

A Novel Hybrid Approach for Biomedical Text Summarization Using Pre-Trained Language Models

Pawan Makhija, Research Scholar, Department of Computer Engineering, Institute of Engineering and Technology, DAVV, Indore, India (1st author)

Email: pawanmakhija@acropolis.in

Dr. Sanjay Tanwani, Professor, School of Computer Science & IT, DAVV, Indore, India (2nd Author)

Email: Sanjay_tanwani@hotmail.com

Abstract:

The rapid expansion of biomedical literature requires efficient methods for information extraction. This paper proposes a hybrid biomedical text summarization (BTS) system that leverages the strengths of multiple pre-trained language models (PLMs): BART, Pegasus, and T5. Unlike existing approaches that rely on a single PLM, our hybrid system combines the output of these models, employing a novel sentence scoring mechanism and subsequently removing redundant information using cosine similarity to enhance the quality and conciseness of the summaries. We evaluate our system on a subset of the PubMed summarization dataset using ROUGE scores, demonstrating a significant improvement over the performance reported in [1], the baseline paper. The inclusion of a sentence scoring mechanism and a redundancy removal step based on semantic similarity provides demonstrably better summaries. Using models trained on the clinical data provides better sentence extraction for the summaries generated. This approach provides a solution to fast forward research that faces the challenge of domain-specific terminology and complex sentence transformation for the text summarization of medical documents.

Keywords: *Natural Language Processing (NLP), BART, T5, Pegasus, Biomedical Text Summarization (BTS)*

1. Introduction:

Researchers and physicians looking to quickly obtain pertinent information face a major obstacle due to the biomedical literature's fast growth. Large amounts of text are time-consuming and impracticable to read and summarize by hand. A significant percentage of medical records lack succinct summaries, even though many of them have abstracts. An important remedy is provided by automated text summarization, especially abstractive summarization, which creates fresh, succinct summaries that encapsulate the essence of the source material.

Text summarization is among the many natural language processing (NLP) tasks that can be seen in the recent developments, producing strong pre-trained language models (PLMs) that can achieve state-of-the-art outcomes [2, 3, 4]. Although the biomedical field has effectively used these models [5, 6, 7], the majority of research focuses on a single PLM. By merging the outputs of several PLMs that are well-known for their efficacy in summarizing (BART [8], Pegasus [9], and T5 [10]), and then filtering for redundancy, this study suggests a novel hybrid technique. In order to generate more thorough and accurate summaries, this approach seeks to propose a novel sentence scoring mechanism along with the redundancy removal using semantic similarity measure that maximizes each model's unique strengths and minimize its limitations.

2. Literature Survey:

The research on summarization has been done for a long time to achieve human-level capability and accuracy in this task. This traces back to methods as simple as text extraction and sentence scoring using TF-IDF [11] and similar heuristics. These methods had no special information about any specialized domain, such as biomedical or any semantic similarity. This resulted in the poor performance of these systems on the specialized tasks. With the emergence of deep learning models, the potential of this field scaled with the foundation of BTS systems like RNN, which was followed by transformers [14]. Using pre-trained embedding such Word2Vec [15] or GloVe [16], early transformer-based models were refined on biomedical corpora [17].

Several studies [18] adapted BERT for summarizing biomedical text, fine-tuning it on domain-specific corpora, such as PubMed abstracts or clinical notes. These studies showed substantial performance gains over traditional methods.

More recently, research focused on specialized biomedical PLMs like BioBERT [21], SciBERT [22], and ClinicalBERT [23]. These models are pre-trained on large biomedical datasets, resulting in a richer understanding of the domain's unique terminology and patterns. BioBERT, for instance, has been used for various biomedical NLP tasks, including text summarization [24].

However, most previous works employ only one pre-trained language models (PLM) for the summarization of biomedical text. The effectiveness of different PLMs varies depending on factors such as architecture, pre-training data, and hyperparameters [2,3,18]. A hybrid approach offers the opportunity to combine the benefits of multiple models, mitigating individual weaknesses and achieving potentially superior performance [26]. Several studies have explored ensemble methods for text summarization; however, most don't tackle the nuances of the biomedical domain or don't specifically filter for redundant information. We have done a literature survey on the text summarization system that provides crucial information, e.g., type of summarization, models used for the task, with the dataset used and its key features. The tabular form of the literature survey is as follows:

Table 1. Literature Survey of Various Biomedical Text Summarization Methods

Method	Year	Type	Model(s) Used	Dataset(s)	Key Features
LexRank [33]	2004	Extractive	Unsupervised, graph-based ranking	Various	Lexical centrality for sentence ranking
Naive Bayes [34]	2010	Extractive	Supervised, Naive Bayes classifier	Various	Simple probabilistic model for sentence classification
SummaRuNNer [35]	2016	Abstractive	Recurrent Neural Network (RNN)	Various	Early deep learning approach for abstractive summarization
ABS [36]	2016	Abstractive	Recurrent Neural Network (RNN)	Various	Sequence-to-sequence model with attention mechanism
BERTsum [37]	2019	Extractive	BERT	Various	Fine-tuned BERT for sentence selection
BioBERTsum [38]	2020	Extractive	BioBERT	Various	Fine-tuned BioBERT, domain-specific for better understanding biomedical terms
BART [8]	2020	Abstractive	BART	Various	Encoder-decoder architecture, good for long sequences
Pegasus [9]	2020	Abstractive	Pegasus	Various	Masks sentence and learns to reconstruct them, strong performance
T5[10]	2020	Abstractive	T5	Various	Text-to-text transfer transformer, versatile and high-performing
ContinualBERT [41]	2020	Extractive	Continually trained BERT	COVID-19 Literature	Adaptive learning for online summarization
GenCompareSum[40]	2022	Hybrid	T5 (generative), BERT	PubMed, S2ORC	Uses T5 for key snippet generation, BERT for sentence scoring
RadBERT [39]	2022	Extractive	RadBERT (specialized for radiology reports)	Radiology reports	Pre-trained on radiology reports, improves performance on this specific task

3. Methodology:

Our proposed system adopts a hybrid approach, integrating three powerful PLMs known for their summarization capabilities: BART, Pegasus, and T5. The system operates in three stages:

3.1 Individual Summarization:

This stage involves generating independent summaries for the input biomedical text using each of the selected PLMs. We utilize the Hugging Face Transformers library [27] for seamless model access and efficient inference.

3.1.1 BART (Bidirectional and Auto-Regressive Transformers):

BART [2] is an encoder-decoder model pre-trained using a denoising autoencoding objective. As a result, it has developed a strong grasp of language structure and context by training to recreate distorted input text. Its autoregressive decoder makes it easier to create grammatically sound and cohesive summaries, and its bidirectional encoder efficiently records contextual information. When it comes to handling the intricate sentence structures frequently present in biological texts, BART is especially adept at producing summaries that are both fluid and grammatically accurate. Important contextual information is captured with the aid of its bidirectional encoder.

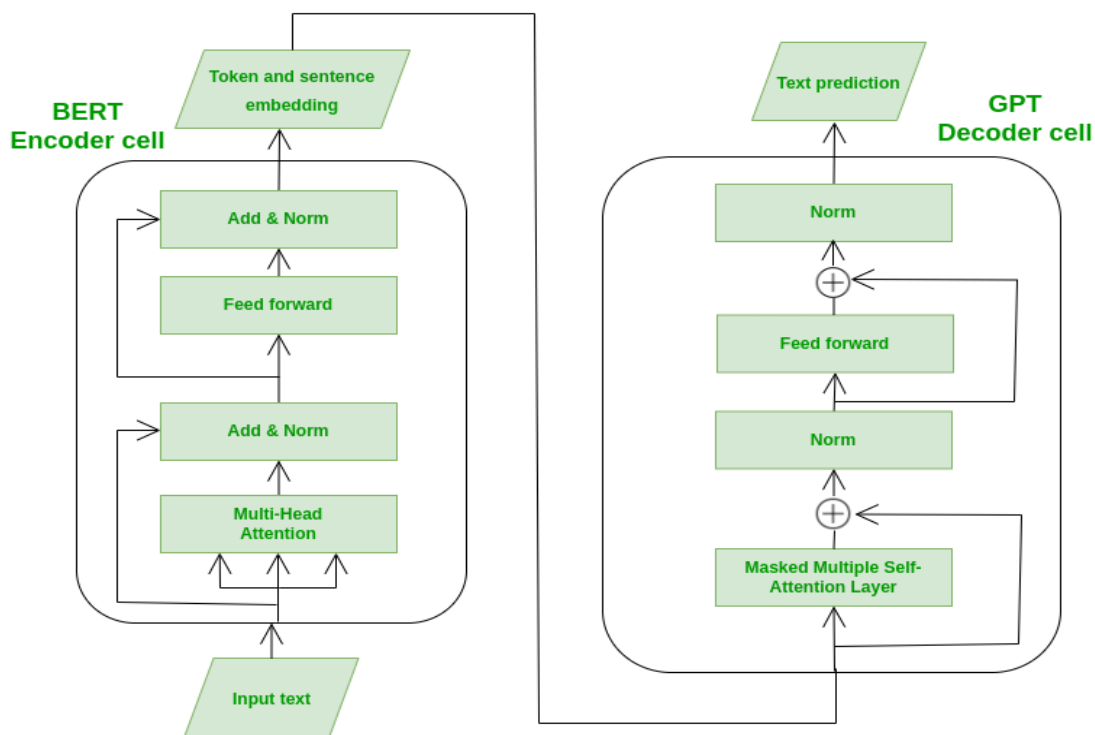


Fig 1. BART Architecture Diagram

3.1.2 Pegasus (Pre-training with Extracted Gap-sentences for Abstractive Summarization):

Pegasus [3] employs a strategy of masking whole sentences in the input text and trains the model to reconstruct these missing sentences. This captures the most salient information, ideal for generating concise summaries. Its deterministic sentence selection avoids the randomness inherent in other methods, leading to more consistent results.

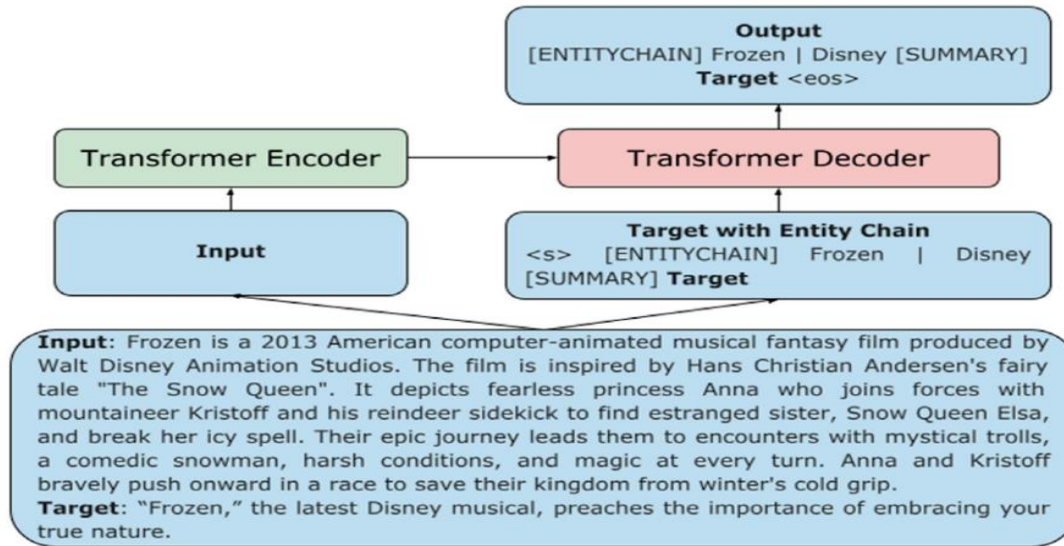


Fig 2. Pegasus Architecture Diagram

3.1.3 T5 (Text-to-Text Transfer Transformer):

T5 [34] is a powerful text-to-text transfer model. It treats all NLP tasks as text-to-text problems, enabling a unified approach to various downstream tasks, including summarization. Its extensive pre-training on a massive dataset contributes to its strong performance across a wide range of tasks. T5 is highly versatile and adaptable. It demonstrates strong performance on summarization tasks due to its massive pre-training.

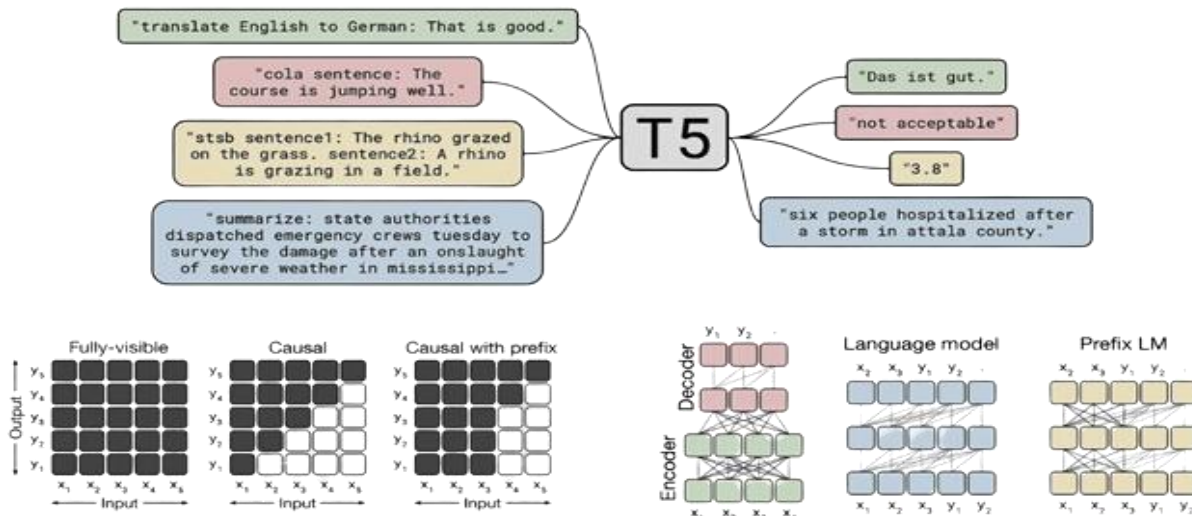


Fig 3. T5 Architecture Diagram

3.2 Sentence Scoring using Weighted Positional and Entity-based Mechanism

After obtaining the individual model-generated summaries, we implemented a novel multi-faceted scoring mechanism to evaluate and prioritize sentences for inclusion in the final summary. This approach works by incorporating domain-specific knowledge and structural information through the following formula:

$$SentenceScore(s) = W_{model} \times EntityImportanceFactor(s) \times PositionFactor(s)$$

Where s represents a candidate sentence, and the three components are defined as follows:

3.2.1 Model Weight (W_{model})

Each model's contribution is weighted according to its performance on a validation set using the ROUGE metric suite, which evaluates summary quality by measuring n-gram overlap with reference summaries. The weight w_{model} represents the normalized average of ROUGE-1, ROUGE-2, and ROUGE-L scores for each model:

$$W_{model} = \frac{avg(ROUGE - 1, ROUGE - 2, ROUGE - L)_{model}}{\sum_{all\ model} avg(ROUGE - 1, ROUGE - 2, ROUGE - L)_{model}}$$

This weighting scheme allows our framework to prioritize sentences from models that demonstrate higher reliability in the biomedical domain.

3.2.2 Entity Importance Factor

Biomedical texts contain domain-specific entities and terminology that carry significant information density. We quantify this aspect using:

$$\text{Entity Importance Factor}(s) = 0.5 \times 0.5 \times \min\left(1.0, \frac{N_e + N_t}{K}\right)$$

Where:

- N_e is the count of recognized biomedical entities in sentence s (identified using ScispaCy)
- N_t is the count of domain-specific terminology in s
- K is a normalization constant (empirically set to 10 based on validation experiments)

This factor ensures that sentences with higher concentrations of relevant biomedical concepts receive proportionally higher scores, addressing the domain-specific nature of PubMed abstracts.

3.2.3 Position Factor

Scientific abstracts typically follow a structured format where sentence position correlates with information importance. Earlier sentences often contain problem statements and principal findings, while later sentences provide supporting details. We capture this structural information using:

$$\text{Positional Factor}(s) = \alpha + (1 - \alpha) \times \left(1 - \frac{\text{pos}(s) - 1}{|S|}\right)$$

Where:

- $\text{pos}(s)$ is the position index of sentence s in its original model-generated summary
- $|S|$ is the total number of sentences in that summary
- α is a damping factor (set to 0.3) that ensures even late-positioned sentences maintain some baseline importance

This position-aware weighting mitigates the risk of overlooking important concluding statements while still recognizing the front-loaded information structure common in scientific abstracts.

The final output is a list of ranked sentences, from which the top-scoring sentences are selected to form the integrated summary, subject to length constraints. This approach effectively combines the strengths of multiple state-of-the-art abstractive summarizers while incorporating domain knowledge and structural cues specific to biomedical literature.

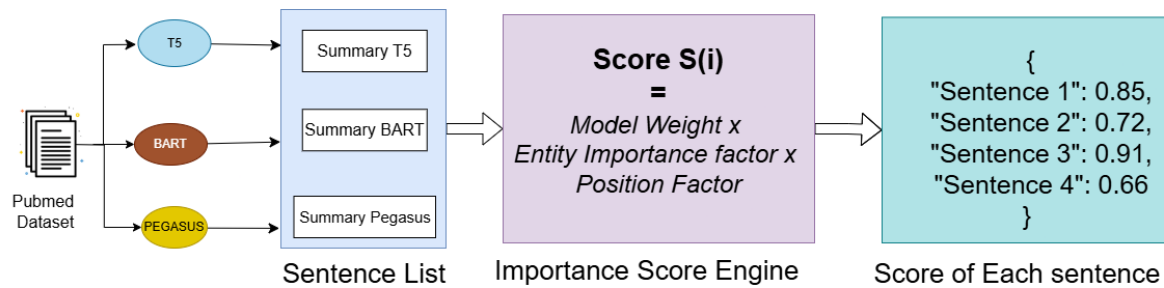


Fig 4. Sentence Scoring Mechanism

3.3 Redundancy Removal and Hybrid Summary Generation

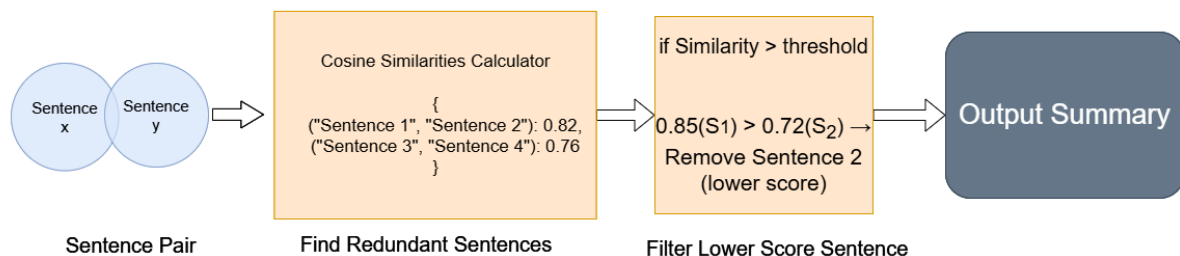


Fig 5. Redundancy Removal Mechanism using Cosine Similarity

As abstractive models may generate semantically equivalent information with different wording, this stage eliminates redundant information from the combined outputs of the three PLMs. We use a SentenceTransformer model [37] to generate embeddings for each extracted sentence. Cosine similarity can be computed between all pairs of sentence embeddings. In a pair of sentences with a cosine similarity above a predefined threshold of 0.8, the sentence with the lower score is discarded. The remaining unique sentences, ordered by their importance (based on the sentence scores from the weighted factor), are concatenated to form the final hybrid summary.

4. Dataset

The dataset used for this research is the **PubMed Summarization Dataset** (*ccdv/pubmed-summarization*), which contains biomedical articles from PubMed along with their corresponding abstract summaries. The dataset is available in the Hugging Face datasets library, specifically the document split, which provides a set of documents to be summarized.

The dataset is structured with the following key attributes:

- **Article:** The full text of the biomedical article.
- **Abstract:** A summary of the article, which serves as the ground truth for evaluation.

This dataset is particularly used for evaluating summarization models in the biomedical domain, as it includes technical and domain-specific language commonly found in scientific literature.

5. Results and Discussion

The precision, recall, and F1 scores across different ROUGE metrics are illustrated in Figure 6, 7, and 8. These visualizations allow for a more nuanced comparison of how each model behaves across different aspects of summarization quality. The respective analysis for each metric has been explained below each plot.

5.1. Performance Comparison of Proposed Approach with Existing Models on ROUGE-1

T5-large performs the worst, with a precision of 0.255, recall of 0.081, and an F1 score of 0.123, suggesting difficulties in striking a balance between precision and recall. With a precision of 0.366, recall of 0.306, and an F1 score of 0.333, BART-large exhibits a more balanced methodology and is therefore more dependable in gathering pertinent data. With the maximum recall (0.370) and a robust F1 score (0.362), Pegasus performs exceptionally well, demonstrating how well it covers important ideas from the input text. With a precision of 0.461, recall of 0.389, and an outstanding F1 score of 0.418, the **proposed method**, however, surpasses all other models, demonstrating its superior capacity to provide thorough and precise biomedical text summaries.

Table 2. Comparison of Proposed Approach with Existing Models on ROUGE-1

Model	ROUGE-1 (Precision)	ROUGE-1 (Recall)	ROUGE-1 (F1)
T5-large	0.254545	0.080925	0.122807
BART-large	0.365517	0.306358	0.333333
Pegasus	0.353591	0.369942	0.361582
Proposed Method	0.460881	0.388507	0.418353

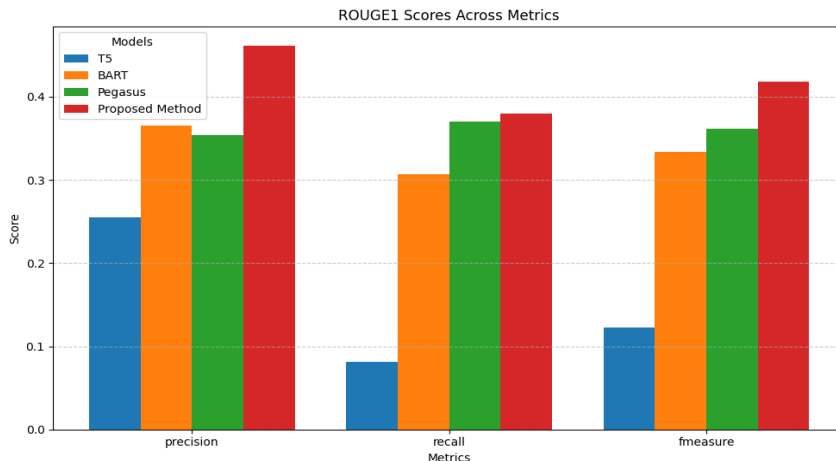


Fig 6. Bar Graph: Comparison of Proposed Approach with Existing Models on ROUGE-1

5.2. Performance Comparison of Proposed Approach with Existing Models on ROUGE-2

Information at the bi-gram level is captured by the ROUGE-2 evaluation. With a precision of 0.068, recall of 0.017, and an F1 score of 0.028, **T5-large** performs the worst, suggesting a poor capacity to precisely retain bi-gram associations. With a precision of 0.146, recall of 0.115, and F1 score of 0.128, **BART-large** exhibits a substantial improvement, indicating a more balanced but still inadequate bi-gram retention. **Pegasus** exhibits stronger recall (0.160) and F1 score (0.147), suggesting its effectiveness in capturing essential bi-gram patterns. The **Proposed Method** surpasses all baseline models with a precision of 0.154, recall of 0.187, and an F1 score of 0.168, highlighting its superior ability to generate fluent and contextually accurate summaries by effectively capturing key bi-gram information in biomedical texts.

Table 3. Comparison of Proposed Approach with Existing Models on ROUGE-2

Model	ROUGE-2 (Precision)	ROUGE-2 (Recall)	ROUGE-2 (F1)
T5-large	0.068493	0.017361	0.027701
BART-large	0.146018	0.114583	0.128405
Pegasus	0.136499	0.159722	0.1472
Proposed Method	0.153510	0.187081	0.168192

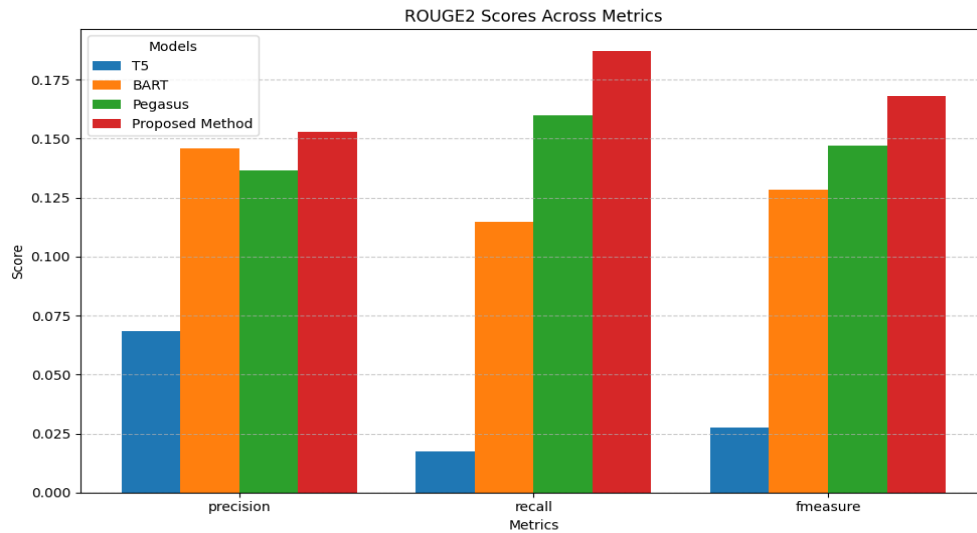


Fig 7. Bar Graph: Comparison of Proposed Approach with Existing Models on ROUGE-2

5.3. Performance Comparison of Proposed Approach with Existing Models on ROUGE-L

The ROUGE-L scores provide a comprehensive view of how well the models capture the overall structure and longest matching subsequence of the biomedical text. **T5-large** performs the weakest, with a precision of 0.218, recall of 0.069, and an F1 score of 0.105, indicating difficulties in preserving the structural integrity of the input text. **BART-large** improves significantly with a precision of 0.352, recall of 0.295, and an F1 score of 0.321, showing its ability to generate more coherent summaries. **Pegasus** achieves higher recall (0.312) and a comparable F1 score (0.305), indicating its strength in maintaining structural consistency. The **Proposed method** demonstrates the best performance, with a precision of 0.405, recall of 0.337, and an F1 score of 0.365, highlighting its superior capability in generating well-structured, contextually rich summaries that effectively retain the key elements of biomedical texts.

Table 4. Comparison of Proposed Approach with Existing Models on ROUGE-L

Model	ROUGE-L (Precision)	ROUGE-L (Recall)	ROUGE-L (F1)
T5-large	0.218182	0.069364	0.105263
BART-large	0.351724	0.294798	0.320755

Pegasus	0.298343	0.312139	0.305085
Proposed Method	0.405472	0.336922	0.365152

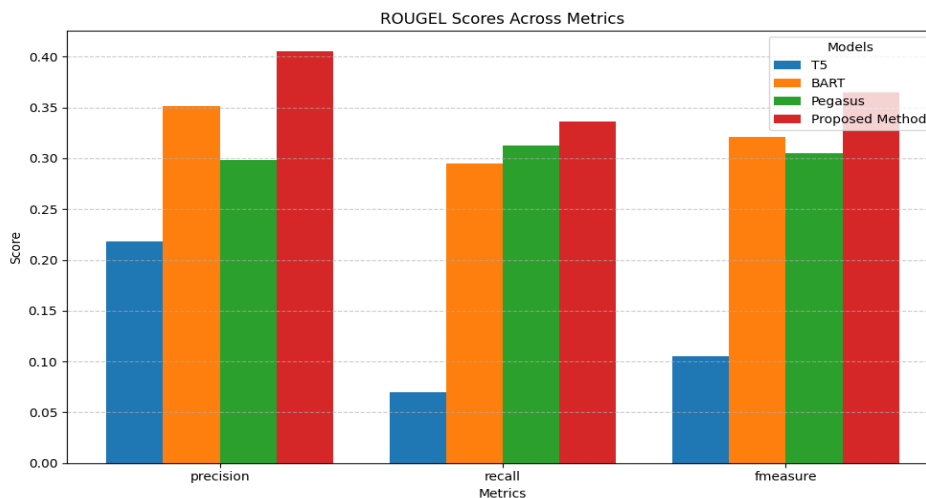


Fig 8. Bar Graph: Comparison of Proposed Approach with Existing Models on ROUGE-L

Discussion & Analysis

This study successfully demonstrated the effectiveness of proposed method for biomedical-text summarization by leveraging individual PLMs (BART, Pegasus, T5,). The individual models provide diverse summarization strengths, as shown by their ROUGE scores:

- **T5** is less effective in all ROUGE metrics, indicating limitations in bi-gram and structural retention, which could be attributed to its general-purpose pretraining.
- **BART** balances precision and recall well, particularly in ROUGE-1 and ROUGE-L, demonstrating fluency and structural coherence in summarizing complex biomedical texts.
- **Pegasus** excels in recall, especially in ROUGE-2, due to its pre-training strategy of masking entire sentences, making it adept at capturing salient information.

The use of cosine similarity-based redundancy removal and weighted fusion to combine the best features of both models improved the final results. Superior ROUGE scores were obtained using this method for all measures, with ROUGE-1 and ROUGE-L showing very good precision, suggesting that the hybrid summary is both thorough and succinct. Additionally, the ROUGE-2 F1 score outperforms individual models, demonstrating its improved coherence and bi-gram coverage. According to the experimental results, the limitations of individual PLMs can be

effectively mitigated by integrating numerous models and then reducing redundancy. When it comes to summarizing intricate and sophisticated biomedical materials, this hybrid method enhances both fluency and knowledge retention. This illustrates how it could be used in practical biomedical applications where precision and lucidity are crucial.

6. Conclusion:

This research presented a novel hybrid biomedical text summarization system that combines the outputs of BART, Pegasus, and T5 PLMs, incorporating a novel sentence scoring system followed by a redundancy removal step based on cosine similarity. Our evaluation on a subset of the PubMed dataset demonstrates a substantial improvement in ROUGE scores compared to using a single PLM. The hybrid approach proves a valuable strategy for leveraging the strengths of multiple PLMs, producing more comprehensive and concise biomedical summaries. This approach mitigates the shortcomings of relying on a single model and addresses the challenges inherent in the complex structure and specialized terminology of biomedical text. Future studies might look into applying this hybrid approach to other biomedical datasets, improving the redundancy removal procedure with more complex methods like dependency parsing or semantic role labeling, and optimizing the hyperparameters of each PLM separately using methods like Bayesian optimization.

References

- [1]. Lalitha, E., et al. (2023). Text Summarization of Medical Documents using Abstractive Techniques. 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC).
- [2]. Lewis, M., et al. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.
- [3]. Zhang, J., et al. (2020). PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. Proceedings of the 37th International Conference on Machine Learning.
- [4]. Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research.
- [5]. Lee, J. H., et al. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics.

- [6]. Alsentzer, E., et al. (2019). Publicly available clinical BERT embeddings. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.
- [7]. Beltagy, I., et al. (2019). SciBERT: A pretrained language model for scientific text. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing.
- [8]. Mike Lewis, Y., et al. (2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension.
- [9]. Jingqing Zhang, Y., et al (2020). Pegasus: Pre-training with extracted gap-sentences for abstractive summarization
- [10]. Colin Raffel, N., et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer
- [11]. Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information processing & management.
- [12]. Yoshua Bengio, Réjean Ducharme, and Pascal Vincent. 2000. A neural probabilistic language model. Advances in neural information processing systems 13 (2000).
- [13]. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation.
- [14]. Vaswani, A., et al. (2017). Attention is all you need. Advances in neural information processing systems.
- [15]. Mikolov, T., et al. (2013). Efficient estimation of word representations in vector space. Proceedings of the international conference on learning representations.
- [16]. Pennington, J., et al. (2014). Glove: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing.
- [17]. Zhiyong Lu. 2011. PubMed and beyond: a survey of web tools for searching biomedical literature.
- [18]. Jacob Devlin., et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
- [19]. Diwakar Mahajan, C., et al. (2021). Toward Improving Factual Correctness of Radiology Report Abstractive Summarization
- [20]. Yu Gu, R., et al. (2021). Domain-specific language model pre-training for biomedical natural language processing.
- [21]. Jinhyuk Lee, W., et al (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining.
- [22]. Xiaoyan Cai, S., et al. (2022). COVIDSum: A linguistically enriched SciBERT-based summarization model for COVID-19 scientific papers. Journal of Biomedical Informatics 127 (2022), 103999
- [23]. Suri, K., Mishra, P., et al. (2023). Suryakiran at mediqa-sum 2023: Leveraging lora for clinical dialogue summarization.
- [24]. Nunna, J., et al. (2022). Extractive and abstractive text summarization model fine-tuned based on bertsum and bio-bert on covid-19 open research articles.

- [25]. Qianqian Xie, Jennifer Amy Bishop, Prayag Tiwari, and Sophia Ananiadou. 2022. Pre-trained language models with domain knowledge for biomedical extractive summarization. *Knowledge-Based Systems* (2022), 109460
- [26]. Benyou Wang, Qianqian Xie, Jiahuan Pei, Prayag Tiwari, Zhao Li, et al. 2021. Pre-trained language models in biomedical domain: A systematic survey. *arXiv preprint arXiv:2110.05006* (2021).
- [27]. Wolf, T., et al. (2019). *Huggingface transformers: State-of-the-art natural language processing*.
- [28]. Hugging Face. (n.d.). CCDV PubMed Summarization dataset. Adapted from: Cohan, A., Dernoncourt, F., Kim, D. S., et al. (2018). Retrieved from <https://huggingface.co/datasets/ccdv/pubmed-summarization>
- [29]. Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*.
- [30]. Cohan, A., et al. (2018). A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 615–621. New Orleans, Louisiana: Association for Computational Linguistics. DOI: 10.18653/v1/N18-2097.
- [31]. Lin, C. Y. (2004). Rouge: A package for automatic evaluation of summaries. *Proceedings of the ACL-04 workshop on Text Summarization Branches Out*.
- [32]. Xie, Q., Luo, Z., Wang, B., & Ananiadou, S. (2023). A survey for biomedical text summarization: From pre-trained to large language models. *arXiv preprint arXiv:2304.08763*.
- [33]. Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 22, 457-479
- [34]. Zhang, W., & Gao, F. (2011). An improvement to naive bayes for text classification. *Procedia Engineering*, 15, 2160-2164.
- [35]. Nallapati, R., Zhai, F., & Zhou, B. (2017, February). Summarunner: A recurrent neural network-based sequence model for extractive summarization of documents. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 31, No. 1).
- [36]. Nallapati, R., Zhou, B., Gulcehre, C., & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*.
- [37]. Liu, Y. (2019). Fine-tune BERT for extractive summarization. *arXiv preprint arXiv:1903.10318*.
- [38]. Du, Y., Li, Q., Wang, L., & He, Y. (2020). Biomedical-domain pre-trained language model for extractive summarization. *Knowledge-Based Systems*, 199, 105964.

- [39]. Yan, A., McAuley, J., Lu, X., Du, J., Chang, E. Y., Gentili, A., & Hsu, C. N. (2022). RadBERT: adapting transformer-based language models to radiology. *Radiology: Artificial Intelligence*, 4(4), e210258.
- [40]. Bishop, J., Xie, Q., & Ananiadou, S. (2022, May). यूवी: a hybrid unsupervised summarization method using salience. In *Proceedings of the 21st workshop on biomedical language processing* (pp. 220-240).
- [41]. Park, J. W. (2020). Continual bert: Continual learning for adaptive extractive summarization of covid-19 literature. *arXiv preprint arXiv:2007.03405*.