

Medical Report Interpretation Using NLP

Ms. Amulya¹, Nagashree Baggon², Niharika Shambhu Gouda³, Prakrathi Shetty⁴, Suchitra M⁵

¹Assistant Professor Shree Devi Institute of Technology, Visvesvaraya Technological University

^{2,3,4,5}Department of Computer Science and Engineering, Shree Devi Institute of Technology, Visvesvaraya Technological University

Abstract—Medical report contain valuable clinical information in unstructured and semi-structured text, making manual interpretation time consuming for patient and healthcare professionals. This paper presents a Natural Language Processing (NLP) based system for medical report based on multiple disease and medical conditions. This system accepts medical documents such as PDF, DOCX, and TXT files and processes them using text processing, clinical entity extraction, terminology normalization, and contextual analysis. The system identify important medical data means symptoms, diseases, medications, investigations, procedures, and clinical findings. The extracted data is converted into semantic representation with patient-friendly explanations. In addition, a doctor-oriented research module allows authorized healthcare professionals to search and review symptoms and related clinical data while maintaining privacy and responsible use of clinical data. The system is intended as an information and clinical decision-support tool rather than an autonomous diagnostic system

Index Terms—Natural Language Processing, Medical Report Interpretation, Clinical information Extraction, Named Entity Recognition, Clinical NLP, Medical Text Mining, Clinical Decision Support.

I. INTRODUCTION

Medical reports contain important data about a patient's health, including symptoms, diseases, medications, laboratory investigations, clinical findings, and treatment related data. These reports often available as unstructured format such as clinical notes, diagnostic reports, discharge summaries, and laboratory data. Manually reviewing large amounts of medical text can require considerable time and may take it difficult to quickly recognize most related clinical data.[1], [3].

Natural Language Processing (NLP) give a technique for processing and understanding human language and become has become an important approach for analyzing clinical data or text. Clinical NLP can used

for task such as named entity recognition, information extraction, text classification, terminology normalization, and summarization [1], [3]. Previous research has demonstrated the application of deep learning and transformer-based methods for extracting useful data from clinical data [5]. These techniques give an opportunity to transform unstructured medical text into structured data that can be searched and analyzed.[1], [3]

The proposed system, Medical Report Interpretation Using Natural Language Processing (NLP), is designed for interpret medical reports for different diseases and medical conditions rather than focusing on one type of disease. The system accepts medical reports in formats such as PDF, TXT, and DOCX and performs perform preprocessing and clinical data extraction. Important entities such as symptoms, disease, medications, investigations, procedures, and clinical findings are recognized from report [1], [3], [9]. Contextual processing, include negation detection and terminology normalization, is used to improve interpretation of extracted information.

In addition to report interpretation the proposed system gives a doctor-oriented research module. Based on data analyze from patient's medical report, an authorized doctor/researcher can search and review related symptoms and associated clinical data. This will help to organize patient-related information and support clinical information retrieval. clinical decision-support system can assist healthcare professionals in accessing and using relevant clinical data while requiring appropriate consideration of safety, privacy, and human oversight [2], [10] The system also generates a concise summary and patient-friendly explanation of the extracted data or information [10], [11]. The objective is not replacing doctors or give a autonomous diagnosis, but support

the organization, interpretation, and retrieval of information contained in medical reports. By combining clinical NLP, information extraction, semantic representation, and doctor-oriented search, the proposed system aims to make medical report analysis more structured, accessible, and efficient.

II. RELATED WORK

Several studies have explored the use of Natural Language Processing (NLP), Machine Learning, and Artificial Intelligence for processing and interpreting medical information. These studies give the foundation for the extracting information from semi-structured medical reports and support to healthcare professionals.[1], [2]

Wu et al. represent a methodical review of deep learning approaches in clinical Natural Language Processing and analyzed a huge number of studies involving clinical text processing. Their work highlighted the use of deep learning for task like clinical data extraction, named entity recognition, and text classification [1]. The study demonstrated the growing importance of deep learning technique for convert unstructured clinical text into useful data

Hahn et al. developed MEDSYNDIKATE, a natural language system extracting clinical information from findings in medical reports. The system demonstrated how NLP techniques will be applied to medical report or documents to identify and structure related clinical data. This work closely relevant to the proposed system because it focuses on extracting information from clinical reports.

Sreenivas et al. investigated clinical Named entity recognition using fine-tuned BERT together with dictionary-based retrieval techniques [9]. Their work shows that transformer-based language models will improve the identification of clinical entities from clinical text. Like approaches are relevant to the proposed system for extracting disease, symptoms, investigation, and other medical concepts

Ladas et al. studied programming technique for improve a readability of rule-based information extraction pipelines for semi-structured and unstructured clinical texts [8]. Their work emphasizes the usefulness of understandable rule-based approaches when extracting data from clinical reports. This supports the use of rule-based processing together

with machine learning technique in the proposed system.

Sutton et al. give an overview of Clinical Decision Support System (CDSS) and discussed their benefits, risks, and strategies for successful implementation [2]. Their work indicates the medical information system should support healthcare professionals rather than replace human decision-making. This concept is considered in the proposed doctor-oriented research module.

Singhal et al. introduced Med-PaLM2, a large language module designed for clinical question answering and evaluation. Their research introduces a potential of advanced language models for processing medical knowledge while also highlighting the importance of safety, evaluation, and human oversight in healthcare applications [13]

Alowais et al. discussed a role of Artificial Intelligence in medical practice and highlighted application of AI in healthcare, along with challenges relevant to implementation, reliability, ethics, and clinical use. This consideration are important for designing a medical report interpretation system.

Fontes et al. reviewed example-based Explainable artificial Intelligence (XAI) support a medical imaging [11]. Although their study focusses in medical image rather than medical text, It emphasizes the importance of providing understandable explanations for AI-supported healthcare systems. This principles is related to the patient-friendly explanation component of the proposed system.

Poli et al. investigated automated coding of electronic health records for rare disease registries [12]. Their work demonstrates a potential of automated processing of medical records for organizing and retrieving clinical information and also highlights challenges associated with medical data quality and standardization.

The reviewed studies demonstrate significant progress in medical NLP, information extraction, transformer-based modes, medical decision support, and explainable AI [1], [2], [8], [9], [12]. However, many existing approaches focuses a specific task, dataset, disease area, or medical applications, The proposed system aims to give a boarder medical report interpretation framework that can process report relevant to different diseases and medical conditions. It combines medical information extraction, contextual

analysis, semantic representation, patient-friendly explanation report summarization, and doctor-oriented research-module for retrieving related symptoms and patient information.

III. PROPOSED SYSTEM

The proposed system is a NLP-based framework to interpreting clinical reports related to multiple diseases and medical conditions. It processes unstructured clinical reports and converted into structured and understandable information for patients and healthcare professionals. The system approaches PDF, TXT, and DOCX formats [1], [3]

A. Medical Report Preprocessing

The upload report is first processed to extract its textual content. Documents in PDF, TXT, and DOCX format are handled using appropriate document-processing techniques. The extracted text data is cleaned and normalized to remove unwanted characters, formatting information, and irrelevant data. NLP preprocessing technique such as sentence segmentation and tokenization are then applied for prepare a text for further analysis [1]

B. Medical Information Extraction

After preprocessing, the system recognizes important medical information from medical report. The extraction process focuses on clinical entities such as diseases, symptoms, medications, laboratory data, clinical findings, and treatment-related information [1], [3], [9]. Medical terminology and related relationships between clinical entities are identified to create structured representation of information contained in the report of extracted clinical information.

C. Semantic Analysis

The system performs semantic analysis to understand a contextual meaning of a extracted medical information. Medical-domain language model such as BioBERT, and ClinicalBERT will be integrated using Hugging Face transformers framework. These models help in understanding clinical terminology and contextual relationships within clinical text. Negation and contextual information are also considered to reduce incorrect interpretation of clinical statements [1].

D. Patient-oriented Interpretation

The extracted clinical information is converted into concise and understandable from patient. The system generates a summary of important recognizing and gives a simple explanation of tough medical terminology. Patient can also see their medical timeline, access general medication reference data, and prepare questions based on the available clinical information for discussion with healthcare professionals [11].

E. Doctor-Oriented Clinical Support

This system gives a additional functionality for healthcare professionals. Based on the available patient information, the system can approach clinical-trial-related activities such as trial matching and trial eligibility analysis [2], [10]. Trial Matching helps to identify potentially related clinical trials, while eligibility analysis compares available patient data with given trial eligibility requirements. The system can also give an clinical-trial site and contact details for further research.

F. Database and security

The system stores user information, uploaded medical documents, and processed results in database. Authentication and access-control mechanisms are used for restrict unauthorized to sensitive medical data. The system is designed for maintain the confidentiality and integrity of stored information while allowing authorized users to access their reports and generated interpretation.

IV. METHODOLOGY

The methodology of a proposed system consists of the sequence of steps for processing and interpreting clinical reports using Natural Language Processing (NLP). The major stages include medical report collection, text extraction, processing, medical information extraction, semantic analysis, interpretation, result generation [1], [3].

A. Medical Report Collection and Text extraction

Medical report is collected in PDF, TXT, and DOCX formats. The text is extracted from the uploaded reports using suitable document-processing libraries. The extracted information is then provided to the NLP processing module.

B. Text Preprocessing

The extracted clinical text is cleaned and normalized to remove unnecessary characters and irrelevant information. Sentence segmentation and tokenization are performed to divide the text into meaningful units. This processed text is used to further clinical data analysis [1].

C. Medical Entity Extraction

The system recognizes important clinical entities like disease, symptoms, medications, laboratory tests, and medical findings [1], [3], [9]. NLP technique and medical-domain vocabularies are used for recognize related information from the processed reports.

D. Semantic and Contextual Analysis

The extracted information are analyzed to understand its contextual meaning. Medical language models like BioBERT or ClinicalBERT can be used for improve understanding of clinical terminology and relationships between clinical entities [5]. Negation and contextual data are considered to reduce incorrect interpretations [1].

E. Report Interpretation and Output Generation

The analyzed data is converted into the structured representation and used to generate a concise clinical report summary. Complex clinical terminology is presented in simpler and understandable language to support patient understanding. The processed data can also be used to doctor-oriented clinical trial matching and eligibility analysis [11].

F. Storage and security

The uploaded documents, user information, and generated results are stored in a databases. Authentication and access-control mechanisms are applied to protect sensitive clinical information and prevent unauthorized in access [2], [10]

V. SYSTEM ARCHITECTURE

The proposed system approaches both patient and doctors in medical report interpretation and clinical trial matching. Medical report are uploaded in PDF, DOCX, and text format and processed using text extraction and NLP to identify medical information such as disease, symptoms, medications, and laboratory conditions. Context and negation analysis

create a structured patient profile, which is semantically matched with clinical trials retrieved from Clinical Trials.gov using Chro-maDB. A rule-based engine evaluates eligibility and generates eligible, blocked, or uncertain result with explanations. The results is presented through a common dashboard for patients and doctors, while search history is stored in SQLite.

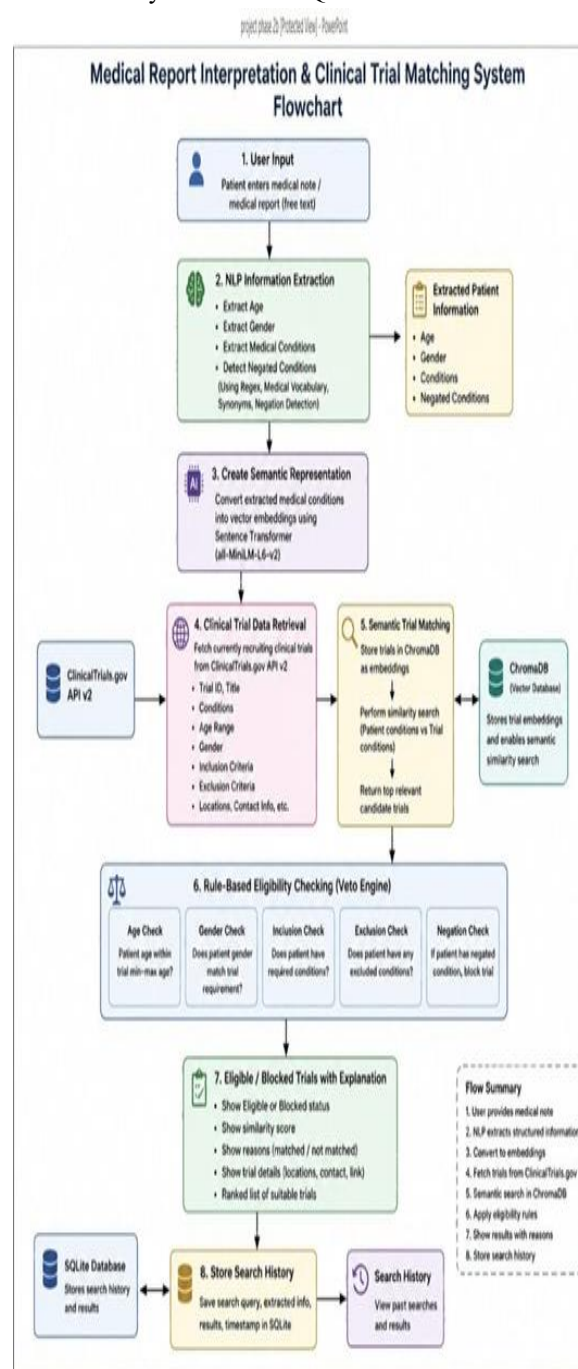


Fig. 1. Doctor-oriented workflow of the Proposed system

The patient uploads a medical report in PDF, DOCX, or TXT format, which is processed by the system for further analysis. The system extracts and preprocesses the text by removing unnecessary information and handling medical terms and abbreviations. NLP techniques are used to identify important information such as diseases, symptoms, medications, test results, and clinical findings. The extracted medical information is interpreted and converted into simple, patient-friendly language for better understanding. The system generates a structured summary containing important findings, medical history or timeline, and information that can help the patient discuss their condition with a doctor.

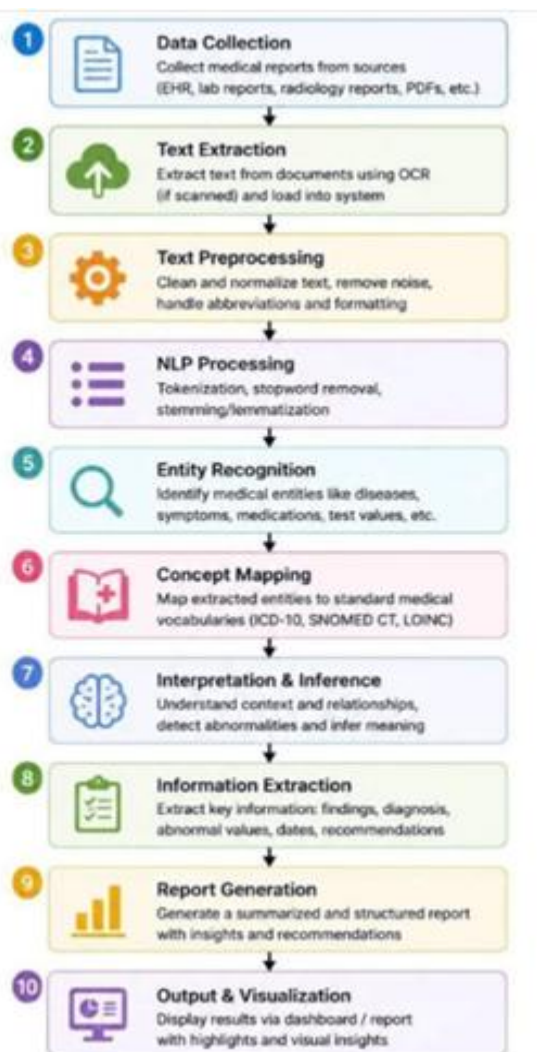


Fig. 2. Patient-Oriented workflow of the System

The doctor can review the patient's medical report because the report undergoes text extraction,

preprocessing, and NLP-based information extraction. The information extracted from the report is converted into semantic representations to identify similar cases. The system then accesses relevant clinical trials and performs semantic similarity modeling and rule-based eligibility scoring using the patient's age, gender, inclusion criteria, exclusion criteria, and negations. The result presents relevant details about the trials, eligibility, reasons for match, location, and contact information for the doctor's reference and research.

VI. RESULT AND DISCUSSION

The system Medical Report Interpretation Using Natural Language Processing (NLP) system was developed to process semi-structured medical reports and give an structured and understandable data. The system accepts clinical report in PDF, TXT, and DOCX, formats and performs text extraction, pre-processing, medical information extraction, semantic analysis, and report interpretation. The obtained results demonstrate a applicability of NLP technique for organizing and interpreting data contained in medical reports [1], [3].

A. Medical Report Processing

The system processes a uploaded reports and extracts the content for further analysis. The preprocessing stage removes unnecessary character, normalizes the extracted content, and perform tokenization and other NLP operation, This improves a quality and consistency of the content give the subsequent analysis modules [1], [8].

B. Medical Information Extraction

The system recognizes important medical information, include diseases, symptoms, medications, laboratory values, and medical findings from medical text. The extracted data is organized into structured report, allowing related medical details to be identified and analyzed more efficiently [1], [3], [9].

C. Report Interpretation and summarization

The processed system provides concise summary of medical report and give a simple explanation of complex medical terms. Semantic- analysis helps the system understand a context of extracted information. This allows medical information to be presented in a more understandable form fora patients and healthcare professionals [5].

D. Patient and doctor-Oriented support

The system provides different features for patients and healthcare professionals. For patients, it give a report summaries, simple clinical explanations, medical timeline data, general medication references. For healthcare professionals, it support clinical analysis, and clinical trial site and contact information [2].

E. Clinical Implementation considerations

Several important factors should be considered when implementing the proposed system:

- 1 Data Quality and Standardization: Medical Report may have different formats, writing styles, and levels of completeness. Proper preprocessing is required to get a reliable information from different reports [1].
- 2 NLP Model Performance: The quality of extracted data depends on a NLP models and medical vocabulary used by the system. Appropriate models should be selected and evaluated using suitable medical datasets [5].
- 3 Context and Ambiguity: Medical terms can have different meanings depend on their context. The system will consider contextual data and negation to reduce incorrect interpretation [1], [3].
- 4 System Integration: The system should give an easy-to-use interface and integrate the patient and doctor modules effectively so that the generated information can be accessed easily [2].
- 5 Privacy and security: Medical report contain sensitive information. Secure storage, authentication, and access-control mechanisms are required for protect patient data and prevent unauthorized access [2], [10].

F. Comparison with Existing systems

Medical processed system combines medical report processing, NLP-based data extraction, semantic analysis, summarization, and clinical support features within a single framework, Compared with system that focus only a medical entity extraction or report summarization, this support both patient-oriented and doctor-oriented functionality[1], [3], [8], [9].These features give a broader approach to medical report interpretation and clinical research support.



Fig. 3. Main Interface for Patient and Doctor

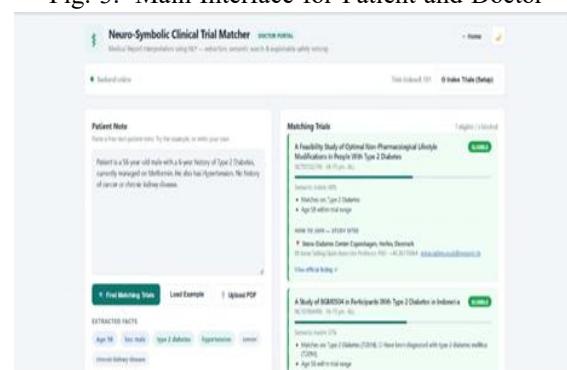


Fig. 4. Doctor Portal for Clinical Trial Matching

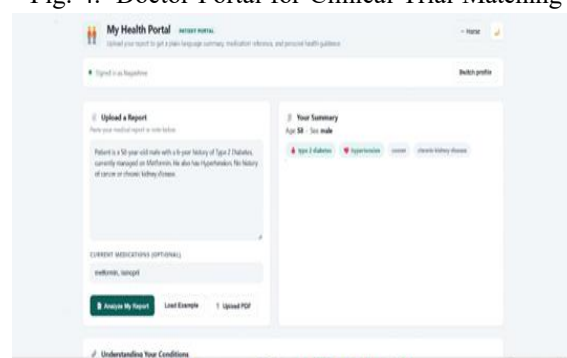


Fig. 5. Patient Portal for Medical Report Interpretation

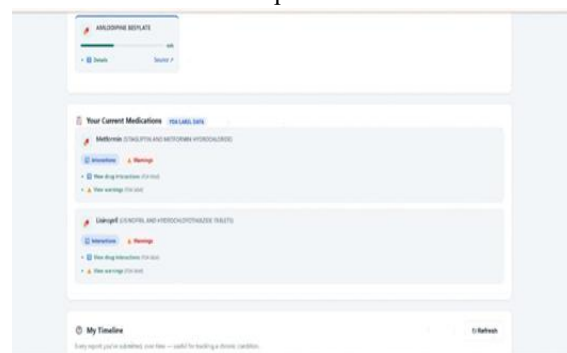


Fig. 6. Medication and Patient timeline.

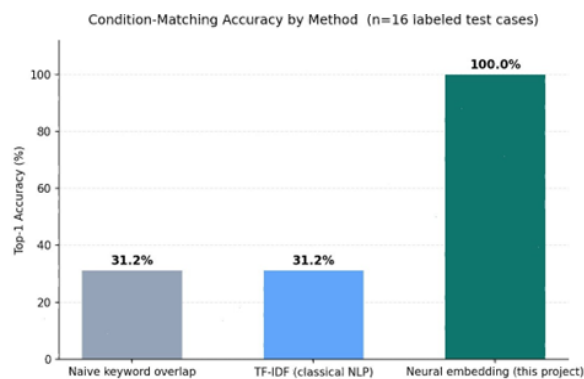


Fig. 7. Performance Results of the Proposed System

VII. CONCLUSION

The Medical Report Interpretation Using Natural Language Processing (NLP) system provides a unified framework for processing, interpreting, and organizing information from unstructured medical reports. The system extracts important medical information such as diseases, symptoms, medications, laboratory values, clinical findings, and other relevant information and presents it in a structured and understandable form [1], [3]. The integration of text preprocessing, medical information extraction, semantic analysis, and report summarization helps in efficiently handling complex clinical text and improves the accessibility of medical information.

The proposed system provides ****patient-oriented features**** such as simplified medical explanations, medical timeline information, report summaries, and general medication references. These features help patients understand the important information contained in their medical reports in a more accessible manner. At the same time, the system provides ****doctor-oriented features**** such as clinical trial matching, trial eligibility analysis, semantic similarity-based matching, and clinical trial site and contact information using available patient medical information [2], [10]. This combination makes the system useful not only for medical report interpretation but also for supporting clinical research activities.

Overall, the proposed system demonstrates the potential of NLP techniques in improving the organization, interpretation, and accessibility of information contained in medical reports [1], [2]. By converting unstructured clinical text into structured

information and providing different outputs for patients and healthcare professionals, the system can reduce the difficulty associated with understanding complex medical documentation. The clinical trial module further extends the system by connecting relevant patient information with available clinical trial data and providing eligibility-related information in an organized manner.

However, the effectiveness of the system depends on the quality and completeness of the input medical reports, the accuracy of information extraction, and the performance of the underlying NLP and semantic models. Future improvements can focus on advanced medical-domain language models, larger and more diverse clinical datasets, improved contextual and negation detection, multilingual medical report processing, better handling of abbreviations and ambiguous medical terms, and stronger privacy and security mechanisms. Further development can also improve the explain ability of extracted information and clinical trial matching results.

The proposed system is intended to function as a supportive information and research tool for patients, doctors, and healthcare-related users. It is not intended to replace professional medical diagnosis, treatment decisions, or clinical judgment. With further validation, improved datasets, and advanced NLP techniques, the system can provide a more effective approach for interpreting medical reports and supporting access to relevant clinical information.

REFERENCES

- [1] S. Wu, K. Roberts, S. Datta, J. Du, Z. Ji, D. Si, S. Wei, and Y. Zhang, "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.
- [2] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. Kroeker, "An overview of clinical decision support systems: Benefits, risks, and strategies for success," *npj Digital Medicine*, vol. 3, p. 17, 2020.
- [3] A. Mulyar, O. Uzuner, and B. McInnes, "MT-clinical BERT: Scaling clinical information extraction with multitask learning," *Journal of the American Medical Informatics Association*, vol. 28, no. 10, pp. 2108–2115, 2021.

- [4] 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.
- [5] E. Alsentzer, J. Murphy, W. Boag, W.-H. Weng, D. Jin, T. Naumann, and M. McDermott, "Publicly available clinical BERT embeddings," in *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pp. 72–78, 2019.
- [6] K. Huang, J. Altosaar, and R. Ranganath, "ClinicalBERT: Modeling clinical notes and predicting hospital readmission," *arXiv preprint arXiv:1904.05342*, 2019.
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [8] D. Ladas et al., "Programming techniques for improving rule readability for rule-based information extraction natural language processing pipelines of unstructured and semi-structured medical texts," *Health Informatics Journal*, 2023.
- [9] S. Sreenivas et al., "Enhancing clinical named entity recognition via fine-tuned BERT and dictionary-infused retrieval-augmented generation," *Electronics*, vol. 14, no. 18, p. 3676, 2025.
- [10] H. Eguia, C. L. Sa'nchez-Bocanegra, F. Vinciarelli, F. Alvarez-Lopez, and F. Saig'i-Rubio', "Clinical decision support and natural language processing in medicine: Systematic literature review," *Journal of Medical Internet Research*, vol. 26, e55315, 2024.
- [11] R. Fontes et al., "Application of example-based explainable artificial intelligence (XAI) for analysis and interpretation of medical imaging: A systematic review," *IEEE Access*, vol. 12, 2024.
- [12] A. Paoli et al., "Opportunities and challenges in automated coding of electronic health records: A pilot study for rare disease registries," *Frontiers in Digital Health*, vol. 8, p. 1807607, 2026.
- [13] K. Singhal et al., "Toward expert-level medical question answering with large language models," *Nature Medicine*, 2025.
- [14] M. Neumann, D. King, I. Beltagy, and W. Ammar, "ScispaCy: Fast and robust models for biomedical natural language processing," in *Proceedings of the 18th BioNLP Workshop and Shared Task*, pp. 319–327, 2019.
- [15] S. Henry, K. Buchan, M. Filannino, A. Stubbs, and O. Uzuner, "2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records," *Journal of the American Medical Informatics Association*, vol. 27, no. 1, pp. 3–12, 2020.
- [16] Q. Wei et al., "A study of deep learning approaches for medication and adverse drug event extraction from clinical text," *Journal of the American Medical Informatics Association*, vol. 27, no. 1, pp. 13–21, 2020.
- [17] M. Ju, N. Nguyen, M. Miwa, and S. Ananiadou, "An ensemble of neural models for nested adverse drug events and medication extraction with subwords," *Journal of the American Medical Informatics Association*, vol. 27, no. 1, pp. 22–30, 2020.
- [18] "Deep contextual multi-task feature fusion for enhanced concept, negation and speculation detection from clinical notes," *Informatics in Medicine Unlocked*, vol. 34, p. 101109, 2022.