Exploring Automated Text Summarization in Clinical Approaches Trials: Towards Explainable AI Solutions

Zakaria Benzadri[1], Maroua Benhamlaoui[2] and Naila Marir[3,1]

[1]University of Constantine 2 - Abdelhamid Mehri,
LIRE Laboratory, Constantine, Algeria
[2]University Fréres Mentouri- Constantine 1, Constantine, Algeria
[3]Effat College of Engineering, Effat University, Jeddah, Saudi Arabia

Abstract
The exponential rise in digital content has made Automatic Text Summarization (ATS) an indispensable tool in information retrieval and knowledge management. Neural network architectures have significantly advanced the efficiency and accuracy of ATS systems, particularly in specialized domains requiring precise and reliable information synthesis. This paper provides a comprehensive review of recent advancements in neural network-based summarization models, highlighting the state-of-the-art methodologies, challenges faced in adapting these models to complex, domain-specific texts, and areas for future research. Additionally, the limitations of current techniques are critically analyzed, alongside the emerging need for more robust evaluation metrics to ensure the practical relevance and effectiveness of ATS systems. This work aims to provide valuable insights to researchers and practitioners dedicated to advancing the capabilities of ATS technologies.

Keywords
Automatic Text Summarization, Neural Networks, Clinical Text Summarization, Explainable AI, Transparent AI,

1. Introduction

The explosive growth of digital data presents significant challenges in managing, retrieving, and synthesizing relevant information efficiently. In response, ATS has emerged as a indispensable tool for distilling large volumes of text into concise, meaningful summaries. By providing users with simplified representations of complex content, ATS systems can greatly enhance information accessibility and utility. However, the complexity and domain-specific nature of certain texts, especially in technical fields, create a demand for more sophisticated summarization techniques.

Recent advancements in neural network architectures, particularly deep learning models such as recurrent neural networks, transformers and generative AI, have revolutionized the field of ATS. These models, which leverage techniques such as attention mechanisms and pre-training on large corpora, have demonstrated substantial improvements over traditional rule-based or

TACC'2024, The 4th Tunisian-Algerian Conference on applied Computing, December 17-18, 2024, Constantine, Algeria
*Corresponding author.
zakaria.benzadri@univ-constantine2.dz (Z.BENZADRI);

statistical methods. By enabling more context-aware and coherent summaries, neural networks have set new standards for both extractive and abstractive summarization tasks.

Despite these advancements, numerous challenges remain. Neural network models, while powerful, often struggle with domain-specific content where terminological precision and factual consistency are critical. This is particularly evident in fields where summarization tasks must meet high standards of accuracy and reliability. Moreover, the inherent complexity of neural networks makes them prone to issues like overfitting and generalization errors, especially when applied to specialized corpora. These limitations call for further research into more robust, adaptable architectures that can handle the intricacies of complex textual data.

As the field continues to evolve, there is also a growing need for more comprehensive evaluation metrics that go beyond surface-level text matching. While metrics like ROUGE and BLEU have been widely adopted, they are often inadequate for assessing the deeper semantic accuracy required in domain-specific summarization. Consequently, developing new metrics that better capture the relevance, accuracy, and completeness of summaries will be crucial for advancing the effectiveness of ATS systems in specialized applications.

2. Literature Review: Neural Network Models for Clinical Text Summarization

In recent years, the volume of clinical data has increased exponentially, creating a need for efficient methods to distill key information from complex medical texts [1]. Clinical text summarization, which involves condensing lengthy medical documents such as patient records, discharge summaries, and radiology reports into concise, meaningful summaries, is essential for improving healthcare delivery. This section presents an overview of the advancements in neural network models for clinical text summarization, highlighting their applications in the medical field, challenges encountered, and current limitations. The focus will be on understanding the specific characteristics of neural network-based approaches, such as sequence-to-sequence models, transformer-based architectures, and their efficacy in extracting meaningful insights from clinical texts. Additionally, the section will address the need for explainable AI techniques to enhance the transparency and trustworthiness of these models in the healthcare context.

2.1. Key Neural Network in Clinical Text Summarization

Recent advancements in neural networks have significantly impacted the field of text summarization. Sequence-to-sequence (Seq2Seq) models, transformers, BERT, and GPT are among the most prominent models used for this purpose. Seq2Seq models, particularly those enhanced with attention mechanisms, have shown substantial improvements in generating coherent summaries [2]. Transformers, with their self-attention mechanisms, have revolutionized text summarization by enabling the processing of entire sequences simultaneously, leading to more contextually aware summaries [3]. BERT, a bidirectional transformer, has been particularly effective in capturing contextual information, making it a powerful tool for both extractive and abstractive summarization tasks [4, 5]. GPT, with its autoregressive nature, excels in generating human-like text, making it suitable for abstractive summarization [6].

In the medical domain, these models have been adapted to handle the unique challenges posed by clinical texts. For instance, BERT-based models have been employed to classify and summarize clinical reports, achieving high accuracy in identifying key findings from radiographic reports [7]. The Biomed-Summarizer framework integrates deep neural networks to provide context-aware summarization of biomedical texts, ensuring that the summaries are both precise and clinically relevant [8]. Additionally, models like T-BERTSum have been developed to handle long text dependencies and latent topic mapping, which are crucial for summarizing extensive clinical documents [4].

2.2. Challenges in Neural Network based Models for Clinical Text Summarization

Clinical text summarization faces several unique challenges, including the complexity of medical jargon, the presence of regulatory language, and the sensitivity of information. Medical texts often contain specialized terminology that can be difficult for general models to understand and accurately summarize [9]. Furthermore, the need to comply with regulatory standards adds an additional layer of complexity, as the summaries must be both accurate and legally compliant.

Current models, while advanced, still face limitations such as generalization and overfitting. For example, BERT's performance can be inhibited by its pretraining and tokenization methods, which may not be well-suited for the specific nuances of clinical texts [9]. Additionally, models like GPT-3.5 and ChatGPT, despite their capabilities, can generate factually inconsistent summaries and struggle with identifying salient information in longer texts [6]. These limitations highlight the need for further refinement and adaptation of these models to better handle the intricacies of clinical text summarization.

3. Explainable AI (XAI) in Clinical Decision Support

Explainable AI (XAI) refers to the development of machine learning models that provide transparent, understandable, and interpretable explanations for their decisions and predictions. In clinical decision support, XAI plays a critical role by offering insights into the reasoning behind AI-driven recommendations, making them more accessible and trustworthy for healthcare professionals. The ability to interpret AI models' outputs is particularly important in healthcare, where decisions can directly impact patient outcomes. This section explores the application of XAI in clinical decision support systems (CDSS) and its potential to enhance decision-making processes in healthcare settings.

3.1. Applications of XAI in Clinical Decision Support

XAI techniques are increasingly being applied to clinical decision support systems to enhance the interpretability of AI models [10, 11]. These systems assist clinicians in making decisions related to diagnosis, treatment recommendations, and patient risk assessments[12, 13]. One of the key areas where XAI is applied is in risk prediction and personalized medicine. AI models are often used to predict patient outcomes, such as the likelihood of disease progression or complications. XAI methods, such as SHAP (SHapley Additive exPlanations) and LIME (Local

Interpretable Model-agnostic Explanations), can identify which factors, such as test results, demographics, or medical history, contributed most to the model's prediction. This information helps clinicians better understand the basis for risk assessments and make more personalized treatment decisions.

XAI is also crucial for diagnostic decision support[14]. By providing reasoning behind an AI model's predictions, XAI helps clinicians understand how a model arrived at a specific diagnosis. For instance, if an AI model suggests a diagnosis of pneumonia, it can explain that the prediction was based on factors such as chest X-ray images, temperature readings, and cough symptoms. This transparency allows healthcare professionals to validate the AI's recommendations and ensure the accuracy of their clinical decisions.

Another area where XAI plays a critical role is in clinical text summarization[14]. In clinical environments, summarizing medical records, radiology reports, or discharge summaries can be a time-consuming task. Transformer-based models, like BERT or GPT, are frequently used to generate summaries, and XAI techniques, such as attention mechanisms, can highlight which parts of the text, such as key phrases or sentences, most influenced the summary. This improves the transparency of AI-driven text processing in clinical contexts, making the results easier to interpret and trust.

3.2. Overview of Existing XAI Techniques

Several XAI techniques, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), have been developed to provide transparency in clinical decision-making and enhance the interpretability of AI models. SHAP and LIME are model-agnostic approaches that explain individual predictions by quantifying the contribution of each feature. In healthcare, these techniques can be used to explain risk scores or diagnosis predictions based on patient data. For example, SHAP can show how a patient's age, blood pressure, and previous medical history contribute to a prediction of heart disease risk, while LIME can generate interpretable explanations for specific predictions by approximating the model locally. Attention mechanisms, often used in deep learning models, highlight which parts of the input data (e.g., text or images) are most important for a prediction. In clinical settings, attention mechanisms can be applied to models processing medical texts or images to pinpoint key features influencing decisions, providing clinicians with a visual representation of the model's focus areas. Additionally, counterfactual explanations offer insights by showing how changes in input data would lead to different model outcomes. In clinical decision support, counterfactuals can demonstrate how modifying a patient's treatment plan or clinical data would affect the risk prediction or diagnostic outcome, helping clinicians understand the potential impact of various decisions. These techniques, applied to NLP models, are crucial in enhancing the transparency of AI-driven decision-making in healthcare, fostering trust and better decision support for clinicians [3].

3.3. Gaps in Explainability for Neural Networks in Text Summarization

Despite the significant progress in the development of neural network models for clinical text summarization, there remain notable gaps in explainability that hinder the full adoption and trust of these models in clinical settings. One of the primary challenges is the "black-box" nature of many advanced neural network models, which makes it difficult for clinicians to understand how the model generates summaries or selects key information. This lack of transparency can undermine the reliability of AI systems, especially in critical clinical contexts where errors can have serious consequences.

A key gap is the limited interpretability of attention mechanisms, which are widely used in models like transformers for summarization tasks. While attention mechanisms can identify which parts of the text are most influential in generating a summary, they do not provide a clear rationale for why certain sections were prioritized over others. For example, a model may highlight specific symptoms or treatments in a summary, but without a deeper explanation, it remains unclear whether the model is making decisions based on medical relevance, statistical patterns, or other factors. This opacity can lead to confusion and reduce clinician confidence in the system's recommendations.

Additionally, current techniques for model interpretation often struggle to account for the complex and context-dependent nature of medical texts. Clinical documents, such as patient histories and discharge summaries, contain nuanced information that can be influenced by a variety of factors, including the patient's unique medical history, current health status, and environmental factors. Neural networks may aggregate these data points in ways that are not always intuitive, and existing interpretability methods may fail to capture the full context of the decision-making process.

Furthermore, while models such as BERT and GPT have demonstrated strong performance in generating clinical summaries, the explanations of these models often do not align with human reasoning. Clinicians may struggle to relate the model's output to their own diagnostic process, which is typically based on years of training and clinical experience. The mismatch between AI-generated summaries and clinical intuition can create barriers to effective collaboration between AI systems and healthcare professionals.

Finally, the absence of standardized frameworks for evaluating the explainability of clinical text summarization models contributes to the gap in understanding. Without clear metrics or guidelines for assessing the transparency of these models, it is difficult for researchers and clinicians to gauge the effectiveness of XAI techniques or to compare different models across studies.

Addressing these gaps is crucial for improving the trustworthiness and effectiveness of neural network-based clinical text summarization systems. By developing more interpretable models, enhancing the transparency of decision-making processes, and aligning AI outputs with human reasoning, researchers can create more reliable and user-friendly tools for clinicians.

4. Evaluation and Metrics for Clinical Text Summarization

The effectiveness of neural network models in automatic text summarization is determined not only by their ability to generate coherent and concise summaries but also by how well they capture essential information from the source text. In the clinical domain, the accuracy and relevance of the generated summaries are paramount, given the complexity and sensitivity of medical data [15]. Therefore, appropriate evaluation metrics are necessary to ensure that these models meet the high standards required for clinical applications.

In this section, we explore various methods for evaluating summarization models, including widely adopted metrics in general text summarization, as well as specialized metrics tailored to the clinical domain. Additionally, with the growing emphasis on explainable AI (XAI), we discuss metrics that assess the interpretability and transparency of model outputs, which are critical for the practical use of AI systems in healthcare [16]. This comprehensive evaluation framework is crucial for identifying the strengths and limitations of existing models and guiding future improvements in neural network architectures for clinical text summarization.

4.1. Evaluating Neural Networks in Text Summarization

The evaluation of neural networks in automatic text summarization is a crucial component in determining the effectiveness of models across different tasks. Both qualitative and quantitative metrics are used to assess how well a model captures the salient points of the source text and generates coherent, accurate summaries. In the context of clinical text summarization, where precision and information retention are paramount, robust evaluation metrics are essential for validating the performance of these models.

4.1.1. Common Evaluation Metrics

Several standard metrics are employed to evaluate text summarization models, most notably:

- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** ROUGE is one of the most widely used metrics in text summarization, especially for extractive methods. It measures the overlap between the generated summary and reference summaries, primarily focusing on n-grams, word sequences, and word pairs [2]. ROUGE-n, ROUGE-L (longest common subsequence), and ROUGE-W (weighted longest common subsequence) variants are often used. While useful, ROUGE can sometimes overestimate similarity due to surface-level matching, making it less effective in evaluating abstractive summaries [5].
- **BLEU (Bilingual Evaluation Understudy):** Originally designed for machine translation, BLEU scores are also employed in text summarization. BLEU evaluates how closely a generated text matches a reference text based on the precision of n-grams [3]. However, BLEU is less commonly used in summarization compared to ROUGE, as it often favors shorter and more concise summaries, which might not be ideal for tasks requiring detailed content preservation.
- **METEOR (Metric for Evaluation of Translation with Explicit ORdering):** METEOR is another metric that, like BLEU, is used primarily in machine translation but has applications in text summarization. METEOR accounts for synonymy and stemming, making it more flexible than ROUGE and BLEU for abstractive summaries. It assigns higher scores to summaries that capture meaning, even when surface-level similarity is lower [8].

4.1.2. Domain-Specific Evaluation in Clinical Summarization

In clinical summarization, accuracy and relevance of the information retained in summaries are critical. Therefore, in addition to the aforementioned metrics, several domain-specific evaluation methods are utilized:

- **Precision and Recall:** These metrics are particularly important in clinical summarization, where false positives (irrelevant information) and false negatives (missing key clinical details) can have serious consequences. Precision evaluates the relevance of the information included, while recall assesses the completeness of the summary [7]. High precision and recall are necessary to ensure the generated summaries provide accurate and complete overviews of clinical reports.
- **F1-Score:** As a harmonic mean of precision and recall, the F1-score provides a balanced metric to evaluate models in clinical domains where both over-inclusion of irrelevant details and omission of critical information are unacceptable. This metric is particularly useful when optimizing both precision and recall [9].
- **Clinical Relevance and Expert Review:** Clinical text summarization models are often evaluated through expert review, where medical professionals assess the relevance and quality of the generated summaries [6]. These evaluations are crucial, as automated metrics alone may not fully capture the clinical importance of the information presented. For example, a summary that scores well on ROUGE may still fail to meet clinical relevance if key findings are omitted.

4.2. Explainability and Interpretability Metrics

With the increasing demand for explainable AI (XAI) in healthcare, new metrics are being developed to evaluate the transparency and interpretability of neural network models. Some emerging metrics include:

- **SHAP (SHapley Additive exPlanations):** SHAP values provide insight into the contributions of individual features or words in a model's decision-making process, allowing healthcare professionals to understand why certain text segments were included or omitted in the summary [3].
- **LIME (Local Interpretable Model-agnostic Explanations):** LIME generates local approximations of a model's behavior, explaining why specific summary decisions were made. This is particularly useful in clinical settings, where trust in AI systems depends on the interpretability of the model's output [3].
- **Transparency and Trustworthiness Scores:** In clinical settings, AI models are increasingly evaluated for their interpretability using trustworthiness scores based on qualitative assessments by medical professionals. These scores help quantify the level of trust that clinicians can place in the AI's decisions, which is vital for the adoption of these models in practice [6].

4.3. Challenges in Evaluation

While existing metrics provide valuable insights into model performance, there are still challenges in evaluating text summarization models, particularly in the clinical domain. For instance, most automated metrics like ROUGE and BLEU do not adequately capture semantic accuracy or the clinical importance of information. Moreover, the subjective nature of clinical relevance often requires expert validation, which can be resource-intensive. Additionally, the increasing complexity of neural network models makes explainability a growing concern, as traditional metrics do not account for the interpretability of model decisions. Therefore, future evaluation efforts should focus on developing more comprehensive metrics that combine automated evaluation with expert review, while also incorporating transparency and interpretability measures to ensure the clinical viability of summarization systems.

5. Discussion: Bridging the Gap Between AI Technology and Clinical Practice

5.1. Key Findings

The studies reviewed demonstrate significant advances in the automatic summarization of clinical trial data, both for extractive and abstractive methods. The *TextRank* algorithm performed well for extractive summarization tasks, offering concise, meaning-preserving synopses of clinical trial descriptions. Similarly, ontological and domain-specific models, such as those incorporating *BERTSUMEXT* and other neural models, have demonstrated improvements in abstractive summarization, as seen in [17].

A key advantage in this domain is the use of neural architectures, particularly *transformers* and *pre-trained models*, to address the challenges of clinical text summarization. In particular,it achieved strong results with multi-objective optimization techniques, leveraging sentence position and similarity metrics like: Term Frequency-Inverse Document Frequency (*TF-IDF*) and *Word Mover's Distance (WMD)*. These methods not only outperform human gold standard summaries in some cases but also offer practical utility in new disease settings, such as during the emergence of novel diseases like COVID-19.

One common challenge noted in many papers was the tendency of abstractive models to introduce factual inaccuracies, particularly when summarizing biomedical evidence across multiple documents.

5.2. Limitations

While the advancements in automatic summarization are promising, there remain significant limitations. First, the lack of transparency and controllability in many neural models limits their practical utility in medical settings. Although systems like *TrialsSummarizer* aim to enhance user verification by allowing traceability of generated tokens, more work is needed to ensure end-users can reliably trust the outputs.

Another critical limitation lies in the evaluation methods for summarization quality. The heavy reliance on *ROUGE scores* is not always aligned with the preferences of healthcare professionals. This gap between automatic and manual evaluations highlights the need for more domain-specific evaluation metrics that better reflect clinical relevance and accuracy. Moreover, inter-rater agreement among human evaluators often varies.

5.3. Future Directions

Looking ahead, the integration of *Explainable AI (XAI)* approaches holds great potential for improving the transparency and usability of clinical summarization models. For instance, enhancing the traceability of neural model outputs—where users can see which portions of the input data contribute to each generated sentence—could improve trust in automated summaries. Additionally, further fine-tuning models with clinical-specific ontologies (e.g., *UMLS*) could improve both the factual accuracy and content relevance of generated summaries, as demonstrated in [17].

Another promising direction is the exploration of *multi-document summarization systems* that can handle vast datasets, such as those found on *ClinicalTrials.gov*. For instance, the *EXACT* system shows that automated extraction can drastically reduce the time needed to summarize and analyze clinical trial data for meta-analyses.

Lastly, improving factual verification mechanisms within neural models should be prioritized, as this will be crucial for the real-world deployment of such systems in clinical decision-making contexts. Collaborating with medical professionals to ensure practical utility and relevance will be key in advancing from experimental models to operational clinical tools.

6. Conclusion

This paper has provided an in-depth exploration of the role of ATS in clinical applications, particularly within the context of neural network-based models. We reviewed the key advancements in these models, emphasizing their potential to enhance clinical decision-making processes by improving the efficiency and accuracy of summarizing complex medical texts. Despite the significant progress made, several challenges remain, including the adaptation of neural networks to the specific demands of the healthcare domain and the need for greater transparency in the decision-making process.

The paper also highlighted the growing importance of Explainable AI (XAI) in clinical decision support. By incorporating XAI techniques, such as SHAP, LIME, and attention mechanisms, into clinical text summarization, AI systems can provide valuable insights into their predictions, fostering trust and aiding clinicians in making informed decisions. However, the gap in explainability for neural networks, particularly in clinical text summarization, persists and demands further investigation. Furthermore, we discussed the current limitations of existing evaluation metrics and the need for domain-specific measures that better capture the nuances of clinical text summarization. As medical data becomes more complex, the development of robust evaluation frameworks will be critical to ensuring that ATS systems deliver high-quality, clinically relevant summaries.

In conclusion, while there is great potential for neural network-based ATS systems in clinical settings, future research must address the challenges related to model explainability, evaluation metrics, and the practical application of these technologies. By doing so, we can further enhance the effectiveness of ATS systems and support clinicians in delivering optimal patient care.

7. Declaration on Generative AI

During the preparation of this work, the authors used the free version of ChatGPT for grammar and spelling checks, as well as for some reformulations to improve readability.

After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content, in accordance with the for the publication's content.

References

[1] A. Abdulnazar, R. Roller, S. Schulz, M. Kreuzthaler, Large language models for clinical text cleansing enhance medical concept normalization, IEEE Access 12 (2024) 147981–147990. doi:10.1109/ACCESS.2024.3472500.

[2] B. Baykara, T. Güngör, Turkish abstractive text summarization using pretrained sequence-to-sequence models, Natural Language Engineering (2022).

[3] K. T. Chitty-Venkata, M. Emani, V. Vishwanath, A. Somani, Neural architecture search for transformers: A survey, IEEE Access (2022).

[4] T. Ma, Q. Pan, H. Rong, Y. Qian, Y. Tian, N. Al-Nabhan, T-bertsum: Topic-aware text summarization based on bert, IEEE Transactions on Computational Social Systems (2022).

[5] M. Moradi, G. Dorffner, M. Samwald, Deep contextualized embeddings for quantifying the informative content in biomedical text summarization, Computer Methods and Programs in Biomedicine (2020).

[6] L. Tang, Z. Sun, B. Idnay, J. Nestor, A. Soroush, P. A. Elias, Z. Xu, Y. Ding, G. Durrett, J. F. Rousseau, C. Weng, Y. Peng, Evaluating large language models on medical evidence summarization, NPJ Digital Medicine (2023).

[7] K. Bressem, L. Adams, R. Gaudin, D. Tröltzsch, B. Hamm, M. Makowski, C. Schüle, J. Vahldiek, S. Niehues, Highly accurate classification of chest radiographic reports using a deep learning natural language model pre-trained on 3.8 million text reports, Bioinformatics (2020).

[8] M. Afzal, F. Alam, K. Malik, G. Malik, Clinical context–aware biomedical text summarization using deep neural network: Model development and validation, Journal of Medical Internet Research (2020).

[9] S. Gao, M. Alawad, M. T. Young, J. P. Gounley, N. Schaefferkoetter, H.-J. Yoon, X.-C. Wu, E. Durbin, J. Doherty, A. Stroup, L. Coyle, G. Tourassi, Limitations of transformers on clinical text classification, IEEE Journal of Biomedical and Health Informatics (2021).

[10] N. Y. Murad, M. H. Hasan, M. H. Azam, N. Yousuf, J. S. Yalli, Unraveling the black box: A review of explainable deep learning healthcare techniques, IEEE Access 12 (2024) 66556–66568. doi:10.1109/ACCESS.2024.3398203.

[11] N. Khan, M. Nauman, A. S. Almadhor, N. Akhtar, A. Alghuried, A. Alhudhaif, Guaranteeing correctness in black-box machine learning: A fusion of explainable ai and formal methods for healthcare decision-making, IEEE Access 12 (2024) 90299–90316. doi:10.1109/ACCESS.2024.3420415.

[12] F. Giuste, W. Shi, Y. Zhu, T. Naren, M. Isgut, Y. Sha, L. Tong, M. Gupte, M. D. Wang, Explain-

able artificial intelligence methods in combating pandemics: A systematic review, IEEE Reviews in Biomedical Engineering 16 (2023) 5–21. doi:10.1109/RBME.2022.3185953.

[13] K. Ahnaf Alavee, M. Hasan, A. Hasnayen Zillanee, M. Mostakim, J. Uddin, E. Silva Alvarado, I. de la Torre Diez, I. Ashraf, M. Abdus Samad, Enhancing early detection of diabetic retinopathy through the integration of deep learning models and explainable artificial intelligence, IEEE Access 12 (2024) 73950–73969. doi:10.1109/ACCESS.2024.3405570.

[14] Z. Sadeghi, R. Alizadehsani, M. A. CIFCI, S. Kausar, R. Rehman, P. Mahanta, P. K. Bora, A. Almasri, R. S. Alkhawaldeh, S. Hussain, B. Alatas, A. Shoeibi, H. Moosaei, M. Hladík, S. Nahavandi, P. M. Pardalos, A review of explainable artificial intelligence in healthcare, Computers and Electrical Engineering 118 (2024) 109370. URL: https://www.sciencedirect.com/science/article/pii/S0045790624002982. doi:https://doi.org/10.1016/j.compeleceng.2024.109370.

[15] M. Kirmani, G. Kour, M. Mohd, N. Sheikh, D. A. Khan, Z. Maqbool, M. A. Wani, A. H. Wani, Biomedical semantic text summarizer, BMC Bioinformatics 25 (2024) 152. doi:10.1186/s12859-024-05712-x.

[16] I. A. E. Madi, A. Redjdal, J. Bouaud, B. Séroussi, Exploring explainable ai techniques for text classification in healthcare: A scoping review, Studies in Health Technology and Informatics 316 (2024) 846–850. doi:10.3233/shti240544.

[17] S. MacAvaney, et al., Ontology-aware clinical abstractive summarization, in: SIGIR'19, 2019, pp. 1013–1016.