IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 1, February 2025, pp. 680~691
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i1.pp680-691

Towards efficient knowledge extraction: Natural language processing-based summarization of research paper introductions

Nikita Chaudhari[1], Deepali Vora[1], Payal Kadam[2], Vaishali Khairnar[3], Shruti Patil[4], Ketan Kotecha[4]
[1]Department of Computer Science Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India
[2]Department of Electronics and Telecommunication Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India
[3]Department of Information Technology, Terna College of Engineering, Navi Mumbai, India
[4]Symbiosis Centre for Applied Artificial Intelligence (SCAAI), Symbiosis Institute of Technology Pune Campus, Symbiosis International (Deemed University), Pune, India

Article Info

Article history:

Received Jan 22, 2024
Revised Jul 10, 2024
Accepted Jul 26, 2024

Keywords:

BART
BERT
Natural language processing
PEGASUS
Text summarization
Text-to-text transfer transformer

ABSTRACT

Academic and research papers serve as valuable platforms for disseminating expertise and discoveries to diverse audiences. The growing volume of academic papers, with nearly 7 million new publications annually, presents a formidable challenge for students and researchers alike. Consequently, the development of research paper summarization tools has become crucial to distilling crucial insights efficiently. This study examines the effectiveness of pre-trained models like text-to-text transfer transformer (T5), bidirectional encoder representations from transformers (BERT), bidirectional and auto-regressive transformer (BART), and pre-training with extracted gap-sentences for abstractive summarization (PEGASUS) on research papers, introducing a novel hybrid model merging extractive and abstractive techniques. Comparative analysis of summaries, recall-oriented understudy for gisting evaluation (ROUGE) and bilingual evaluation understudy (BLEU) score evaluations and author evaluation help evaluate the quality and accuracy of the generated summaries. This advancement contributes to enhancing the accessibility and efficiency of assimilating complex academic content, emphasizing the importance of advanced summarization tools in promoting the accessibility of academic knowledge.

Corresponding Author:

Deepali Vora
Department of Computer Science Engineering, Symbiosis Institute of Technology
Symbiosis International (Deemed University)
Near Lupin Research Park, Gram: Lavale, Tal: Mulshi, Pune, 412 115, India
Email: deepali.vora@sitpune.edu.in

1. INTRODUCTION

The first set of scholarly literature dates to the 17th century [1], beginning with the Journal des sçavans [2] and the Philosophical Transactions of the Royal Society [3] in 1665. By 2019, Fire and Guestrin [4] analyzed over 120 million papers, estimating yearly publications at over 7 million, including 1.8 million scientific papers with more than five references. Researchers face significant challenges in keeping up with the increasing number and complexity of new scholarly papers, making it difficult to survey existing literature or move to new research areas efficiently. Automatic summarization, first proposed by Luhn in the 1950s with the heuristic method, identifies key sentences by their high-frequency words [5]–[7]. Advancements in the

1980s and 1990s introduced natural language processing techniques to improve the understanding and interpretation of sentence significance in summarization [8].

Long documents, unlike shorter ones, are typically structured into sections to aid comprehension [9], and as automatic text summarization gains popularity, there's a rising need to summarize extensive research papers due to the time-intensive nature of processing information [10]. However, challenges such as topic identification and misinterpretation persist in long document summarization [11]. Nan *et al*. [12] demonstrated that evaluating candidate summaries from the convolutional neural networks-daily mail (CNN-DM) dataset on a single NVIDIA V100 GPU takes over four days. This study evaluates the effectiveness of pre-trained transformers—T5 [13], pre-training with extracted gap-sentences for abstractive summarization (PEGASUS) [14], bidirectional encoder representations from transformers (BERT) [15], and bidirectional and auto-regressive transformer (BART) [16]—and explores a hybrid approach combining extractive and abstractive summarization to enhance accuracy in summarizing complex documents like academic papers. The approach blends techniques from BERT and PEGASUS to create a customized hybrid model tailored for research paper introductions, aiming to strike a balance between extractive and abstractive methods.

Text summarization techniques, extensively studied in natural language processing and computational linguistics, cover both extractive and abstractive methods [17]. Fattah and Ren [18] discusses challenges and advancements in automatic text summarization. Moreno [5] emphasizes the importance of summarization across diverse domains. Denomme *et al.* [19] categorize summary types and provide illustrative examples.

In specialized domains, researchers have adapted summarization techniques to specific contexts. Ghosh *et al.* [20] focus on Indian legal text summarization using text normalization approaches. Jaafar and Bouzoubaa [21] propose a hybrid approach for abstractive summarization. Gupta *et al*. [22] explore transformer-based models for automated news summarization.

Hybrid approaches combining extractive and abstractive methods have gained popularity. Fabbrizio *et al*. [23] focused on opinion extraction in multi-document summarization, while Fattah [24] proposed a diverse machine learning model for enhanced summarization. Celikyilmaz and Tur [25] introduced a hierarchical clustering-based hybrid model for organizing textual content, and Mandal *et al*. [26] combined strategies for generating concise summaries.

Despite these advancements, comprehensive evaluations are limited. Surveys in [27], [28] offer insights into summarization techniques, stressing the need for rigorous evaluation. Christensen *et al.* [29] emphasized style-specific summarization for broadcast news, while Allahyari *et al.* [30] provided an overview of text summarization methodologies. Our study introduces a novel hybrid method integrating extractive and abstractive techniques, aiming to advance text summarization through rigorous evaluation and tailored approaches to address information overload.

2. METHOD

Over the years, multiple methods have been implemented for automatic text summarization by researchers which are presented in Figure 1. Unfortunately, no method is capable of accurately summarizing documents of different domains [31]–[36]. Based on the output and purpose, the methods and approaches for summarization can vary.

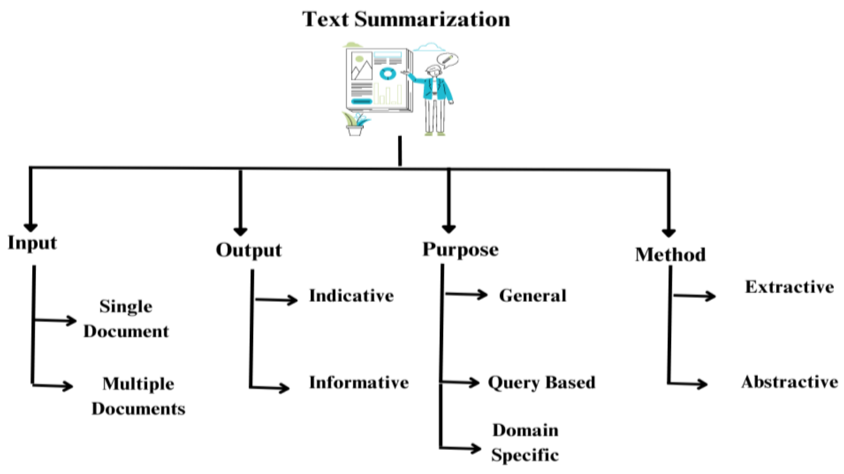

Figure 1. Text summary methodology [19]

To generate a summary, the extractive-based approach selects informative sentences directly from the input document based on predefined criteria, scoring each sentence according to its features to determine its relevance [17]. In contrast, abstractive summarization rewrites the original document using internal semantic representations and natural language processing (NLP) techniques to produce coherent and human-readable summaries [37]. This method is more complex due to the challenges of NLP and the need for effective compression techniques [38]. Figure 2 illustrates the basic flow of extractive and abstractive text summarization methods.

Figure 2. Basic extractive and abstractive flowchart [39]

Generic text summarization treats all inputs uniformly, aiming to provide a comprehensive understanding while minimizing redundancy [40]. However, without domain-specific knowledge or queries, such models may overlook critical topics, especially evident when summarizing complex domains like medical reports [41]. Query-based text summarization focuses on extracting information that directly answers specified queries within a limited word count, often lacking coherence compared to other methods [42]. Domain-based text summarization utilizes domain-specific learning to produce more precise summaries tailored to fields such as biomedical research and legal documents [43], highlighting the importance of identifying domain-specific terms. Indicative text summarization captures key topics from the original text, useful for scholarly articles where essential information is condensed [44]. Informative text summarization provides metadata about a document's scope, aiding readers in deciding whether to delve deeper into the content [45], often found on book covers or report headlines. Academic papers, structured with sections like title, abstract, introduction, methods, results, discussion, and references, can benefit from segment-specific summarization to effectively cover essential points [46].

Scientific papers poses a challenge due to their complex language. The authors explored a hybrid method using both extractive and abstractive techniques to summarize papers and evaluated transformer models' performance on paper introductions. Extractive techniques were used to select important sections or sentences from the input data, integrated into automatically generated abstractive summaries. Fabbrizio *et al.* noted that hybrid summaries are more readable and compact compared to purely extractive methods, as depicted in Figure 3 [47]–[49].

The hybrid summaries offer broad coverage of input data, are concise, avoid redundancy, and maintain high linguistic quality in readability and coherence. They are also tailored directly to readers' needs, enhancing usability. The study tested and evaluated pre-trained transformers including T5, BERT, PEGASUS, and BART to develop the hybrid model.

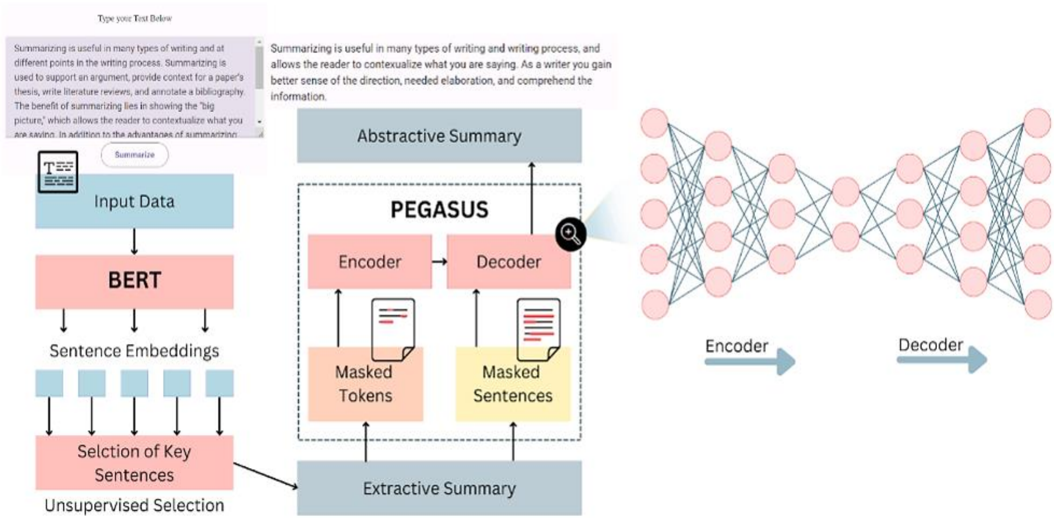

Figure 3. A hybrid model for automatic text summarization

2.1. Transformer models overview

This section provides an overview of several prominent Transformer-based models used in natural language processing tasks. T5, or text-to-text transfer transformer, serves as a versatile architecture for tasks like translation, question answering, and classification [13] as shown in Figure 4. PEGASUS which is depicted using Figure 5 is specialized for abstractive text summarization by generating summaries through a masking mechanism that creates an extractive-like summary sequence [14]. BERT employs bidirectional context to understand ambiguous language in various contexts, initially trained on Wikipedia and later adapted for Q and A tasks [15] as shown in Figure 6. BART which is shown in Figure 7 functions as a denoising autoencoder designed for sequence-to-sequence tasks, utilizing a transformer-based architecture similar to BERT. but with a specific focus on text reassembly from corrupted inputs [16]. Additionally, the hybrid model, the one being used here, integrates BERT for extractive summarization and PEGASUS for abstractive summarization, combining both approaches to enhance summary quality and achieve higher recall-oriented understudy for gisting evaluation (ROUGE) scores [12]. To begin, the raw text was passed through the BERT transformer model to obtain an extractive summary. This summary was then fed into the PEGASUS model, which produced a final abstractive summary.

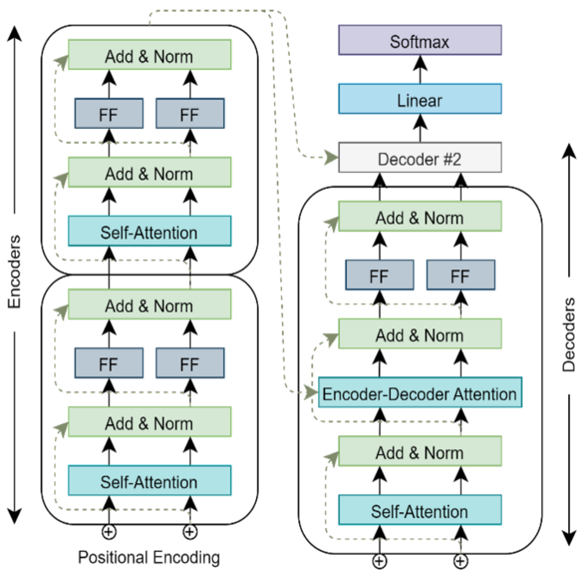

Figure 4. T5[13]

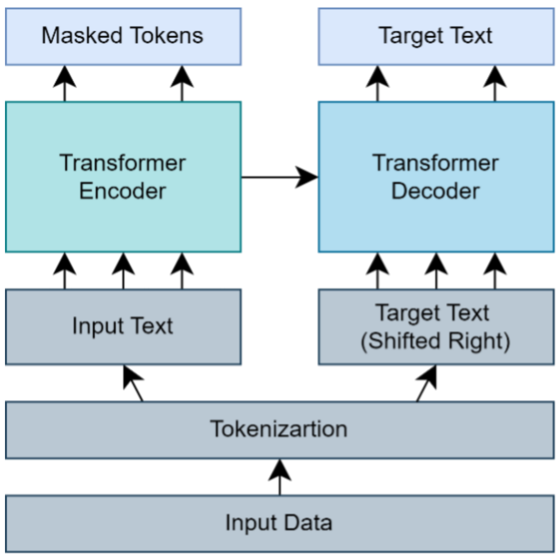

Figure 5. PEGASUS [14]

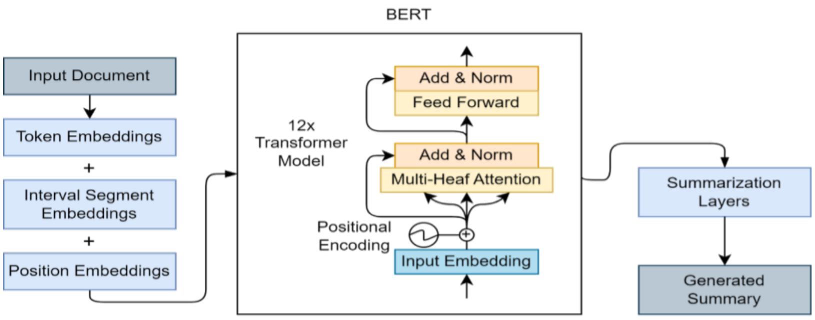

Figure 6. BERT [15]

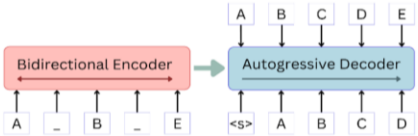

Figure 7. BART [16]

2.2. Data collection and preprocessing

During data collection, introductions from scientific papers authored by Singh *et al.* [47] formed the dataset. To ensure uniformity, rigorous preprocessing techniques were applied. This included removing white spaces, eliminating stop words, and clearing special characters to enhance text quality. All text was converted to lowercase to standardize case, punctuation marks were reduced, and numeric characters were omitted. Optionally, stemming or lemmatization could further homogenize the text. Tables 1 and 2 showcase a dataset sample featuring introductions from papers like "Stock prediction using machine learning: a review paper," with corresponding references and generated summaries.

Table 1. Information on the sample

Authors	Reference summary
Singh *et al.* [47]	In machine learning, data is obtained by knowledge extraction. Machine learning techniques learn from past predictions and give an outcome. Machine learning is used to describe the process of extracting knowledge from a code. Data is provided to the generic algorithms and logic is developed on the basis of that data. Machine learning can be defined as the data which is obtained by knowledge extraction. Instead of writing a code for every specific problem, data is provided to the generic algorithms, and logic is developed on the basis of that data

Table 2. Generated summaries for the sample

Algorithm	Generated summary
T5	Machine learning can be defined as the data which id obtained by knowledge extraction. The technique for most accurate prediction is by learning from past instances and to make a program to do this is best possible with machine learning techniques
PEGASUS	Machine learning is a branch of computer science that deals with the extraction of data from the environment and its subsequent use for the purpose of decision making. It is a rapidly growing field that is finding its way into various fields such as the data that is obtained by knowledge extraction, instead of explicitly programmed instead they are trained to make decisions that are driven by data.
BERT	Machine learning can be defined as the data which is obtained by knowledge extraction, Instead of writing code for every specific problem, data is provided to the generic algorithms and logic is developed on the basis of that data.
BART	Machines dont have to be programmed explicitly instead they are trained to make decisions that are driven by data. Data is provided to the generic algorithms and logic is developed on the basis of that data, When a machine improves its performance based on its past experiences it can be said that the machine has truly learnt.
Hybrid (BERT+PEGASUS)	The term learning is used to describe the process of extracting knowledge from a code. Learning can be defined as that data is obtained by knowledge extraction from a specific problem, Geb=neric algorithms are provided for the knowledge extraction from a code, which is provided on the basis of logic that is developed for writing a specific problem.

2.3. Hyperparameter setting

To enhance the summarization quality, hyperparameters for each method were thoughtfully established based on a combination of default values and domain expertise as mentioned in Table 3. These hyperparameter settings are specifically tailored for the hybrid model. Ensuring that it effectively combines extractive and abstractive summarization techniques to produce high-quality summaries while considering both source content and fine-tuned abstractions.

Table 3. Hyperparameter settings

Model	Learning rate	Batch size	Number of epochs	Weight decay	Optimizer type
BERT	2e-5	32	3	0.01	AdamW
BART	3e-5	16	4	0.001	Adam
PEGASUS	1e-5	8	3	0.01	AdamW
T5 (TF)	0.001	64	5	0.0001	Adam
Hybrid model	1e-5	8	4	0.001	Adam

3. RESULTS AND DISCUSSION

The evaluation of the model summaries followed the standard pattern for assessing automatically generated summaries, employing ROUGE scores, bilingual evaluation understudy (BLEU) scores, and expert evaluation. The classification of these assessments is depicted in Figure 8. Reference summaries provided by the authors of the sample papers were utilized for comparison and computation of the ROUGE and BLEU scores for the model-generated summaries.

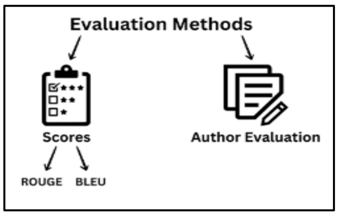

Figure 8. Evaluation methods

The ROUGE-1, ROUGE-2, and ROUGE-L scores were computed to assess summarization performance. As depicted in Table 3, the ROUGE precision, recall, and F-scores for T5, PEGASUS, BART, BERT, and the hybrid (BERT+PEGASUS) models are presented concerning the summary of samples. Table 4 provides an overview of the ROUGE scores (1, 2, and L) for all four samples. The hybrid, BERT, and T5 models demonstrated the highest scores, while BART exhibited the lowest performance. The hybrid model excelled in the first sample, while BERT performed well in the first two samples, and T5 achieved exceptional results for Sample 1 and Sample 3. It is essential to note that the ROUGE score, while informative, does not fully measure semantic accuracy in summarization. Therefore, two additional evaluation methods have been employed. Figure 9 illustrate the average ROUGE scores for each model, offering a concise representation of their comparative performance.

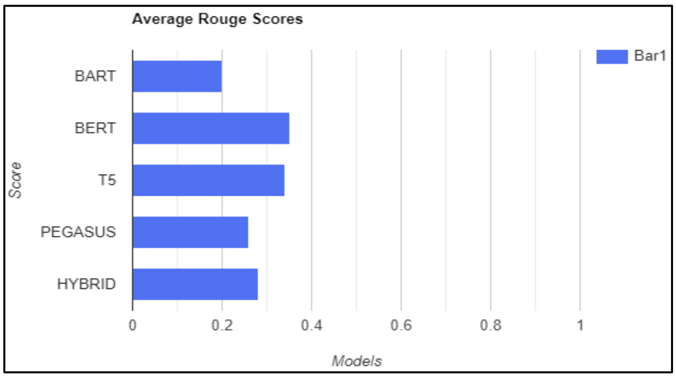

Figure 9. Average ROUGE score for each model

Table 4. ROUGE scores

	Sample no.	Recall ROUGE-1	ROUGE-2	ROUGE-L	Precision ROUGE-1	ROUGE-2	ROUGE-L	F-1 ROUGE-1	ROUGE-2	ROUGE-L
T5	1	0.571	0.300	0.524	0.353	0.150	0.324	0.436	0.200	0.400
	2	0.122	0.000	0.122	0.132	0.000	0.132	0.127	0.000	0.127
	3	0.870	0.682	0.870	0.526	0.357	0.526	0.656	0.469	0.656
	4	0.133	0.000	0.100	0.105	0.000	0.079	0.118	0.000	0.088
PEGASUS	1	0.476	0.150	0.476	0.175	0.038	0.175	0.256	0.061	0.256
	2	0.390	0.229	0.390	0.500	0.250	0.500	0.438	0.239	0.438
	3	0.174	0.000	0.130	0.148	0.000	0.111	0.160	0.000	0.120
	4	0.167	0.033	0.167	0.185	0.031	0.185	0.175	0.032	0.175
BART	1	0.286	0.000	0.238	0.136	0.000	0.114	0.185	0.000	0.154
	2	0.244	0.042	0.171	0.250	0.047	0.175	0.247	0.044	0.173
	3	0.174	0.000	0.174	0.085	0.000	0.085	0.114	0.000	0.114
	4	0.333	0.067	0.267	0.238	0.042	0.190	0.278	0.051	0.222
BERT	1	0.429	0.300	0.429	0.273	0.154	0.273	0.333	0.203	0.333
	2	0.439	0.354	0.439	0.621	0.459	0.621	0.514	0.400	0.514
	3	0.217	0.000	0.174	0.058	0.000	0.047	0.092	0.000	0.073
	4	0.433	0.200	0.367	0.317	0.122	0.268	0.366	0.152	0.310
Hybrid	1	0.381	0.250	0.381	0.211	0.094	0.211	0.271	0.137	0.271
	2	0.122	0.000	0.122	0.208	0.000	0.208	0.154	0.000	0.154
	3	0.391	0.045	0.217	0.220	0.020	0.122	0.281	0.027	0.156
	4	0.200	0.000	0.167	0.171	0.000	0.143	0.185	0.000	0.154

Table 5 illustrates the summarization performance based on BLEU scores across all models. Notably, the hybrid and BERT models demonstrate exceptional performance, surpassing BART, T5, and PEGASUS. This performance difference is attributed to BLEU scores' tendency to favor extractive summarization methods, aligning with BERT's strengths in this context. Figure 10 visually depicts the mean BLEU scores for each model, offering a concise overview of their performances. However, it's important to note that while BLEU and ROUGE scores are valuable metrics, they may not fully capture the semantic accuracy of summarization. Therefore, an author's evaluation is conducted, as depicted in Figure 11, assessing summaries based on conciseness, accuracy, and detail preservation. This evaluation is critical as it provides a comprehensive perspective on the quality of summarization, considering factors beyond quantitative scores.

Table 5. Average BLEU Scores

BART	BERT	T5	PEGASUS	Hybrid
		Sample 1		
0.28662	1	0.24225	0.12015	0.41739
		Sample 2		
0.16532	0.81044	0.22855	0	0.18402
		Sample 3		
0	0.56930	0	0.06130	0.64468
		Sample 4		
0.56292	0.79863	0.09943	0	0.72343

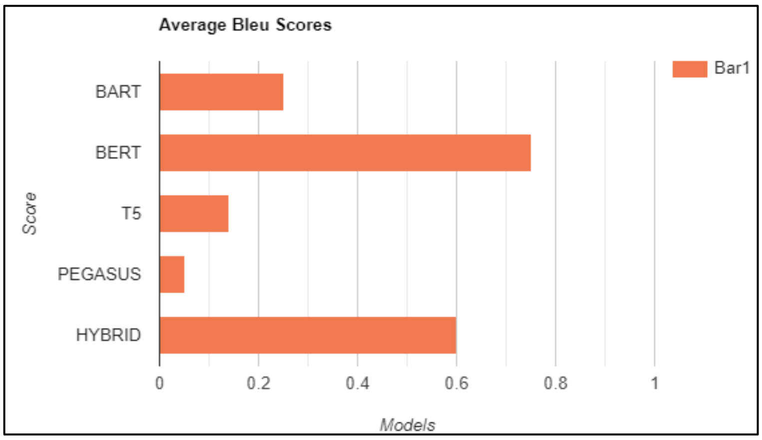

Figure 10. Average BLEU score for each model

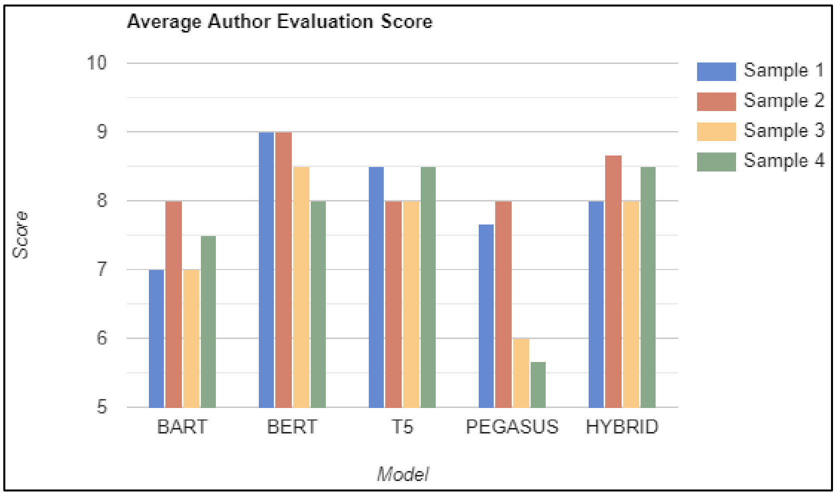

Figure 11. Average author evaluation scores

It is both computationally and theoretically challenging to achieve flawless ROUGE scores. For evaluation, the method involved calculating the ROUGE scores for all generated summaries using a reference summary approved by the authors of the papers. In the evaluation, PEGASUS, BERT, hybrid, T5, and BART were assessed across multiple samples for summarization performance. In Sample 1, PEGASUS, BERT, and hybrid achieved ROUGE 1 F1 scores of 0.256, 0.333, and 0.271, respectively, indicating satisfactory summarization quality with potential for improvement. In Sample 2, BERT achieved an F1 score of 0.514, outperforming other models, while hybrid followed with 0.185; T5 had multiple inconsistencies. Sample 3 highlighted T5 and BART's effectiveness with ROUGE 1 F1 scores of 0.656 and 0.114, respectively, compared to BERT and PEGASUS. Sample 4 showed BART's superior coherence despite lower ROUGE scores than BERT, which excelled in Bleu scores alongside hybrid. Overall, T5 and BART consistently provided competitive summaries, demonstrating their efficacy in diverse summarization tasks.

ROUGE and BLEU scores are standard metrics in natural language processing tasks, including text summarization and machine translation. Despite their usefulness, they are limited in assessing semantic and structural aspects of text, focusing mainly on surface-level similarity and lacking consideration for subjective human understanding. Author evaluation proves to be more accurate, as demonstrated in this study comparing the Hybrid and BERT models. The evaluation involved processing random samples from the dataset using these models, and assessing summaries based on criteria such as conciseness, accuracy, and detail preservation. Aggregated scores from ROUGE, BLEU, and author evaluations were weighted and presented in Table 6, with Table 7 showing the overall average scores across all models based on these evaluation metrics.

Table 6. Evaluation parameter distribution

Evaluation method	Weighted percentage (%)
ROUGE	30
BLEU	20
Author evaluation	50

Table 7. Total average scores for all models

Sample	BART	BERT	T5	PEGASUS	Hybrid
Sample 1	4.627	7.499	6.02	4.84	5.6
Sample 2	5.0607	7.662	4.834	5.31	5.16
Sample 3	3.83	5.6637	5.965	3.6	6.13
Sample 4	5.7068	6.6952	4.7998	3.36	6.25
Average	4.8	6.88	5.4	4.28	5.79

By comparing figures (Figure 8, Figure 9, Figure 11, and Figure 12), it can be concluded that the highest average scores for all samples are presented by BERT and the hybrid model; the latter providing the best abstracting summaries in the majority of cases, according to Figure 12. Although the hybrid model garners a score slightly lower than other models in both classic ROUGE and BLEU metrics, authors' overall assessment confirmed its effectiveness in generating summary of research paper abstracts that are accurate and relevant. The comparison in Table 7 indicates the superiority of BERT on the considered metrics; however, it is critical to note the shortcomings of these indicators for evaluating the quality of summaries in general. Some of the difficulties that were present were data heterogeneity, meaning that the styles, content, and structures of research papers that were received were different, and caused an issue because they could not be standardized to have one effective summarization model for them all. Furthermore, the concept of subjectivity in evaluation emerged as a result of the issues with evaluation procedures that rely on relative approaches, including BLEU and ROUGE, and formulas, as the latter are incapable of considering such qualitative criteria for summaries as coherence or informativeness deemed to be vital for evaluation and may potentially impose systematic bias.

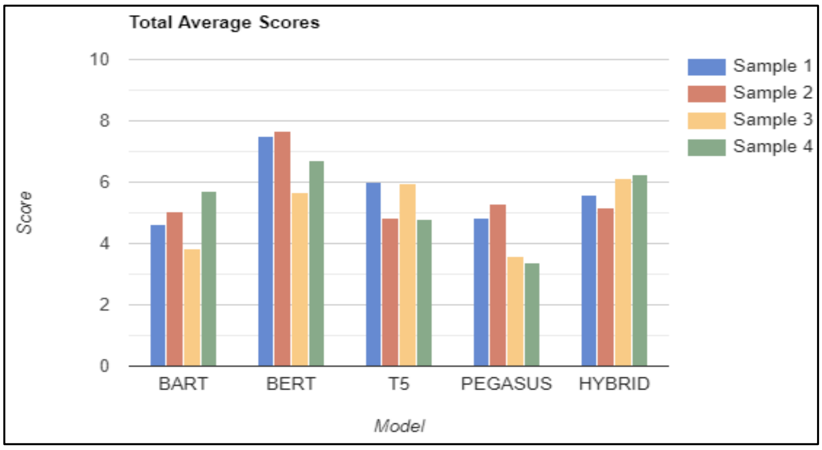

Figure 12. Total average scores for each sample

The evaluation metrics, such as ROUGE and BLEU are used to evaluate the quality of the summaries and these metrics tend to reward the models that focus more on the word occurrence, which is very much in line with extractive mechanisms, such as BERT. However, summarization quality does not refer to the basic task of literally copying contents. The post-processing combines content selection from BERT with the summarization's abstractive nature to provide short, logical, and comprehensible summaries that contain essential ideas as authors determine, as established. This balanced scheme improves the coherency and the overall structure of the summaries, where UmBERT even out BERT for developing user-orientated summaries that fit the readers' requirements. Thus, whereas in the traditional measures it is beneficial for BERT, the method that is being developed for reader-oriented summarization with the help of author assessment is more beneficial for the same purpose of summarizing research papers.

3.1. Applications of text summarization methods and future work

As for the further development of this research, it is noteworthy to expand the use of this model beyond identifying the introductions of research papers. The CAST approach [50] effectively classifies intra-section and cross-section relations in scientific articles, improving multi-article summarization and addressing redundancy and readability issues. Thus, the proposed hybrid approach can be further extended to synthesize summaries for short text documents using pretraining language models such as BERT [51] and PEGASUS [52]. This adaptation prove useful when space is a concern such as individual tweets, post headlines or news briefs. Moreover, the method's ability to generate abstractive summaries can be helpful in developing prompts for identifying backdoor attacks on language models. Finally, the strengths of both extractive and abstractive elements integrated in the approach allow for advancing the seq2seq models with the benefits of higher effectiveness and accuracy in multiple NLP applications. This work provides a foundation for future work in the creation of stronger summarization systems as well as the enhancement of the NLP discipline [53]-[55].

4. CONCLUSION

The study develops a machine-learning model specifically for summarizing research and academic papers. It applied summarization models—BART, T5, PEGASUS, and BERT—to eight samples from academic works' Introduction sections. BERT achieved the highest BLEU scores, demonstrating superior performance. A hybrid model merging BERT and PEGASUS was created to generate summaries, combining abstractive and extractive techniques. This hybrid model showed excellent ROUGE and BLEU scores and received positive feedback. Future enhancements could further optimize the model for practical applications in aiding academic research. Exploring different iterations of hybrid models integrating diverse techniques represents a promising avenue for future research, including the summarization of scientific study textbooks and other complex documents.

REFERENCES

[1] Wikipedia Contributors, "Academic publishing," *Wikipedia*, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Academic_publishing

[2] T. Volpe and J. Schopfel, "Dissemination of knowledge and copyright: An historical case study," *Journal of Information, Communication and Ethics in Society*, vol. 11, no. 3, pp. 144–155, 2013, doi: 10.1108/JICES-06-2013-0018.

[3] L. Partridge, "Celebrating 350 years of philosophical transactions: life sciences papers," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1666, 2015, doi: 10.1098/rstb.2014.0380.

[4] M. Fire and C. Guestrin, "Over-optimization of academic publishing metrics: observing goodhart's law in action," *GigaScience*, vol. 8, no. 6, 2019, doi: 10.1093/gigascience/giz053.

[5] J. T. -Moreno, "Why summarize texts?," *Automatic Text Summarization*, pp. 1–21, 2014, doi: 10.1002/9781119004752.ch1.

[6] A. Nenkova and K. McKeown, "Automatic summarization," *Foundations and Trends in Information Retrieval*, vol. 5, no. 2–3, pp. 103–233, 2011, doi: 10.1561/1500000015.

[7] A. Vashisht, "Luhn's heuristic method for text summarization," *OpenGenus IQ: Computing Expertise & Legacy*, 2019. [Online]. Available: https://iq.opengenus.org/luhns-heuristic-method-for-text-summarization/

[8] L. Gonçalves, "Automatic text summarization with machine lLearning — an overview," *Luisfredgs*, 2020, [Online]. Available: https://medium.com/luisfredgs/automatic-text-summarization-with-machine-learning-an-overview-68ded5717a25

[9] A. Cohan *et al.*, "A discourse-aware attention model for abstractive summarization of long documents," *2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 2, pp. 615–621, 2018, doi: 10.18653/v1/n18-2097.

[10] H. Y. Koh, J. Ju, M. Liu, S. Pan, and S. Pan, "An empirical survey on long document summarization: datasets, models, and metrics," *ACM Computing Surveys*, vol. 55, no. 8, 2022, doi: 10.1145/3545176.

[11] U. Hahn and I. Mani, "Challenges of automatic summarization," *Computer*, vol. 33, no. 11, pp. 29–36, 2000, doi: 10.1109/2.881692.

[12] F. Nan *et al.*, "Improving factual consistency of abstractive summarization via question answering," *ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, pp. 6881–6894, 2021, doi: 10.18653/v1/2021.acl-long.536.

[13] C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, 2020.

[14] J. Zhang, Y. Zhao, M. Saleh, and P. J. Liu, "PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization," *37th International Conference on Machine Learning, ICML 2020*, pp. 11265–11276, 2020.

[15] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 4171–4186, 2019, doi: 10.18653/v1/N19-1423.

[16] M. Lewis *et al.*, "BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 7871–7880, 2020, doi: 10.18653/v1/2020.acl-main.703.

[17] N. Andhale and L. A. Bewoor, "An overview of text summarization techniques," *2nd International Conference on Computing, Communication, Control and Automation, ICCUBEA 2016*, 2017, doi: 10.1109/ICCUBEA.2016.7860024.

[18] M. A. Fattah and F. Ren, "Automatic text summarization," *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 2, no. 1, pp. 90–93, 2008.

[19] D. Denomme, N. Taccone, and M. Firestone, "Summary definition, types & examples," *Study Learn Lesson*, 2023. [Online]. Available: https://study.com/learn/lesson/what-is-a-summary.html

[20] S. Ghosh, M. Dutta, and T. Das, "Indian legal text summarization: a text normalization-based approach," *INDICON 2022 - 2022 IEEE 19th India Council International Conference*, 2022, doi: 10.1109/INDICON56171.2022.10039891.

[21] Y. Jaafar and K. Bouzoubaa, "Towards a new hybrid approach for abstractive summarization," *Procedia Computer Science*, vol. 142, pp. 286–293, 2018, doi: 10.1016/j.procs.2018.10.496.

[22] A. Gupta, D. Chugh, Anjum, and R. Katarya, "Automated news summarization using transformers," *Sustainable Advanced Computing*, vol. 840, pp. 249–259, 2022, doi: 10.1007/978-981-16-9012-9_21.

[23] G. D. Fabbrizio, A. J. Stent, and R. Gaizauskas, "A hybrid approach to multi-document summarization of opinions in reviews," *INLG 2014 - Proceedings of the 8th International Natural Language Generation Conference - Proceedings of the INLG and SIGDIAL 2014 Joint Session*, pp. 54–63, 2014, doi: 10.3115/v1/w14-4408.

[24] M. A. Fattah, "A hybrid machine learning model for multi-document summarization," *Applied Intelligence*, vol. 40, no. 4, pp. 592–600, 2014, doi: 10.1007/s10489-013-0490-0.

[25] A. Celikyilmaz and D. H. -Tur, "A hybrid hierarchical model for multi-document summarization," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, vol. 2010, pp. 815–824, 2010, doi: 10.26713/jims.v9i3.760.

[26] S. Mandal, G. K. Singh, and A. Pal, "A hybrid text summarization approach," *Journal of Informatics and Mathematical Sciences*, vol. 9, no. 3, pp. 547–555, 2017.

[27] M. Kirmani, N. M. Hakak, M. Mohd, and M. Mohd, "Hybrid text summarization: a survey," *Advances in Intelligent Systems and Computing*, vol. 742, pp. 63–73, 2019, doi: 10.1007/978-981-13-0589-4_7.

[28] I. Awasthi, K. Gupta, P. S. Bhogal, S. S. Anand, and P. K. Soni, "Natural language processing (NLP) based text summarization-a survey," *The 6th International Conference on Inventive Computation Technologies*, pp. 1310–1317, 2021, doi: 10.1109/ICICT50816.2021.9358703.

[29] H. Christensen, B. Kolluru, Y. Gotoh, and S. Renais, "From text summarisation to style-specific summarisation for broadcast news," *Advances in Information Retrieval (ECIR 2004)*, vol. 2997, pp. 223–237, 2004, doi: 10.1007/978-3-540-24752-4_17.

[30] M. Allahyari *et al.*, "Text summarization techniques: a brief survey," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, Jul. 2017, doi: 10.48550/arxiv.1707.02268.
[31] A. Nenkova and K. McKeown, "A survey of text summarization techniques," *Mining Text Data*, vol. 9781461432234, pp. 43–76, 2012, doi: 10.1007/978-1-4614-3223-4_3.
[32] J. Shi, L. Hellesoe, G. Wang, W. Li, and Q. Bai, "Focus-based text summarisation with hybrid embeddings," *AI 2021: Advances in Artificial Intelligence*, pp. 705–715, 2022, doi: 10.1007/978-3-030-97546-3_57.
[33] C. Y. Lin, "ROUGE: a package for automatic evaluation of summaries," in *Proceedings of the workshop on text summarization branches out (WAS 2004)*, 2004, pp. 74–81.
[34] K. Papineni, S. Roukos, T. Ward, and W. J. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, USA: Association for Computational Linguistics, 2002, pp. 311–318, doi: 10.3115/1073083.1073135.
[35] E. Hermansson and C. Boddien, "Using pre-trained language models for extractive text summarisation of academic papers," *M.Sc. Thesis*, Department of Mechanics and Maritime Sciences, Chalmers University of Technology, Gothenburg, Sweden, 2020.
[36] Y. K. Meena and D. Gopalani, "Domain independent framework for automatic text summarization," *Procedia Computer Science*, vol. 48, no. C, pp. 722–727, 2015, doi: 10.1016/j.procs.2015.04.207.
[37] V. N. Gudivada, "Natural language core tasks and applications," *Handbook of Statistics*, vol. 38, pp. 403–428, 2018, doi: 10.1016/bs.host.2018.07.010.
[38] A. Khan and N. Salim, "A review on abstractive summarization methods," *Journal of Theoretical and Applied Information Technology*, vol. 59, pp. 64–72, 2014.
[39] 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, 2021, doi: 10.1016/j.eswa.2020.113679.
[40] K. Chauhan, "Unsupervised text summarization using sentence embeddings," *Medium*, 2019. Accessed: Apr. 05, 2024. [Online]. Available: https://medium.com/jatana/unsupervised-text-summarization-using-sentence-embeddings-adb15ce83db1
[41] Y. Gong and X. Liu, "Generic text summarization using relevance measure and latent semantic analysis," *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*, pp. 19–25, 2001, doi: 10.1145/383952.383955.
[42] M. Damova and I. Koychev, "Query-based summarization: a survey," *Second International Conference S3T*, Varna, Bulgaria: Sofia University, 2010.
[43] S. Afantenos, V. Karkaletsis, and P. Stamatopoulos, "Summarization from medical documents: a survey," *Artificial Intelligence in Medicine*, vol. 33, no. 2, pp. 157–177, 2005, doi: 10.1016/j.artmed.2004.07.017.
[44] S. Gholamrezazadeh, M. A. Salehi, and B. Gholamzadeh, "A comprehensive survey on text summarization systems," *Proceedings of the 2009 2nd International Conference on Computer Science and Its Applications, CSA 2009*, 2009, doi: 10.1109/CSA.2009.5404226.
[45] A. NithyaKalyani and S. Jothilakshmi, "Speech summarization for tamil language," *Intelligent Speech Signal Processing*, pp. 113–138, 2019, doi: 10.1016/B978-0-12-818130-0.00007-6.
[46] G. R. VandenBos, "Publication manual of the American psychological association (6th ed.)," Washington, DC: American Psychological Association, pp. 41–60.
[47] N. Singh, N. Khalfay, V. Soni, and D. Vora, "Stock prediction using machine learning a review paper," *International Journal of Computer Applications*, vol. 163, no. 5, pp. 36–43, 2017, doi: 10.5120/ijca2017913453.
[48] N. Schluter, "The limits of automatic summarisation according to ROUGE," *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017*, vol. 2, pp. 41–45, 2017, doi: 10.18653/v1/e17-2007.
[49] R. Guetari and N. Kraiem, "CoMod: an abstractive approach to discourse context identification," *IEEE Access*, vol. 11, pp. 82744–82770, 2023, doi: 10.1109/ACCESS.2023.3302179.
[50] N. I. Altmami and M. E. B. Menai, "CAST: a cross-article structure theory for multi-article summarization," *IEEE Access*, vol. 8, pp. 100194–100211, 2020, doi: 10.1109/ACCESS.2020.2997881.
[51] S. A. -Salam and A. Rafea, "Performance study on extractive text summarization using BERT models," *Information*, vol. 13, no. 2, 2022, doi: 10.3390/info13020067.
[52] M. Alsuhaibani, "Fine-tuned PEGASUS: exploring the performance of the transformer-based model on a diverse text summarization dataset," *Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science*, 2023, doi: 10.11159/cist23.117.
[53] M. M. Saiyyad and N. N. Patil, "Text summarization using deep learning techniques: a review," *Engineering Proceedings*, vol. 59, no. 1, 2023, doi: 10.3390/engproc2023059194.
[54] H. A. -Bakr and S. A. Mohamed, "Automatic multi-documents text summarization by a large-scale sparse multi-objective optimization algorithm," *Complex and Intelligent Systems*, vol. 9, no. 4, pp. 4629–4644, 2023, doi: 10.1007/s40747-023-00967-y.
[55] S. Patil, L. Chavan, J. Mukane, D. Vora, and V. Chitre, "State-of-the-art approach to e-Learning with cutting edge NLP transformers: implementing text summarization, question and distractor generation, question answering," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, pp. 445–453, 2022, doi: 10.14569/IJACSA.2022.0130155.

BIOGRAPHIES OF AUTHORS

Nikita Chaudhari is a research intern at the Symbiosis Centre of Applied AI and a Master's in Science student at Carnegie Mellon University's School of Computer Science. She earned her Bachelor of Technology in Computer Science and Engineering at Symbiosis Institute of Technology, SIU, in 2023. Her research interests span natural language processing (NLP), machine learning (ML), deep learning, computer vision, and interactive AI. She has co-authored 2 papers in reputed journals and is actively working on many ongoing projects at SCAAI. She can be contacted at email: nikitarajeevchaudhari26@gmail.com.

Dr. Deepali Vora (Senior Member, IEEE) completed her Ph.D. in Computer Science and Engineering from Amity University, Mumbai. Currently working as Professor and Head at Department of Computer Science, and Engineering, Symbiosis Institute of Technology Pune, Symbiosis International University (Deemed), Pune, India. She has more than 24 years of experience in total in teaching, research, and Industry. She has published more than 75 research papers in reputed national, international conferences, and journals. She has co-authored three books and many book chapters and delivered various talks in data science and machine learning. She received grants from government bodies such as AICTE, ISTE, DST and the industry. She can be contacted at email: deepali.vora11@gmail.com.

Payal Kadam is a research scholar at Symbiosis International University, Pune and working as an Assistant Professor at Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune. she has completed Master of Technology from Bharati Vidyapeeth (Deemed to be University), Pune in Electronics (VLIS) and Bachelor of Engineering from Shivaji University in 'Electronics and Telecommunication'. Her area of interest includes image processing and deep learning. She has published more than 18 research papers in reputed national, international journals and conferences. She can be contacted at email: payalskadam94@gmail.com.

Vaishali Khairnar is a Professor and Head of Department of Information Technology at Terna Engineering College, Navi-Mumbai. She has total 21 years of teaching experience. She is Board of Studies Member in Information Technology, University of Mumbai. She has guided many Ph.D. and P.G. studies. Her areas of interest are wireless communication, connected vehicles, VANET, and storage. She has published more than 50 plus papers in Scopus journal, Springer, and IEEE. She has published 3 patents. She has written and published more than five books under Wiley publication. She has completed one consultancy project and currently working on research funded project in area of connected vehicles under Department of Science and Technology. She has received Best Research Award in 2021. She can be contacted at email: vaishalikhairnar@ternaengg.ac.in.

Dr. Shruti Patil is an industry professional currently associated with the Symbiosis Institute of Technology as a professor and with SCAAI as a research associate in Pune, Maharashtra. She completed her M.Tech. in Computer Science and Ph.D. in the domain of Data Privacy from Pune University. She has 3 years of industry experience and 10 years of academic experience. She has expertise in applying innovative technology solutions to real world problems. Her research areas include applied artificial intelligence, natural language processing, acoustic AI, adversarial machine learning, data privacy, digital twin applications, GANS, and multimodal data analysis. She is currently working in the application domains of healthcare, sentiment analysis, emotion detection, and machine simulation in which she is also guiding several UG, PG, and Ph.D. students as a domain expert. She has published 75+ research articles in reputed international conferences and Scopus and Web of Science indexed journals, as well as books with 90+ citations. She can be contacted at email: shruti.patil@sitpune.edu.in.

Dr. Ketan Kotecha received a Ph.D. and M.Tech. from IIT Bombay, India, and currently holds position as the Head of the Symbiosis Center for Applied AI (SCAAI), the Director of Symbiosis Institute of Technology, and Dean of the Faculty of Engineering, Symbiosis International (Deemed University). He has gained expertise and experience in cutting-edge research and projects in AI and deep learning over the last 25+ years. He can be contacted at email: director@sitpune.edu.in.