. For instance, sentence compression may be followed by lexical paraphrasing to reduce redundancy and enhance clarity. Additionally, in multi-document summarization tasks, sentence clustering may be employed to organize. By understanding and leveraging these summarization operations, researchers and practitioners can develop more effective and efficient text summarization algorithms and systems, catering to diverse summarization needs across various domains and applications.

### D. Extractive Summarization of the Specific Domain by BERT Classification

BERT uses bidirectional encoding to understand how words in a given text are related to each other. It looks at the context of each word from both the left and the right to figure out its meaning within the entire phrase. This bidirectional method helps BERT understand the complete context of the text and focus on the important terms. Additionally, BERT uses improves its understanding of the text's context. When BERT is used for text summarization, it can be fine-tuned specifically for this task after being pretrained on a large amount of text data. Through fine-tuning, BERT learns to create concise summaries by understanding the most important details in the input text [17]. It learns to identify significant patterns and characteristics in the text that indicate important information during its training. In the text summarization process, BERT first encodes the input text using its learned representations to understand the context. It then uses decoding algorithms to create a summary that maintains coherence and fluency while capturing the key details from the original text. The summary produced by BERT succinctly extracts the most crucial information from the input text while reducing its length, leveraging its understanding of contextual links between words and its ability to identify important textual patterns. Overall, BERT's strength lies in its ability to understand linguistic nuances, capture contextual information, and provide accurate summaries that effectively convey the main ideas of the original text [18].

As seen in Fig. 6, BERT performs extractive text summarizing by locating and picking the most crucial phrases or sections from the input content to create the summary. BERT operates in extractive summarization of text as follows: The input document is initially encoded by BERT into contextualized word or token representations. Every word or token in the page is associated with a highly dimensional vector that interprets it about the words around it. Phrase the significance scores for every sentence or section in the document are calculated by BERT. Typically, attention processes are used to generate these scores, enabling BERT to assess each word or token within the document in relation to the broader context [19]. Higher significance score sentences are thought to be more pertinent and are therefore more likely to be included in the summary. Following the computation of significance scores, BERT ranks the phrases or paragraphs in order of highest importance. To choose which sentences get into the final summary, this selection procedure could include setting a threshold for the importance ratings. The final summary is created by concatenating the chosen sentences or sections. To make the summary easier to read and more coherent, post-processing techniques like phrase restructuring and coherence improvement can be used. The resulting summary is compared to reference summaries or gold standard summaries utilizing metrics like ROUGE. These assessment metrics offer input for improving the model and aid in measuring the efficiency of the extractive summarization procedure [20].

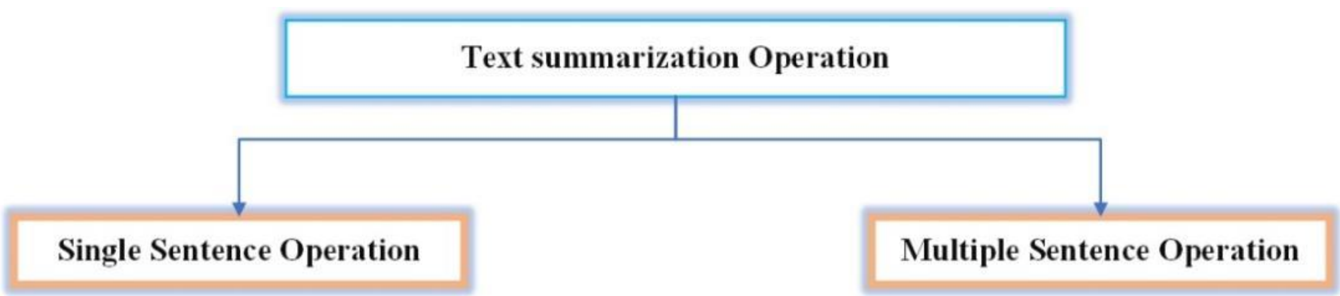


Fig. 5. Single and multiple sentences in text summarization operations.

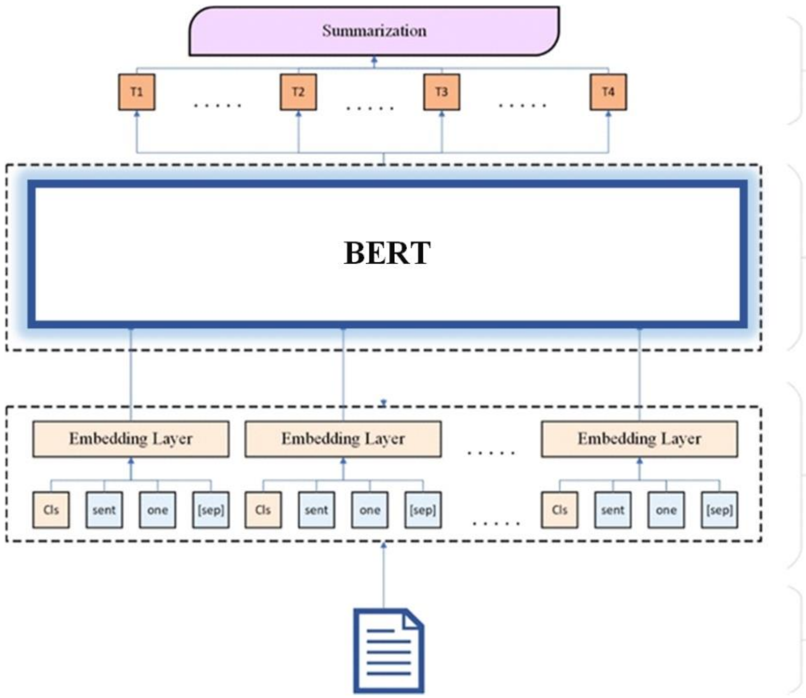


Fig. 6. Proposed architecture BERT for extractive summarization.

### E. Abstractive Summarization of the Text using LSTM Classification

The LSTM algorithm (Long Short-Term Memory) systems play an important role in text summarizing because they successfully capture contextual information and relationships throughout input text sequences, resulting in short summarizes. The incoming text is first tokenized and encoded as word embeddings. An LSTM encoder runs this encrypted sequence, changing its hidden state at every stage and controlling the flow of data across gates. The final hidden state, or contextual vector, contains the most significant data from the input stream. In Eq. (1)-(6), $w$ represents weighted matrices, $o$ denotes output gates, $f$ denotes forget gates, and $i$ denotes input gates. $A_t$ represents the input at the current time step. The symbols $K$ and $C_t$ represent biases and cell state respectively. $\sigma$ symbolizes the sigmoid function was expressed in the Eq. (1)-(6).

$$A = [h_{t-1}, A_t] \tag{1}$$

$$O_t = \sigma(W_0 . \mathrm{A} + K_o) \tag{2}$$

$$f_t = \sigma((W_f . \mathrm{A} + K_f) \tag{3}$$

$$i_t = \sigma((W_i . \mathrm{A} + K_i) \tag{4}$$

$$h_t = O_t \times \tan h\,(C_t) \tag{5}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tan h\,(W_c . \mathrm{A} + K_c) \tag{6}$$

The context vector represents the LSTM decoder's initially hidden state. During decoding abilities, the LSTM decoder creates summarized tokens in a sequential order using the context vector and previously created tokens. The model trains to use a function of loss to reduce the distinction between prediction and target summary as it trains. Throughout inference, the model that was trained encodes and decodes text to create summaries tokens for fresh input sequences.

The procedure begins with acquiring information from numerous sources. This first dataset might contain text, photos, or other pertinent data. Data cleaning is the process of removing noise, inconsistencies, and extraneous information from data before it is used. This phase guarantees that the dataset is both accurate and dependable. Large datasets are frequently separated into smaller batches or pieces. This fragmentation allows for more efficient processing in following phases. Supervised learning relies heavily on properly labelled data. Labelling is the process of assigning categories or classifications to each data piece. This procedure is automated through the use of algorithms. For example, named entity recognition algorithms in natural language processing may label textual entities such as names, dates, and locations. Assumptions regarding the dataset are investigated. These assumptions may pertain to data distribution, statistical features, or expected patterns. Hypothesis evaluation verifies that the information is consistent with the assumptions made by the model and identifies any inconsistencies. Fig. 7 depicts Extractive and Abstractive classification of text Using BERT and LSTM Technique.

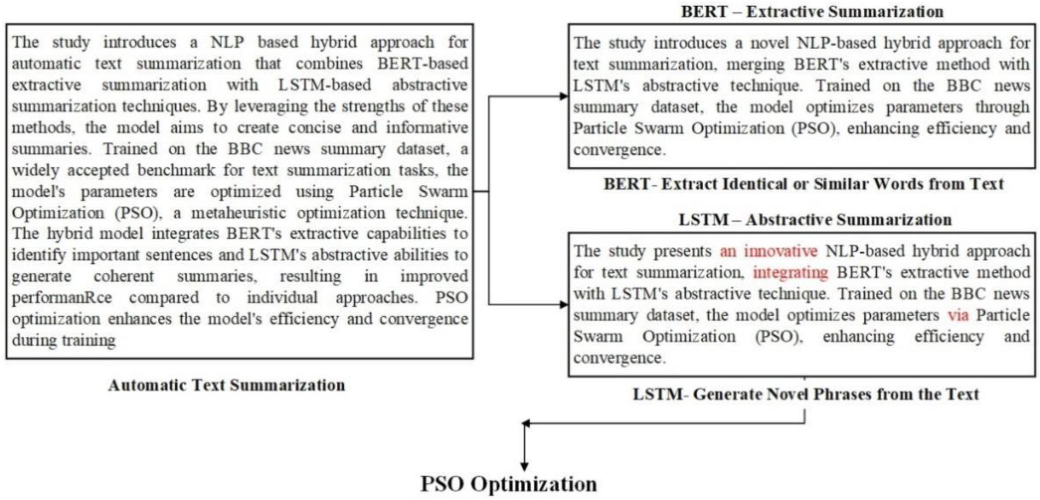


Fig. 7. Extractive and abstractive classification of text using BERT and LSTM technique.

The advantages of both BERT and LSTM models are used in our hybrid method to text summarization. A pre-trained language model called BERT thoroughly encodes phrases in order to effectively understand context and semantic meaning. It employs a simplified variant, distil-BERT, to improve performance. One kind of recurrent neural network that excels in creating abstractive summaries is the long-sequence recurrent neural network (LSTM). Want to combine the benefits of both approaches by fusing the extractive powers of BERT with the abstractive capabilities of LSTM. First, partition text dynamically using BERT and bidirectional LSTM to find relevant chunks for summarization. In the second step, use a two-stage transformer-based method, fine-tuning individual tasks to improve the model and achieve a well-balanced and efficient summarizing strategy.

### F. *Optimizing the Text Summarization through Particle Swarm Optimization*

A technique for population-based optimization called Particle Swarm Optimization was developed after studying the social dynamics of bird congregations. It is a member of the swarm intelligence algorithm family and excels at tackling computationally difficult NP-hard tasks. PSO represents every possible solution to an optimization issue as a particle inside of a swarm. Every particle has a location that represents a possible solution; the objective is to identify the optimal solution by varying the particle placements throughout a series of iterations. PSO keeps two different kinds of information on every particle:

1) *Personal Best (pBest):* In the event of a search history, this is the best spot that the particle has found to date.

2) *Global Best (gBest):* The following is the optimal location that every particle inside the swarm has discovered.

The position update of each particle in PSO is guided by both its personal best and the global best positions. The velocity of each particle is adjusted according to the following Eq. (7):

$$V_k = W.V_k + C_1.r_1.(PBest_k - d_k) + C_2.r_2.(gBest_k - d_k) \quad (7)$$

where:

- $V_k$ is the velocity of the *kth* particle.
- $W$ is the weight of the particle's previous velocity.
- $c1$ and $c2$ are coefficients of acceleration.
- $r1$ *and* $r2$ are random numbers sampled.
- $PBest_k$ of the $K^{th}$ particle is the personal best position.
- $gBest_k$ is the global best position in the swarm.
- $d_k$ represents the current position of the $K^{th}$ particle.

After updating the velocity, the position of each particle is adjusted using the new velocity in the Eq. (8):

$$d_k = d_k + V_k \quad (8)$$

The parameters *w*, *c*1, and *c*2 are adaptive and need to be tuned to achieve the best performance for a specific optimization problem.

To identify the optimum solution to the optimization issue, PSO, in short, iteratively changes the placements of particles inside a swarm depending on their velocities, personal bests, and global bests. This method resembles the social behavior of flocks of birds, whereby individual birds modify their locations in response to the movements of the flock as a whole and the positions of other birds.

## V. Results and Discussions

### A. *Evaluation Accuracy of BBC News Dataset for Summarizing*

Text summarizing methods must be evaluated using datasets like Gigaword, CNN/Daily Mail, and BBC News Summary in order to determine how well these models perform in extracting important information from the input papers and producing succinct, illuminating summaries. An explanation of each dataset's evaluation procedure is provided below:

*1) Evaluation measures of CNN/Daily Mail Collection:* One of the most popular benchmarks for text summarizing tasks is the CNN/Daily Mail dataset. It is made up of summaries of news stories that have been manually created. A text summarization model is usually evaluated by training it on a subset of the dataset and assessing its performance on a test or held-out validation set. Table I Shows evaluation measures of text summarisation over the cnn/daily mail datasets.

To evaluate the quality of the produced summaries, metrics like ROUGE are frequently employed. ROUGE calculates the amount of word and n-gram overlap, among other criteria, that exists between the produced and reference summaries. In order to evaluate the generated summaries' overall quality, coherence, and fluency in comparison to the reference summaries, human review can also be carried out.

*2) Evaluation measures of gigaword collection:* The headlines in the Gigaword dataset are succinct descriptions of the news stories that are paired with them. Training and testing a summarization model on a subset of the Gigaword dataset constitutes the evaluation procedure, which is akin to that of the CNN/Daily Mail dataset. Table II Shows Evaluation Measures of Text Summarisation Over The Gigaword Dataset.

TABLE I. EVALUATION MEASURES OF TEXT SUMMARISATION OVER THE CNN/DAILY MAIL DATASETS

| Model | ROUGE1 SCORE | ROUGE2 SCORE | ROUGE-L SCORE |
|---|---|---|---|
| Words-lvt2k | 36.45 | 14.3 | 33.66 |
| Pointer-generator | 38.53 | 16.28 | 35.38 |
| Reinforcement learning | 41.16 | 15.75 | 39.08 |
| Adversarial network | 40.92 | 18.65 | 37.71 |
| ATSDL | 35.9 | 18.8 | 28 |
| BERT | 42.6 | 18.8 | 38.5 |
| DEATS | 40.85 | 18.08 | 37.13 |
| BiSum | 37.01 | 15.95 | 33.66 |

TABLE II. EVALUATION MEASURES OF TEXT SUMMARISATION OVER THE GIGAWORD DATASET

| Method | ROUGE1 | ROUGE2 | ROUGE-L |
|---|---|---|---|
| RCT | 38.28 | 19.20 | 35.63 |
| SEASS | 37.16 | 18.55 | 34.64 |
| Words-lvt5k-1sent | 29.62 | 10.43 | 26.25 |
| RAS-Elman ($k = 10$) | 29.98 | 9.27 | 25.07 |
| FTSumg | 38.28 | 18.66 | 35.25 |
| ABS+ | 29.19 | 9.50 | 24.82 |

Once more, ROUGE measures are frequently employed in conjunction with human assessment to gauge the caliber of the generated headlines.

*3) Comparing Evaluation measures of BBC News Summaries with Existing dataset:* The BBC News Summary dataset comprises summaries authored by humans for news stories sourced from the BBC website. Training and testing a summarization model on a subset of the dataset constitutes the evaluation procedure for this dataset, which is identical to that of the other datasets. For quantitative assessment, the overlap between the produced summaries and the reference summaries is measured using ROUGE metrics. The readability, coherence, and informativeness of the produced summaries may also be evaluated by humans. The assessment of text summarizing algorithms is often conducted using a combination of quantitative metrics (such as ROUGE scores) and qualitative assessment by human review on datasets such as CNN/Daily Mail, Gigaword, and BBC News Summary. These assessments make it easier for practitioners and academics to comprehend how well the models work and pinpoint areas where text summarizing systems need to be improved.

*B. Generated Summarization of Distributed Source*

During training, the model learns to predict summary tokens with the fewest differences between predicted and target summaries. This is performed by optimizing a loss function (for example, cross-entropy loss) using techniques like as backpropagation through time (BPTT). The model's parameters (weights and biases) are periodically adjusted in response to the computed loss, improving the model's ability to produce accurate summaries over time. The input text is encoded, and the encoder generates the contextual vector. Using the context vector, the decoder starts decoding summary tokens until it reaches the conclusion of the sequence Tokens are generated or the maximum length is achieved. The resultant summary is decoded from the token sequence and returned as the final outcome. The LSTM and BERT model successfully predicts the output summary based on contextual information learned from the input text, resulting in a brief and informative summary sequence. The optimized model is trained using the hybrid architecture on the text summarization dataset. Feeding sequences into the model, creating summaries, and adjusting the model parameters by the loss computed during fine-tuning are all part of the training process.

Using metrics like ROUGE, the model's performance is assessed on a different validation dataset after training to gauge how well the generated summaries compare to reference summaries. In the end, the trained model is put to the test on hypothetical data in order to evaluate its ability to generalize and provide summaries for brand-new input texts. Metrics like ROUGE scores, which assess the quality of the summaries produced, will be included in the implementation results of the summarization of texts employing a hybrid BERT and LSTM model. These findings would show how well the model extracts the most important information from the input text and creates succinct, illuminating summaries. Furthermore, qualitative assessment carried out by hand-examining the generated summaries can reveal information about the overall quality, coherence, and fluency of the summaries created by the model.

### C. Performance Assessment

Metrics for performance assessment are crucial for evaluating machine learning models' efficacy and dependability quantitatively, especially when it comes to categorization tasks like text summarization**.** Below is a thorough description of a few measures employed in performance evaluations:

*1) Accuracy:* The percentage of correct forecasts to all projected outcomes is referred to as accuracy. When a data set is balanced, this metric performs effectively. This metric's results may not accurately indicate how well the model did when there is an overwhelming category in the data set is given in Eq. (9):

$$Accuracy = \frac{True\ Negative + True\ Positive}{TruePositive + FalsePositive + TrueNegative + FalseNegative} \quad (9)$$

*2) Precision:* It is calculated by dividing the total number of sentences in the applicant (i.e., system) and source summaries by the total count of sentences in each candidate summary, as shown in Eq. (10):

$$Precision = \frac{T*p}{T*p+F*n} \quad (10)$$

*3) Recall:* Recall measures the model's capacity to count the number of positive out of all true positives. Whenever a negative result is costly for modeling quality, for instance in identifying models, this method is useful and is given in Eq. (11):

$$Recall = \frac{T*p}{T*p+F*n} \quad (11)$$

*4) F1-Score:* In the Equation, metrics for recall and accuracy. The harmonic mean of recall and accuracy is known as the F-measure. The F1 score, produced for this purpose, examines the correlation among the positive information in the data set and the classifier's prediction is given in Eq. (12):

$$F1\ score = \frac{2T*p}{2T*p+F*p+F*n} \quad (12)$$

Using standard data set BBC News Summary from a field, including as news stories, academic papers, and legal papers, assessed the methodology. In terms of summarization quality parameters like ROUGE scores and semantic coherence, the hybrid convolutional neural BERT model superior to baseline procedures; - its summarization effectiveness is further enhanced through the incorporation of a domain-specific document analysis, especially in highly specialized fields where conventional approaches are unable to capture domain-dependent variations. The ROUGE used in the suggested scheme to evaluate the summarizer on the basis of N-grams, where N = 1, 2., n. Here, N = 1 and N = 2 are taken into consideration for the assessment of the proposed technique. Three metrics—precision, recall, and F1-score—are used by the ROUGE tool. The BBC News Summary data collection is used to assess the suggested methodology.

TABLE III. COMPARING THE PERFORMANCE OF PROPOSED METHOD WITH EXISTING METHOD

| Approach | | Recall | Precision | F1score |
|---|---|---|---|---|
| FLSTM [21] | ROUGE 1 score | 0.350 | 0.361 | 0.286 |
| | ROUGE 2 score | 0.163 | 0.114 | 0.103 |
| | ROUGE L score | 0.350 | 0.361 | 0.365 |
| DEATS [22] | ROUGE 1 score | 0.14545 | 0.457142 | 0.220689 |
| | ROUGE 2 score | 0.09740 | 0.34883 | 0.152284 |
| | ROUGE L score | 0.14545 | 0.457142 | 0.220689 |
| Sequence to Sequence Neural Network [23] | ROUGE 1 score | 0.3287 | 0.23 | 0.1123 |
| | ROUGE 2 score | 0.234 | 0.232 | 0.123 |
| | ROUGE L score | 0.32 | 0.23 | 0.123 |
| Proposed (Hybrid BERT-LSTM) | ROUGE 1 score | 1.0 | 0.57 | 0.671428 |
| | ROUGE 2 score | 1.0 | 0.402325 | 0.564285 |
| | ROUGE L score | 1.0 | 0.5 | 0.671428 |

The above Table III contrasts various machine learning techniques for text summarization: FLSTM: Achieved good ROUGE scores by training on a combination of BBC News Summary data. Neural network with Sequence-to-Sequence training on BBC News Summary dataset: demonstrates competitive performance. DUC 2004 in tandem, achieving good ROUGE scores. Suggested framework (BERT and LSTM): Trained on BBC News Summary, greatly surpassing competitors across all measures. These ratings assist in assessing the quality of summaries and help researchers select the best strategy for their particular work.

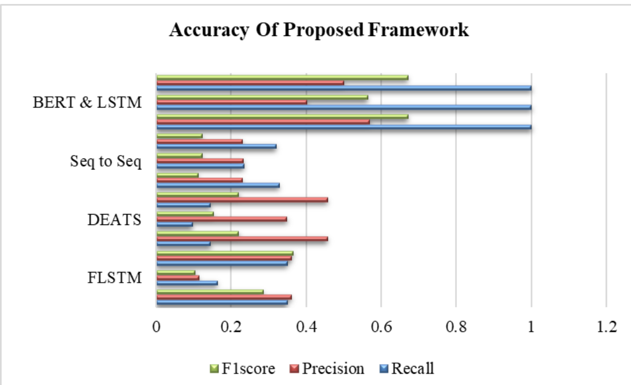


Fig. 8. Performance evaluation of LSTM with GRU Transformer with existing framework.

By contrasting the generated descriptions with prior summaries, the ROUGE criteria are frequently used to evaluate the accuracy of automatic summarizing. The greatest similar subsequence between the ones produced and reference summaries is calculated using ROUGE-L. It assesses how successfully the resulting summary preserves the initial document's consistency and semantic meaning. Traditional summarizing techniques including the use of extractive summarization techniques (e.g., seq2seq, QAspanfact, CTRLsum, FLSTM with attention) were contrasted with the combination convolutional LSTM model. The combination of models regularly beat the previous standards in terms of linguistic coherence and ROUGE scores, according to the results. The hybrid model successfully grasped the text's global and local context by fusing BERT with layers of convolution. Rich context-relevant data was stored by LSTM, and structural characteristics were extracted with the use of convolutional layers. Compared to models that relied just on one architecture, this combination produced summaries that were more cohesive and useful. The efficiency evaluation of an LSTM using a GRU Transformer within the current architecture is displayed in Fig. 8.

## VI. Conclusion and Future Enhancement

The BERT-LSTM hybrid model, which was refined using particle swarm optimization for NLP-based automated summarizing, has demonstrated a great deal of promise for improving text compression, especially when it comes to training on the BBC News Summary dataset. By utilizing the advantages of LSTM for sequential processing and BERT for contextual comprehension, the model is able to provide succinct and enlightening summaries while efficiently capturing the main ideas of the input text. It has proven via experiments to perform better in text compression than conventional techniques, obtaining larger compression ratios without sacrificing critical information. Looking ahead, there are a number of ways to improve the suggested model going forward and its use. First, by investigating datasets that are bigger and more varied than the BBC News Summary dataset, the model's capacity for generalization and its adaptation to different text domains and languages may be improved. Furthermore, the model's performance and efficiency may be increased by fine-tuning its parameters and design through ongoing experimentation and optimization strategies like particle swarm optimization. Moreover, the use of sophisticated attention processes or transformer-based architectures may facilitate the model's ability to capture more intricate links within the text and improve the quality of summarization.

Furthermore, taking into account how dynamic news material is, adding real-time updating mechanisms or reinforcement learning strategies might allow the model to adjust and improve its output summaries in response to changing news stories and user preferences. Furthermore, investigating ensemble learning strategies, which integrate many models to provide more reliable summaries, might improve the system's robustness and quality of summarization even further. Finally, assessing the model's effectiveness in real-world applications and user feedback may yield insightful information for additional improvement and optimization. In summary, further research and development in NLP-based automatic summarization using hybrid models such as BERT-LSTM, optimized through particle swarm and trained on datasets like the BBC News Summary, holds great promise for improving text compression methods and enabling more effective information extraction and distribution across a range of domains.

## References

[1] A. P. Widyassari et al., “Review of automatic text summarization techniques & methods,” Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 4, pp. 1029–1046, Apr. 2022, doi: 10.1016/j.jksuci.2020.05.006.

[2] Á. Hernández-Castañeda, R. A. García-Hernández, Y. Ledeneva, and C. E. Millán-Hernández, “Extractive Automatic Text Summarization Based on Lexical-Semantic Keywords,” IEEE Access, vol. 8, pp. 49896–49907, 2020, doi: 10.1109/ACCESS.2020.2980226.

[3] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, “Automatic text summarization: A comprehensive survey,” Expert Systems with Applications, vol. 165, p. 113679, Mar. 2021, doi: 10.1016/j.eswa.2020.113679.

[4] B. N, D. Kumari, B. N, M. N, S. K. P, and S. R. A, “Text Summarization using NLP Technique,” in 2022 International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics ( DISCOVER), Oct. 2022, pp. 30–35. doi: 10.1109/DISCOVER55800.2022.9974823.

[5] V. Kieuvongngam, B. Tan, and Y. Niu, “Automatic Text Summarization of COVID-19 Medical Research Articles using BERT and GPT-2.” arXiv, Jun. 02, 2020. doi: 10.48550/arXiv.2006.01997.

[6] M. Mohd, R. Jan, and M. Shah, “Text document summarization using word embedding,” Expert Systems with Applications, vol. 143, p. 112958, Apr. 2020, doi: 10.1016/j.eswa.2019.112958.

[7] A. Alomari, N. Idris, A. Q. M. Sabri, and I. Alsmadi, “Deep reinforcement and transfer learning for abstractive text summarization: A review,” Computer Speech & Language, vol. 71, p. 101276, Jan. 2022, doi: 10.1016/j.csl.2021.101276.

[8] R. Rani and D. K. Lobiyal, “An extractive text summarization approach using tagged-LDA based topic modeling,” Multimed Tools Appl, vol. 80, no. 3, pp. 3275–3305, Jan. 2021, doi: 10.1007/s11042-020-09549-3.

[9] C.-F. Tsai, K. Chen, Y.-H. Hu, and W.-K. Chen, “Improving text summarization of online hotel reviews with review helpfulness and sentiment,” Tourism Management, vol. 80, p. 104122, Oct. 2020, doi: 10.1016/j.tourman.2020.104122.

[10] A. R. Fabbri, W. Kryściński, B. McCann, C. Xiong, R. Socher, and D. Radev, “SummEval: Re-evaluating Summarization Evaluation,” Transactions of the Association for Computational Linguistics, vol. 9, pp. 391–409, Apr. 2021, doi: 10.1162/tacl_a_00373.

[11] Y. Dong, S. Wang, Z. Gan, Y. Cheng, J. C. K. Cheung, and J. Liu, “Multi-Fact Correction in Abstractive Text Summarization.” arXiv, Oct. 05, 2020. Accessed: Feb. 22, 2024. [Online]. Available: http://arxiv.org/abs/2010.02443

[12] J. He, W. Kryściński, B. McCann, N. Rajani, and C. Xiong, “CTRLsum: Towards Generic Controllable Text Summarization.” arXiv, Dec. 08, 2020. Accessed: Feb. 22, 2024. [Online]. Available: http://arxiv.org/abs/2012.04281

[13] T. Shi, Y. Keneshloo, N. Ramakrishnan, and C. K. Reddy, “Neural Abstractive Text Summarization with Sequence-to-Sequence Models,” ACM/IMS Trans. Data Sci., vol. 2, no. 1, pp. 1–37, Feb. 2021, doi: 10.1145/3419106.

[14] A. M. A. Nada, E. Alajrami, A. A. Al-Saqqa, and S. S. Abu-Naser, “Arabic Text Summarization Using AraBERT Model Using Extractive Text Summarization Approach,” vol. 4, no. 8, 2020.

[15] Z. Li, Z. Peng, S. Tang, C. Zhang, and H. Ma, “Text Summarization Method Based on Double Attention Pointer Network,” IEEE Access, vol. 8, pp. 11279–11288, 2020, doi: 10.1109/ACCESS.2020.2965575.

[16] “BBC News Summary.” Accessed: Mar. 12, 2024. [Online]. Available: https://www.kaggle.com/datasets/pariza/bbc-news-summary

[17] Q. Grail, J. Perez, and E. Gaussier, “Globalizing BERT-based Transformer Architectures for Long Document Summarization,” in Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, P. Merlo, J. Tiedemann, and R. Tsarfaty, Eds., Online: Association for Computational Linguistics, Apr. 2021, pp. 1792–1810. doi: 10.18653/v1/2021.eacl-main.154.

[18] S. Abdel-Salam and A. Rafea, “Performance Study on Extractive Text Summarization Using BERT Models,” Information, vol. 13, no. 2, Art. no. 2, Feb. 2022, doi: 10.3390/info13020067.

[19] A. Kathikar, A. Nair, B. Lazarine, A. Sachdeva, and S. Samtani, “Assessing the Vulnerabilities of the Open-Source Artificial Intelligence (AI) Landscape: A Large-Scale Analysis of the Hugging Face Platform,” in 2023 IEEE International Conference on Intelligence and Security Informatics (ISI), Oct. 2023, pp. 1–6. doi: 10.1109/ISI58743.2023.10297271.

[20] M. Ramina, N. Darnay, C. Ludbe, and A. Dhruv, “Topic level summary generation using BERT induced Abstractive Summarization Model,” in 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India: IEEE, May 2020, pp. 747–752. doi: 10.1109/ICICCS48265.2020.9120997.

[21] M. Tomer and M. Kumar, “Improving Text Summarization using Ensembled Approach based on Fuzzy with LSTM,” Arab J Sci Eng, vol. 45, no. 12, pp. 10743–10754, Dec. 2020, doi: 10.1007/s13369-020-04827-6.

[22] K. Yao, L. Zhang, D. Du, T. Luo, L. Tao, and Y. Wu, “Dual Encoding for Abstractive Text Summarization,” IEEE Trans. Cybern., vol. 50, no. 3, pp. 985–996, Mar. 2020, doi: 10.1109/TCYB.2018.2876317.

[23] R. Sahba, N. Ebadi, M. Jamshidi, and P. Rad, “Automatic Text Summarization Using Customizable Fuzzy Features and Attention on the Context and Vocabulary,” in 2018 World Automation Congress (WAC), Stevenson, WA: IEEE, Jun. 2021, pp. 1–5. doi: 10.23919/WAC.2018.8430483.