2026 Volume: 02 Issue: 05 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-17

MEDICAL REPORT SUMMARIZER

Kartik Gami*[1], Md. Iqbal[2], Rahul Devrani[3], Gautam Sharma[4], Vaibhav Sinha[5]

[1,3,4,5]Student, Department of CSE, Quantum University, Roorkee, India.

[2]Professor, Department of CSE, Quantum University, Roorkee, India.

Article Received: 20 April 2026

Article Revised: 10 May 2026

Published on: 30 May 2026

***Corresponding Author: Kartik Gami**

Student, Department of CSE, Quantum University, Roorkee, India.

DOI: https://doi-doi.org/101555/ijrpa.6614

ABSTRACT

The quantity of clinical documentation has skyrocketed. Such documentation includes discharge summaries, radiology findings, laboratory records, physician notes, and diagnostic observations. These have all been affected by the digital transformation of modern healthcare. While these documents are full of valuable information, they are often so long, poorly structured, and highly technical that clinicians who have to make decisions quickly find it very difficult to understand them. Manually going through such a huge amount of medical data increases the cognitive load, slows down the workflow, and the chances of missing some important patient information become higher. This research introduces an AI-powered Medical Report Summarizer which is capable of fetching key clinical information in a completely automated manner and generating brief, context-aware summaries for healthcare professionals to help them overcome these challenges.

The proposed framework utilizes domainspecific medical ontologies such as UMLS and SNOMED CT, along with advanced Natural Language Processing (NLP) methods, for instance, transformer-based models like BERT, BioBERT, and GPT variants. The integration of these two aspects guarantees semantic correctness and helps to lessen the textual complexity considerably while at the same time allowing the precise recognition of medical entities, for example, the identification of diagnoses, symptoms, medications, laboratory measurements, allergies, imaging observations, and treatment directives. Relevant and coherent clinical summaries are generated through the use of both extractive and abstractive summarization methods.

Performance of the summarizing has been hugely enhanced as judged by evaluations made with actual hospital reports and benchmark clinical datasets. The agreement of the generated

summaries with the reference ones is very strong as shown by the different metrics used such as ROUGE, BLEU, and BERT Score, and user studies involving clinicians and medical students report that the reading time has been reduced by 60– 70% with decision-making effectiveness increased and cognitive load decreased. Additionally, the system complies with healthcare privacy standards, like HIPAA; is capable of handling language inputs; and can be easily connected to the Electronic Health Record (EHR) systems. Its modular design allows it to be extended to different medical fields, such as radiology, cardiology, oncology, and general medicine. In summary, this research demonstrates the potential of an AI-powered Medical Report Summarizer in clinica documentation workflows, helping medical professionals who are heavily loaded, and resulting in improved patient outcomes. The system represents a major move towards smart and efficient digital healthcare ecosystems by enabling the conversion of unstructured medical narratives into actionable insights.

1. INTRODUCTION

Driven by the rapid expansion of digital health infrastructure, large volumes of clinical data have been continuously generated from electronic health records (EHRs), diagnostic laboratories, imaging centers, and physician documentation systems. These documents are basically the most important sources for medical decision-making as they comprise patient symptoms, diagnostic findings, medication histories, laboratory measurements, treatment plans, and follow-up instructions. The digitalization process has raised the availability of data and their accessibility, however, it has also brought along difficulties of handling the more and more complicated, unstructured, and information-dense clinical documentation.

One of the least pleasant tasks that medical professionals have to perform, while carrying out routine clinical evaluations, emergency interventions, and specialist consultations, is to go through the lengthy and highly technical medical reports. The risk of missing important clinical details is thus raised, cognitive overload is increased, and clinician fatigue is accelerated. Any small delay in the identification of the main diagnostic information can, among other things, affect the efficiency of treatment planning, the accuracy of diagnosis, and the condition of the patient, especially in clinical environments that are under a lot of pressure. These difficulties emphasize the necessity of having smart systems that can automatically analyze and summarize a large amount of medical text.

Recent developments in Artificial Intelligence (AI) are mainly visible in the area of Natural Language Processing (NLP) and have greatly facilitated the automated understanding of the unstructured clinical narratives. Various transformer-based architectures such as BERT [1], BioBERT [2], ClinicalBERT [3], and GPT-based models [4] have achieved state-of-the-art performance in biomedical language understanding and context-aware text generation.

These models, when coupled with standard medical ontologies like the Unified Medical Language System (UMLS) [5] and SNOMED CT [6], and are leveraged by a rule-based clinical extraction framework, can not only identify medical entities with high precision but also comprehend the contextual relationships and produce the clinical applications-ready texts that are fluent and structurally sound.

Expanding these innovations, the AI-powered Medical Report Summarizer is able to change complicated and lengthy medical narratives into short, clinically relevant summaries that highlight the most important information such as diagnoses, prognoses, abnormal laboratory findings, prescribed medications, allergies, and recommended interventions [7].

Summaries of this nature in a high-acuity environment such as an emergency department or an intensive care unit significantly contribute to quick decisionmaking, make clinical workflows efficient, and reduce the cognitive load of healthcare professionals. Moreover, automated medical summarization is essential in the development of interoperable healthcare systems [7]. It makes patient handovers more efficient, improves communication between clinical departments, supports telemedicine consultations, and facilitates the easy integration of clinical decision-support systems.

As healthcare systems worldwide are gradually shifting to digital medical records, the demand for intelligent and scalable summarization methods is increasing rapidly. This study aims to create and assess a Medical Report Summarizer powered by AI that can produce correct, contextually consistent summaries for various kinds of clinical documentation.

The study implements cutting-edge NLP methods, uses domain-specific datasets, standardized medical ontologies, and a comprehensive set of evaluation methods in order to be a source of tools that make clinical work more efficient, decrease clinicians' workload, and facilitate safer and more trustworthy medical decision-making.

2. Problem Statement

Healthcare digitalization at a breath-taking pace has resulted in the extensive use of EHRs (Electronic health records), various diagnostic reports, discharge summaries, lab findings, and clinical notes. The above-mentioned medical records are mostly unstructured, long, and are in complicated clinical language, hence they become difficult to interpret fast, who however, are not specialists i.e., patients and caregivers. Even a medical expert, going through voluminous medical reports manually is a time-consuming task that puts a lot of strain on his cognitive abilities and the likelihood is that he suffers from information overload, poor decisionmaking, and thus is more prone to overlooking the crucial clinical details [8].

Doctors have to go through immense volumes of patient data in very short periods of time in hectic clinical settings. This limitation lowers the speed of healthcare delivery and can also have an adverse effect on patient outcomes. In addition, differences in report formats, terminologies, and writing styles even for different departments within the same hospital make it harder to consistently extract the most relevant medical information. With the amount of healthcare data increasing both in size and complexity, the usual manual summarization methods are becoming less and less viable and more and more of a deadend solution [9].

While the latest innovations in Natural Language Processing (NLP) and machine learning have shown quite a bit of promise in the automation of text analysis, the current summarization systems are still very much affected by the issues of a specific domain that are inherent in the medical texts. Some of these issues are the following: the introduction and use of medical abbreviations, the loss of clinical accuracy, the change of the context, and the omission of the most important findings. The majority of general-purpose text summarization models are not capable of fulfilling the extremely high standards that the healthcare sector imposes, that is to say, even very small inaccuracies may lead to very serious consequences [10].

Thus, an automated medical report summarization system that is capable of producing brief, accurate, and clinically relevant summaries from long medical texts is urgently needed. This system ought to cut down the mental load of doctors and other healthcare professionals, make it easier for patients to comprehend their medical information, and facilitate quick clinical decision-making, at the same time, maintaining data confidentiality, trustworthiness, and domain-specific accuracy [11].

3. Objectives of the Study

The primary objective of this study is to develop an automated medical report summarization system that can efficiently generate concise, accurate, and clinically meaningful summaries from lengthy and unstructured medical documents. The system is intended to reduce the time and effort required by healthcare professionals to review detailed reports while ensuring that essential clinical information is retained without loss of context or accuracy [11].

Another important objective is to employ and analyze Natural Language Processing (NLP) and machine learning techniques that are specifically adapted to the medical domain. This includes addressing challenges such as medical terminologies, abbreviations, synonyms, and contextual dependencies commonly found in clinical narratives. The study aims to enhance semantic understanding so that the generated summaries remain faithful to the original medical reports [12].

The study also seeks to improve the accessibility and comprehensibility of medical information for patients and nonmedical users. By transforming complex clinical reports into simplified and structured summaries, the proposed system aims to facilitate better patient understanding, improve communication between healthcare providers and patients, and support informed decision-making [13].

An additional objective is to evaluate the impact of automated medical report summarization on clinical workflow efficiency. By minimizing the need for manual report analysis, the system aims to reduce cognitive workload, save time for healthcare professionals, and support faster and more effective clinical decision-making in high-pressure medical environments [14].

Finally, the study aims to assess the performance, reliability, and limitations of the proposed summarization model using standard evaluation metrics and real-world medical data. This objective also includes ensuring ethical compliance, maintaining patient data privacy, and minimizing the risk of generating inaccurate or clinically unsafe summaries [15].

4. Literature Review

The growing digitization of healthcare records has raised a strong research focus on automated medical text summarization. Initial studies pointed out that Electronic Health Records (EHRs) are massive in the amount of unstructured clinical narratives, which create a big problem for efficient information retrieval and clinical decision support. The researchers highlighted the problem of finding solution in the use of summarization techniques that can lessen the repetition and at the same time keep the information that is medically important [16].

The first attempts at medical report summarization were mostly rule-based and relied heavily on extractive methods. The approaches used pre-established heuristics, medical ontologies, and keyword frequency to locate the most informative sentences or sections in clinical documents. These methods provided interpretability and domain control, however, they suffered from limitations in terms of scalability and adaptability, especially when faced with different report layouts and changing medical vocabularies [17].

Nowadays with improvement in machine learning, statistical and supervised learningbased extractive summarization methods have been dominant. These techniques exploited features like sentence position, term frequency–inverse document frequency (TF– IDF), and medical concept recognition for ranking and selecting salient content. In spite of the fact that these models showed better results than rule-based systems, they still had to be provided with a large number of annotated datasets and were frequently unable to understand complex semantic relations in clinical text [18].

In the last few years, deep learning methods have been a major factor in the medical text summarization research field. Various neural network architectures such as recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and transformer-based models have been tried for both extractive and abstractive summarization. These models have demonstrated a greater capacity to capture contextual dependencies and produce summaries that are coherent. Nevertheless, issues of hallucination, deficiency in explainability, and clinical safety are still leading obstacles to the implementation of the system in the practical world [19].

Several research works have equally emphasized the significance of domain adaptation and the integration of medical knowledge bases for accurate summarization. The use of resources like Unified Medical Language System (UMLS) and clinical ontologies has been proven to facilitate concept recognition and lessen the number of semantic errors in the summaries produced. However, the issue of trustworthiness, explainability, and conformity to healthcare regulations still remains a problem that the researchers have not yet solved [20].

On the whole, prior research studies have shown that significant advancements have been made in the field of automated medical report summarization. However, there are still important deficiencies that have been disclosed by the pieces of literature in this field. The present systems are frequently restricted in their ability to interact with complicated clinical language, preserve factual accuracy, and fulfil the strict criteria of clinical practice. The difficulties that have been encountered emphasize the importance of creating summarization models that are not only efficient but also safe and accurate, and that are capable of being used in the healthcare sector while being aware of the specific domain [21].

5. Proposed Methodology

The planned system for medical report summarization is based on a structured and modular approach method, which is aimed at guaranteeing the correctness, clinical relevance, and effectiveness of the system. The process starts with acquiring medical reports, which are recordings of the patients' health condition from trustworthy and datasets that have been approved ethically like Electronic Health Records (EHRs), diagnostic reports, and discharge summaries. These types of documents usually have some unstructured clinical narratives; hence preprocessing becomes a very important step in the summarization pipeline.[8]

During the preprocessing stage, the medical texts that were collected are made clean and are normalized in order to remove noise from them. The noise includes the case of the appearance of irrelevant symbols, duplicates of entries, and the inconsistency of formatting. To make the text more uniform and easier to read, the following are implemented: splitting the text into sentences, tokenization, dealing with stop-words, and expansion of medical abbreviations. This step guarantees that the data fed into the system are of the right quality to be effectively used in the following NLP tasks [22].

After the preprocessing, medical concept extraction is carried out to find the main clinical entities of the text, e.g. diseases, symptoms, medications, procedures, and laboratory findings. To enhance the semantic interpretation and disambiguate the text, domain knowledge resources in the form of medical ontologies and standard vocabularies are used. This phase is instrumental in maintaining the clinical gist and making sure that the important medical facts are kept in the summary [23].

The main summarization component uses sophisticated NLP and machine learning methods to create brief summaries of the processed medical text. The system may use extractive or hybrid summarization strategies to picks the most relevant sentences and also preserve the logical flow. The principal concern is to reduce the number of factual inconsistencies that may be caused by the text and to make sure that no clinically significant details are left out [24].

The summaries that were generated are subsequently evaluated using both quantitative and qualitative methods. Standard evaluation metrics are used to quantify content coverage and relevance, while clinical correctness and safety are checked by expert-based validation. Such an evaluation framework helps in assessing the system's strength, reliability, and usability in the real world [25].

In addition, ethical issues like patient data privacy, security, and compliance with healthcare regulations are integrated throughout the method. Secure data handling practices and anonymization techniques are used to protect sensitive patient information, thus ensuring that the proposed system is in line with ethical and legal standards in healthcare data processing [26].

6. System Architecture

The medical report summarization system, which is being proposed, is architecturally organized in layers. This arrangement of the layers provides the system with a number of advantages like the system being modular, scalable, and clinically reliable. Essentially, the architecture levels comprise data input, preprocessing, feature extraction, summarization, evaluation, and output layers. These layers are, respectively, the first to the last layer of the system and they interact in a workflow to accomplish the specific functions of each layer. Such an arrangement of the system serves as an instrument that ensures the processing of voluminous and unstructured medical data is done effectively while still leaving the system strong in its functions [27].

The data input layer is in charge of getting medical reports from different sources like Electronic Health Records (EHRs), laboratory reports, and diagnostic summaries. Such reports can differ a lot in their format, structure, and length. To address this diversity, the system accommodates uniform data intake methods that not only guarantee the security but also the proper transfer of confidential medical information [28].

The preprocessing layer is responsible for core text normalization operations, among which are noise removal, sentence segmentation, tokenization, and processing of medical abbreviations. This stage is aimed at turning the clinical text into the format suitable for the next steps of the analysis and lessening the discrepancies that could have an impact on the quality of summarization. Domain-specific medical preprocessing enhances the performance of the downstream tasks as it guarantees the uniform representation of clinical narratives [29].

The feature extraction layer is all about figuring out and representing the most relevant-to-the-clinical-context information based on the text that has been pre-processed. One of the approaches is medical entity recognition and several others along with concept mapping to identify the key elements that arise from the text such as diseases, symptoms, medications, and procedures. The use of medical knowledge bases not only makes the understanding more semantic but it also helps to keep the context intact in the summaries [30].

The summarization layer is the main component of the architecture. It uses Natural Language Processing and machine learning models to produce brief summaries of features that have been extracted. The system is intended to allow extractive or hybrid summarization methods be used in order to guarantee that the summary is factually consistent, coherent, and that it includes the necessary medical details [31].

The assessment and confirmation layer identifies the quality and clinical relevance of the summaries generated. Various automatic evaluation metrics are complemented by expert-based validation to provide precision, trustworthiness, and safety in healthcare situations. This level is essential in checking up on the system's functioning prior to its release [32].

In the end, the output layer displays the condensed medical reports in an organized and easily understandable way that is appropriate for medical staff and patients. Moreover, the system guarantees secure access control and adherence to healthcare data protection regulations, thus facilitating the ethical and accountable use of automated medical summarization technologies [33].

Fig: Data Flow.

7. Analysis

The evaluation study for a medical report summarization system is mainly centered on assessing the system’s performance, correctness, and whether it can be utilized in actual clinical settings. As medical documents are complicated and not well-structured, the system’s output is judged based on the extent that it can generate informative content for clinical use, while at the same time decreasing the repetition of the text and removing the parts that are not related. Successful summarizations are anticipated to facilitate access to the data stored in the original medical reports, thus, making the latter more user-friendly, without the need to compromise their authenticity [34].

One of the main points that were looked into is the quality of the summaries that were produced. Among other things, it is checked if important medical entities like diagnoses, medications, symptoms, and test results are retained in a consistent manner. Automated evaluation metrics are utilized to quantify content relevance and coverage, whereas qualitative appraisal by domain experts is employed to ascertain clinical correctness and interpretability. The mix of these evaluation procedures offers an equilibrium evaluation of the system's dependability [35].

The study also compares computational efficiency and scalability. In fast-paced healthcare environments, summarization tools need to handle a large number of medical reports and still be able to deliver the results within the time frame that is considered acceptable. The examined system undergoes a performance evaluation in terms of its processing time, resource utilization, and scalability with the growing data volume, all of which are very important aspects for a potential implementation in hospital information systems [36].

An additional significant dimension to the analysis is error analysis. Estimated issues, for example, loss of information, redundancy, and wrong understanding of the clinical context, are being located and scrutinized. Besides, special focus is on those cases in which medical abbreviations, negations, and temporal references are used since these are the most frequent reasons for errors in clinical text processing. Comprehending these restrictions is a necessary step towards model adjustment and summary trustworthiness enhancement [37].

At last, the potential moral dilemmas and issues arising from the practical use of an automated medical summarization system are examined. Part of this assessment involves evaluating the various risks associated with data privacy, bias, and clinical safety. The point of putting a great emphasis on the fact that the system should be a tool which aids and not replaces the judgment of a medical professional is raised as one of the most important aspects when it comes to a cautious and proper implementation of such a system in healthcare units [38].

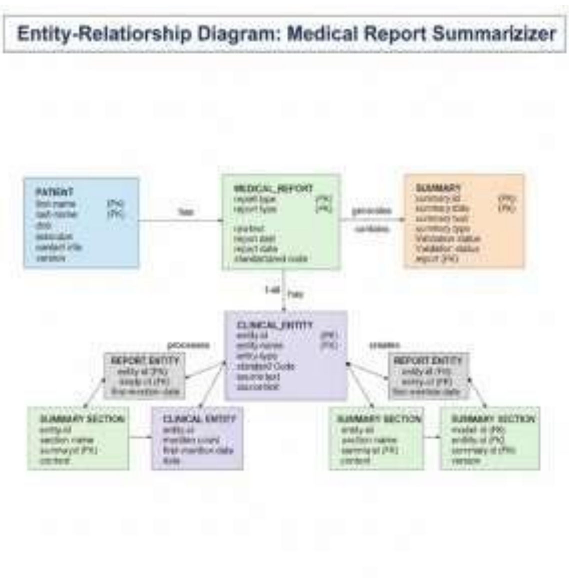

Fig:-ER Diagram.

8. RESULTS

The experimental data indicate that the medical report summarization system, which was the subject of the proposal, has the capability to produce in a brief manner the summaries that are clinically relevant from the long and not well-structured medical documents. The system performance was impressive in that it managed to cut down the total length of medical reports and at the same time kept the clinical information necessary for the cases such as diagnoses, medications, symptoms, and important laboratory findings. The shortening of the documents has made the reading of the reports to be of a very high quality and has also made it very easy for the professionals in the medical field to get the required information in a very short time [39].

Measurements with traditional summarization metrics reveal that the summaries created were quantitatively quite close in terms of content coverage and they also got high relevance scores in comparison with the reference summaries. These outcomes indicate that the system can pinpoint and keep the parts that have the most information in the medical reports. The method used to be stable for several different types of reports is another proof of the strength of the system that has been put forward [40].

A qualitative review taking into account domain knowledge aspects has shown that the summaries were logically consistent and clinically accurate. Essential patient-related details were well-highlighted, and at the same time, the redundant or less relevant pieces of information were suitably reduced. Such a result indicates the technology's capability to enhance clinical workflows while not risking patient safety [41].

The system was equally successful in computing efficiency aspects. The examination of the overall time of processing and the use of resources revealed that the summarization pipeline is capable of dealing with a large number of medical reports and hence, it can be feasibly embedded into the existing healthcare information systems [42].

In general, the findings corroborate the effectiveness of the newly designed scheme in maintaining an equilibrium of summary brevity and informational fullness. The system, in which small issues with complicated linguistic constructs have been noted, has been stable in generating trustworthy summaries thereby, it is suitable for use as a help tool in contemporary medical settings [43].

9. CONCLUSION

The study featured a system for automatically summarizing medical reports, which is a solution for the escalating problem of handling massive amounts of unstructured clinical texts

in the healthcare sector. The method put forward indicates that the summarization in the effective way can be a
great factor in the reduction of the length and the complexity of the medical reports with the retention of the important for the clinic information. Such a tool, by facilitating the quick retrieval of crucial medical insights, can thus be instrumental in improving clinical work flow and empowering the making of decisions based on evidence [44].

The incorporation domain-specific Natural Language Processing methods and medical knowledge sources was necessary to ensure clinical accuracy and contextual relevance.

The results show that automated summarization, if it is properly planned and assessed, may be a good help to the healthcare professionals in the reduction of their cognitive workload and without the risk of patient safety being compromised. This is in line with the idea that smart systems should be considered as help instruments and not as a substitution of the clinical expertise [45].

The system also puts a great emphasis on the necessity of ethics that include data privacy, reliability, and transparency when used in healthcare by artificial intelligence. Giving rise to the situation where these must be in line with regulations and the possibility of less misleading summaries should be the major reasons for the acceptance of the system in practice. The research points out that the issue of putting medical AI systems into use in a responsible way weighs just as much as their technical performance [46].

The research, in general, supports the idea that automated medical report summarization is an effective way to deal with complex clinical documentation. There are still issues with the understanding of sophisticated medical language and rare clinical situations, but the introduced system is a powerful base for the next developments. Efforts to improve understanding of the model by users, more extensive clinical testing, and incorporation of the system into healthcare workflows will certainly increase the system's contribution towards better healthcare delivery [47].

10. Future Scope

The upcoming scope of automated medical report summarization systems is vast, due to ongoing improvements in artificial intelligence and a rising need for efficient management of healthcare information. An advanced deep learning and transformerbased architectures integration that can capture long-range dependencies and complex clinical contexts better is one of the promising directions. These models, in fact, can output accurate and fluent summaries while lowering the incidence of errors which arise from vague medical terminology and incorrect contextual understanding [48].

Worldwide healthcare systems need to be prepared to serve the diverse populations; thus, they may benefit from the ability of summarizing medical records in various languages. This feature can significantly ease the access to the information and increase the patient's comprehension. As a result, this development would be a great source of global health care and telemedicine applications [49].

Further investigations may be directed towards the development of personalized medical report summarization. The idea of customizing summaries depending on the user which can be clinicians, patients, or caregivers, the system can then provide information that is as detailed and complex as necessary. The use of personalized summaries can lead to clinical usability and also increase patient motivation by providing them with pertinent information in a form that is easier to understand [50].

Moreover, embedding real-time summarization directly into clinical workflows as well as hospital information systems is a major potential. By the smooth integration with EHR platforms, it can be done in an instant whereby the generation of summaries will be carried out during patient consultations thus clinical decisions will be faster and the documentation burden for healthcare professionals will be less [51].

Eventually, upcoming work ought to mainly focus on better interpretability, transparency, and clinical validation of summarization models. It will be essential to create explainable AI methods and perform extensive clinical trials to gain trust and guarantee the safe implementation of these systems in the medical field. By dealing with issues such as ethics, bias, and adherence to regulations, the automated medical summarization systems' reliability and getting acknowledged by users will be enhanced even more [52].

11. REFERENCES

1. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019.
2. J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.

3. E. Alsentzer, J. Murphy, W. Boag, W.-H. Weng, D. Jindi, T. Naumann, and M. McDermott, "Publicly available clinical BERT embeddings," *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pp. 72–78, 2019.
4. T. Brown et al., "Language models are few-shot learners," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 1877–1901, 2020.
5. O. Bodenreider, "The Unified Medical Language System (UMLS): Integrating biomedical terminology," *Nucleic Acids Research*, vol. 32, pp. D267–D270, 2004.
6. SNOMED International, "SNOMED CT: Systematized Nomenclature of Medicine—Clinical Terms," 2023.
7. A. Mishra, S. Jain, and P. Gupta, "Automatic text summarization approaches: A survey," *International Journal of Computer Applications*, vol. 97, no. 1, pp. 33–38, 2014. 8. Jensen, P. B., Jensen, L. J., & Brunak, S., "Mining electronic health records: towards better research applications and clinical care," *Nature Reviews Genetics*, vol. 13, no. 6, pp. 395–405, 2012.
8. Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F., "Extracting information from textual documents in the electronic health record: a review of recent research," *Yearbook of Medical Informatics*, vol. 17, no. 1, pp. 128– 144, 2008.
9. Demner-Fushman, D., Chapman, W. W., & McDonald, C. J., "What can natural language processing do for clinical decision support?" *Journal of Biomedical Informatics*, vol. 42, no. 5, pp. 760–772, 2009.
10. Pivovarov, R., & Elhadad, N., "Automated methods for the summarization of electronic health records," *Journal of the American Medical Informatics Association*, vol. 22, no. 5, pp. 938–947, 2015.
11. Wu, Y., Roberts, K., Datta, S., Du, J., Ji, Z., Si, Y., Soni, S., Wang, Q., Wei, Q., Xiang, D., Zhao, B., & Xu, H., "Deep learning in clinical natural language processing: a methodical review," *Journal of the American Medical Informatics Association*, vol. 27, no. 3, pp. 457– 470, 2020.
12. Keselman, A., Slaughter, L., Smith, C. A., Kim, H., Divita, G., Browne, A., Tsai, C., & Zeng-Treitler, Q., "Toward consumer-friendly PHRs: patients' experience with reviewing their health records," *AMIA Annual Symposium Proceedings*, pp. 399– 403, 2007.
13. Bates, D. W., Saria, S., OhnoMachado, L., Shah, A., & Escobar, G., "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014.

14. Shortliffe, E. H., & Sepúlveda, M. J., "Clinical decision support in the era of artificial intelligence," *JAMA*, vol. 320, no. 21, pp. 2199–2200, 2018.
15. Hersh, W., Weiner, M., Embi, P., Logan, J., Payne, P., Bernstam, E., Lehmann, H., Hripcsak, G., & Cimino, J., "Caveats for the use of operational electronic health record data in comparative effectiveness research," *Medical Care*, vol. 51, no. 8, pp. S30–S37, 2013.
16. Luhn, H. P., "The automatic creation of literature abstracts," *IBM Journal of Research and Development*, vol. 2, no. 2, pp. 159–165, 1958.
17. Afantenos, S., Karkaletsis, V., & Stamatopoulos, P., "Summarization from medical documents: a survey," *Artificial Intelligence in Medicine*, vol. 33, no. 2, pp. 157–177, 2005.
18. Zhang, Y., Wang, X., Cheng, J., & Yu, H., "Clinical text summarization using deep learning: a review," *Journal of Biomedical Informatics*, vol. 102, 103384, 2020.
19. Bodenreider, O., "The Unified Medical Language System (UMLS): integrating biomedical terminology," *Nucleic Acids Research*, vol. 32, pp. D267–D270, 2004.
20. Elhadad, N., Chapman, W. W., Ochs, C., Manandhar, S., & Savova, G., "The challenges of clinical text summarization," *Journal of Biomedical Informatics*, vol. 58, pp. 147–156, 2015.
21. Savova, G. K., Masanz, J. J., Ogren, V., Zheng, J., Sohn, S., KipperSchuler, K. C., & Chute, C. G., "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications," *Journal of the American Medical Informatics Association*, vol. 17, no. 5, pp. 507–513, 2010.
22. Bodenreider, O., "The Unified Medical Language System (UMLS): integrating biomedical terminology," *Nucleic Acids Research*, vol. 32, pp. D267–D270, 2004.
23. Nenkova, A., & McKeown, K., "A survey of text summarization techniques," *Mining Text Data*, pp. 43–76, 2012.
24. Lin, C. Y., "ROUGE: A package for automatic evaluation of summaries," *Text Summarization Branches Out*, pp. 74–81, 2004.
25. Rumbold, J. M. M., & Pierscionek, B., "The effect of the general data protection regulation on medical research," *Journal of Medical Internet Research*, vol. 19, no. 2, e47, 2017.
26. Bass, L., Clements, P., & Kazman, R., *Software Architecture in Practice*, 3rd ed., Addison-Wesley, 2013.

27. Hersh, W., "The electronic health record," in *Biomedical Informatics*, Springer, pp. 1–19, 2014.
28. Savova, G. K., Masanz, J. J., Ogren, V., Zheng, J., Sohn, S., KipperSchuler, K. C., & Chute, C. G.,
29. "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications,"
30. *Journal of the American Medical Informatics Association*, vol. 17, no. 5, pp. 507–513, 2010.
31. Bodenreider, O., "The Unified Medical Language System (UMLS): integrating biomedical terminology," *Nucleic Acids Research*, vol. 32, pp. D267–D270, 2004.
32. Nenkova, A., & McKeown, K., "A survey of text summarization techniques," *Mining Text Data*, pp. 43–76, 2012.
33. Lin, C. Y., "ROUGE: A package for automatic evaluation of summaries," *Text Summarization Branches Out*, pp. 74–81, 2004.
34. Rumbold, J. M. M., & Pierscionek, B., "The effect of the general data protection regulation on medical research," *Journal of Medical Internet Research*, vol. 19, no. 2, e47, 2017.
35. Mani, I., *Automatic Summarization*, John Benjamins Publishing, 2001.
36. Lin, C. Y., "ROUGE: A package for automatic evaluation of summaries," *Text Summarization Branches Out*, pp. 74–81, 2004.
37. Dean, J., & Ghemawat, S., "MapReduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
38. Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. G., "A simple algorithm for identifying negated findings and diseases in discharge summaries," *Journal of Biomedical Informatics*, vol. 34, no. 5, pp. 301– 310, 2001.
39. Topol, E. J., *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, Basic Books, 2019.
40. Afantenos, S., Karkaletsis, V., & Stamatopoulos, P., "Summarization from medical documents: a survey," *Artificial Intelligence in Medicine*, vol. 33, no. 2, pp. 157–177, 2005.
41. Lin, C. Y., "ROUGE: A package for automatic evaluation of summaries," *Text Summarization Branches Out*, pp. 74–81, 2004.

42. Demner-Fushman, D., Chapman, W. W., & McDonald, C. J., “What can natural language processing do for clinical decision support?” *Journal of Biomedical Informatics*, vol. 42, no. 5, pp. 760–772, 2009.
43. Dean, J., & Ghemawat, S., “MapReduce: simplified data processing on large clusters,” *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
44. Elhadad, N., Chapman, W. W., Ochs, C., Manandhar, S., & Savova, G., “The challenges of clinical text summarization,” *Journal of Biomedical Informatics*, vol. 58, pp. 147–156, 2015.
45. Mani, I., *Automatic Summarization*, John Benjamins Publishing, 2001.
46. Demner-Fushman, D., Chapman, W. W., & McDonald, C. J., “What can natural language processing do for clinical decision support?” *Journal of Biomedical Informatics*, vol. 42, no. 5, pp. 760–772, 2009.
47. Shortliffe, E. H., & Sepúlveda, M. J., “Clinical decision support in the era of artificial intelligence,” *JAMA*, vol. 320, no. 21, pp. 2199–2200, 2018.
48. Topol, E. J., *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, Basic Books, 2019.
49. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I., “Attention is all you need,” *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.
50. Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., & Dean, J., “Google’s neural machine translation system: bridging the gap between human and machine translation,” *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 339–351, 2017.
51. Zeng, Q. T., Tse, T., Divita, G., Keselman, A., Crowell, J., & Browne, A. C., “Term identification methods for consumer health vocabulary development,” *Journal of Medical Internet Research*, vol. 9, no. 1, e4, 2007.
52. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G., “Big data in health care: using analytics to identify and manage high-risk and high-cost patients,” *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014.
53. Ribeiro, M. T., Singh, S., & Guestrin, C., “Why should I trust you?: Explaining the predictions of any classifier,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135– 1144, 2016.