International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 2.102

A Systematic Analysis of Accuracy and Trust in Clinical Decision Making with NLP

Chandra Prakash[1], Musa Shaikh[2], Pavan Mutha[3], Praveen Kumar Parimi[4]

[1]School of Computer Information Sciences, University of the Cumberlands, Williamsburg, KY, USA
Email*: cprakash[at]outlook.com*

[2]School of Computer Information Sciences, University of the Cumberlands, Williamsburg, KY, USA
Email: *musa.a.shaikh[at]gmail.com*

[3]School of Computer Information Sciences, University of the Cumberlands, Williamsburg, KY, USA
Email: *pavan.mutha[at]hotmail.com*

[4]School of Computer Information Sciences, University of the Cumberlands, Williamsburg, KY, USA
Email: *pkumarparimi[at]gmail.com*

Abstract: ***This systematic review examines the role of Natural Language Processing (NLP) in enhancing the efficiency and accuracy of clinical documentation. The study investigates key NLP techniques utilized in the automated generation and summarization of clinical notes. It analyzed 152 peer-reviewed articles from the past 24 months, as recent studies are crucial given the rapidly evolving landscape of machine learning, AI, and their applications in the medical field. The findings show that though NLP significantly enhances clinical documentation, increasing efficiency and accuracy through automation, there remains a concern about Trust and Transparency. This systematic review proposes novel methods to address the trust and transparency issues in clinical decision-making with NLP. This research is vital for developing NLP applications that aid clinical decision-making, reduce clinician burnout, improve patient outcomes, and support personalized healthcare. The review concludes with suggestions for future research to create adaptable and context-aware NLP models.***

Keywords: Natural Language Processing (NLP), Clinical Documentation, Decision Support, Healthcare

1. Introduction

The healthcare industry has experienced a profound transformation due to the integration of Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies, particularly in clinical documentation and decision support systems. Natural Language Processing (NLP), a branch of artificial intelligence, facilitates the extraction of valuable insights from unstructured clinical data, including Electronic Health Records (EHRs), clinical notes, and diagnostic reports (1–3). Through automated processing of complex medical narratives, NLP enhances the accuracy of clinical decision-making and streamlines healthcare workflows, providing interpretable insights that clinicians can understand and trust, thereby reducing the cognitive burden on healthcare providers (4). NLP can be leveraged for various tasks, such as disease prediction, clinical decision support, and patient risk stratification (1,4). However, the clinical domain presents unique challenges for NLP, including the complexity of medical terminology, the need to protect patient privacy, and the scarcity of labeled data (5–8).

Recent advancements in transformer-based models like DistilBERT, BioClinicalBERT, BioBERT, and Bio+ Discharge Summary BERT have significantly improved the performance of NLP applications in various healthcare tasks, such as clinical text classification, sentiment analysis, and named entity recognition (NER) (5,7,9–11). These models have demonstrated superior capabilities in recognizing medical entities, understanding the context within clinical texts, and supporting predictive analytics for patient outcomes. For example, NLP systems are increasingly being utilized to automate disease classification, identify adverse drug reactions, and assist in evaluating clinical trial eligibility, enhancing the delivery of timely and personalized patient care (12).

Despite the considerable progress, challenges persist, including the need for large annotated datasets, ethical concerns regarding patient data privacy, and the inherent complexity of medical language. This research explores the current landscape of NLP applications in healthcare, consolidating existing evidence on how these technologies reduce clinician workload, improve data accuracy, and streamline healthcare delivery. The review also focuses on their role in enhancing clinical decision support and improving the quality of clinical documentation. It highlights emerging trends, key techniques, and the future potential of NLP in transforming healthcare delivery. By systematically reviewing the literature, this research can uncover critical insights into the current state of NLP in healthcare, highlight areas for improvement, and propose future directions for advancing clinical documentation and decision support.

The primary objective of this research is to systematically investigate how Natural Language Processing (NLP) is transforming clinical documentation in healthcare. Specifically, it aims to assess the impact of NLP and evaluate the extent to which NLP has improved the efficiency and accuracy of clinical documentation by reducing manual data entry, enhancing data completeness, and minimizing errors in Electronic Health Records (EHRs). The research also focused on identifying and critically examining the key NLP techniques, such as deep learning models and transformer architectures, used in automating the generation and summarization of clinical notes, highlighting their effectiveness and limitations.

2. Literature Review

Historically, conversations have been a primary method for gathering clinical information, gradually evolving into systematic record-keeping. Clinical systems such as electronic medical records, labs, and other documentation require manual entry and review of the data, bringing inefficient and inaccurate processes and causing faulty decision-making, leading to delays in patient care. Over time, efforts have been made to bring in automated decision-making support systems to improve clinical decisions. Digital systems brought various techniques to organize and synthesize the available information, including the use of natural language processing (13). It is imperative to gain insights into patients' physical, psychological, and health-related quality of life, which are critical factors to analyze in clinical decision-making. NLP is one of the disciplines of machine learning that has brought revolutionary changes in how clinical data is analyzed, and decisions are made.

2.1 NLP

Natural Language Processing (NLP), a subset of Artificial Intelligence (AI), holds significant promise for advancing healthcare through its integration into clinical decision support systems (1). By analyzing large volumes of unstructured clinical text, these systems use NLP to provide healthcare professionals with personalized recommendations, real-time alerts, and evidence-based insights, allowing them to enhance patient care (14). Although NLP has achieved substantial progress in tasks like automatic translation, virtual assistants, and healthcare applications, the complexity and nuances of human language continue to pose challenges for AI. However, clinical NLP (C-NLP) has become essential for improving doctor-patient communication and streamlining healthcare documentation processes (15).

2.2 Clinical Documentation

Clinical documentation is vital to modern healthcare, offering numerous advantages for patient care and the broader medical field. It acts as a key cognitive tool for clinicians, aiding them in organizing a patient's case, pinpointing critical details, and supporting reasoning and decision-making (5,16). Also, it fosters collaboration among healthcare teams by promoting a shared understanding of cases and improving care coordination. Beyond its role in patient care, the extensive data captured in clinical documentation, such as medical issues, treatments, outcomes, and costs, are valuable for various other purposes, including research.

2.3 NLP in Clinical Documentation

Traditionally, clinicians gathered information about patient conditions through conversations, a practice that has evolved into systematic record-keeping with the introduction of clinical documentation tools. Modern Electronic Health Records (EHRs) manage vast clinical data. However, their implementation has created administrative burdens, primarily due to the extensive use of free-text formats that do not conform to structured information or coding frameworks like ICD-10 (International Classification of Diseases, tenth revision) and UMLS (Unified Medical Language System). This lack of standardization in EHRs complicates data retrieval, posing challenges for clinicians and contributing to burnout (13). Integrating natural language processing (NLP) into clinical decision support systems enhances healthcare by analyzing unstructured clinical text, such as medical records and physician notes. NLP extracts provide key insights, thus enabling context-aware recommendations, alerts, and evidence-based guidelines at the point of care (17). It helps healthcare providers make informed decisions, personalize patient care, and comprehensively understand medical histories and treatment options (18).

Natural Language Processing (NLP) algorithms offer researchers tools to develop models that effectively understand, interpret, and extract valuable insights from textual data. By leveraging the power of Transformers and the bidirectional semantic architecture of BERT, the algorithms have achieved remarkable results in various NLP tasks (6). BERT's ability to capture context from both directions is invaluable for advancing text analysis and enhancing our understanding of language (11). Variants of BERT, such as DistilBERT, BioClinicalBERT, BioBERT, and Bio+Discharge Summary, build upon the foundational architecture of BERT to enhance its performance for specific tasks or improve efficiency (10). These models refine pretraining strategies, adapt to domain-specific text, or reduce computational requirements while retaining BERT's contextual understanding capabilities.

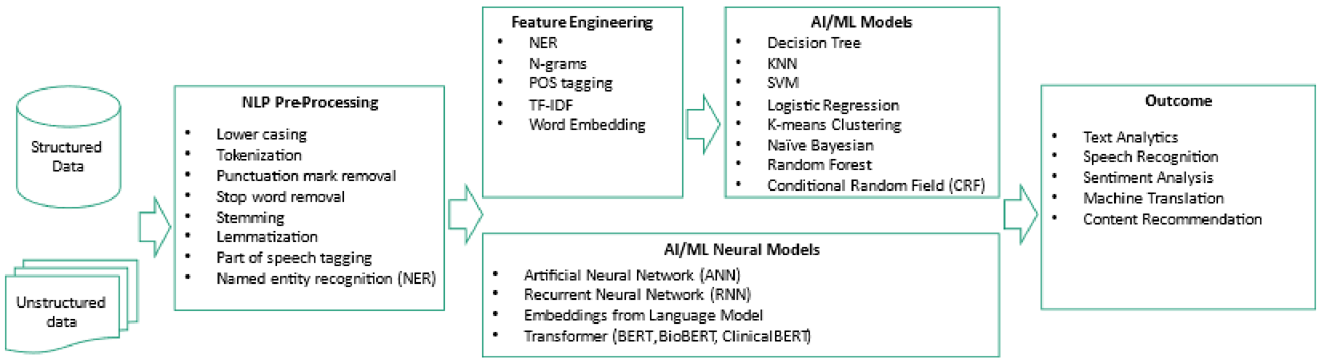

Figure 1: NLP flow with components to process structured and unstructured clinical data

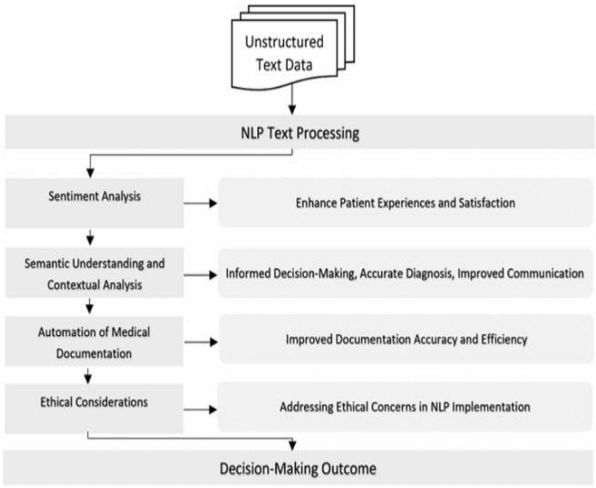

Figure 2: Flowchart of NLP integration in healthcare Decision-Making

2.4 Enhancement of Clinical Decision Support Systems (CDSS) Through NLP

Clinical Decision Support Systems (CDSS) are computer-based tools that assist healthcare professionals, patients, and others by providing relevant, personalized information at the right time. These systems support a range of healthcare activities, including diagnosing diseases, predicting treatment outcomes, personalizing treatments, estimating patient prognosis, and prioritizing care based on risk (19).

NLP's integration into CDSS enhances healthcare delivery by extracting key information from unstructured clinical texts such as physician notes, radiology reports, and pathology reports. These systems provide clinicians with context-aware recommendations, alerts, and evidence-based guidelines. NLP algorithms analyze clinical narratives to identify symptoms, diagnoses, and medications, ensuring that crucial data is organized and accessible for accurate decision-making (3,10).

2.5 Context-Specific Recommendations and Alerts

NLP-powered CDSS can issue real-time alerts, identifying contraindications and potential risks, such as drug interactions or allergies. For instance, if a prescribed medication conflicts with a patient's documented allergies, the system generates an alert to prevent harm. Additionally, these systems dynamically deliver evidence-based guidelines, ensuring clinicians access to the latest research and clinical best practices, thereby improving patient outcomes (3,10).

2.6 Seamless Data Integration and Holistic Patient Views

NLP bridges the gap between structured and unstructured data, synthesizing diverse data sources such as patient demographics, lab results, and clinical narratives. This holistic integration enables providers to anticipate complications, identify risks, and make well-informed decisions based on the most comprehensive and current information available, ultimately improving care efficiency and quality (20).

NLP is transforming healthcare by alleviating the burdens of clinical documentation and enhancing clinical decision support. Its ability to extract, structure, and analyze unstructured clinical data reduces administrative workloads and improves diagnostic accuracy, patient safety, and treatment personalization. As NLP technologies advance, their potential to revolutionize healthcare documentation and decision-making will continue to expand, driving improvements in clinical outcomes and healthcare delivery (18,21).

3. Research Questions

This systematic review aims to explore the current applications of NLP in healthcare, focusing on its role in clinical documentation and decision-making. The review explored the current usage of NLP and its contribution to the efficiency and accuracy of clinical documentation. The review further explored the key clinical documentation and decision-making methods. The researchers have identified the following research questions based on the literature review.

RQ1: How has NLP improved the efficiency and accuracy of clinical documentation in healthcare settings?

RQ2: What are the transparency, trust challenges, and concerns of using NLP in clinical decision-making?

RQ3: What are the novel approaches in NLP to improve transparency and trust in clinical decision-making?

4. Methodology

This systematic review has utilized the methodology presented by Tranfield et al. (22) to ensure transparency and objectivity while reporting recent progress in the literature review on the NLP's usage in clinical decision-making. This systematic review aims first to identify how NLP has improved the efficiency and accuracy of clinical decision-making within healthcare and advanced NLP techniques utilized for clinical notes summarization and decision-making.

This study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA). PRISMA model allows for a comprehensive and transparent approach to systematic reviews and inclusion of relevant studies. Table 1 highlights the inclusion and exclusion criteria for the paper selection. Figure 3 describes the PRISMA approach taken for system literature review selection.

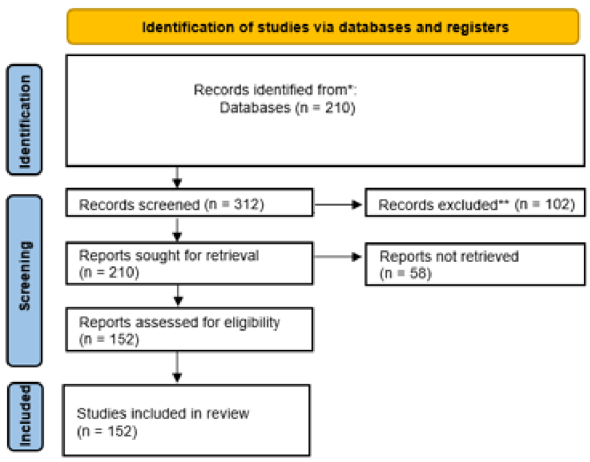

Figure 3: PRISMA Flow Chart. Adoption from Page et al. (23)

4.1 Search Strategy

This paper followed a systematic search strategy to answer the identified research questions and determine the relevant studies. The search terms utilized included "NLP in Clinical Decision Support," "Clinical Decision Support," "NLP for clinical documentation," and a combination of words with and/or criteria to search for better results. The search criteria were applied to IEEE Xplore, Google Scholar, and PubMed to search for relevant articles. The authors selected studies published in the last 24 months to ensure the study covers the advancement in fast-moving machine learning and AI space.

4.2 Data Extraction and Analysis

Data extraction utilized a thoughtful analysis of individual papers to identify the relevancy of research questions. The study considered the research subject as the key criteria and extracted the studies based on relevancy with abstract, title, authors, and research technique. Key findings and methodological details were also collected. The retrieved data was classified into themes or subjects related to the study questions throughout the data analysis phase.

4.3 Validity

The study performed a literature review to answer the identified research questions. The validity and depth of the review paper are equally important, and to ensure validity, Shaheen et al. (24) have recommended using multiple databases, screening articles, and corroborating the findings with other well-established sources.

4.4 Inclusion and Exclusion Criteria

Table 1 defines the inclusion and exclusion criteria utilized during the study for research and filtering out the selected papers. Inclusion and exclusion criteria are extremely helpful in narrowing down the relevancy of suitable papers for the study.

Table 1: Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Papers published in the last 24 months	Papers unrelated to the research question.
Peer-reviewed papers published in journals or presented in conference papers.	Paper written in languages other than English.
Papers using English as the language for NLP analysis.	Paper using languages other than English for NLP analysis.
Papers using various NLP models for analysis purposes.	Papers using non-NLP models for analysis.
Papers focused on clinical documentation using one of the NLP techniques for decision support.	Papers not directly related to healthcare or decision support.

5. Results and Discussions

The earlier literature review and methodology sections provided empirical evidence of NLP usage and its wider adoption in clinical decision-making. The results and discussion section provides an in-depth exploration of the findings related to the research questions posed in the study.

The study is organized into four distinct parts; the first section provides the various keywords associated with the research papers. The second section offers discussions to answer the research questions. The third section discusses the challenges and limitations of the research paper, and the fourth section offers the gaps and direction for future researchers.

5.1 Overview of Included Studies

Table 2 summarizes the keywords from the identified literature and their frequency. The table provides a strong indicator of how NLP is being utilized in clinical documentation for various purposes.

Table 2: Summary of Keywords and Frequency in the documentations

Keyword	Frequency
Natural language processing	78
Medical services	53
NLP	40
Transformers	38
Deep learning	32
Training	30
Biological system modeling	28
Computational modeling	26
Data models	23
Predictive models	22

5.2 Discussions

NLP is revolutionizing clinical documentation by automating labor-intensive processes, ensuring data accuracy, enhancing interoperability, and supporting data-driven decision-making. NLP integration into healthcare workflows improves documentation quality and drives efficiencies that enhance patient care, safety, and outcomes. The results and discussion section provides an in-depth exploration of the findings related to the research questions posed in the study.

- **RQ1:** How has NLP improved the efficiency and accuracy of clinical documentation in healthcare settings?

In the last decade, Natural Language Processing (NLP) has significantly improved clinical documentation in healthcare, automating transcription and enhancing data integrity. Techniques like pseudo-labeling have reduced the time and resources needed for manual assessment (Xu et al., 2023). This automation decreases documentation time and helps alleviate clinician burnout, allowing healthcare providers to prioritize patient care over administrative tasks (2,3,25). NLP systems are highly effective in extracting clinical information from unstructured data sources like physician notes, ensuring that key details such as diagnoses, medications, and demographics are accurately captured in electronic health records (EHRs). This enhancement improves data completeness and accuracy, which is vital for reliable and actionable health records. By reducing human error, NLP supports clinical decision-making and aids research efforts, further demonstrating its value in healthcare (3,16). Comprehensive clinical documentation aids clinical decision support systems, population health management, and research efforts, further amplifying NLP's value.

The research by Berloco et al. (2023) indicates that NLP

Paper ID: SR25106182822 DOI: https://dx.doi.org/10.21275/SR25106182822

algorithms can map clinical language to standardized coding systems like ICD-10 and SNOMED CT, ensuring consistency and interoperability across healthcare institutions. This standardization is essential for efficient data exchange and compliance with regulatory requirements, benefiting value-based care models by improving analytics for performance indicators. NLP models also detect errors and inconsistencies in clinical documentation, which enhances the integrity of patient records (16,26). By flagging potential issues in real time, NLP helps reduce medical errors and supports clinical decisions, particularly in complex cases where accuracy is directly linked to patient outcomes. Additionally, NLP synthesizes large volumes of clinical text into concise summaries that highlight key findings, facilitating quicker information retrieval for clinicians and enhancing communication and continuity of care (27). It also contributes to predictive analytics by improving the datasets available for machine learning models, which can identify trends and support population health initiatives (16,25). As NLP technologies continue to develop, their impact on clinical documentation and healthcare processes is likely to grow, offering new opportunities for innovation.

Table 3: Outlines a summary of the research and respective algorithms and accuracy

Authors	Accuracy	Algorithm	Key Findings	Limitations
J. Wu et al., 2023	96.04%	Word2Vec	Semantic relationship, contextual representation	Limited to basic ML models and utilized the EDAIC dataset only.
Xu et al., 2023	89.11%	DeBERTa	Attention-driven, best for sequential data	No limitations mentioned
Arqam et al. 2023	92.65%	CharacterBERT	Combined BERT and CHARACTER-BERT embeddings achieved an F1 score of 92.65% in identifying cardiovascular disease risk factors from EHRs.	Future research will explore deep learning and ensemble learning for improved risk assessment.
Sudarshan et al., 2023	95%	BioBERT	Pretrained on clinical text, superior due to understanding terminology and relationship	BioBERT is resource intensive.
Chang et al., 2024	97%	LLMs	LLMs are performing better than the specialized domain-trained model	Slower than the pre-trained model
Belkadi et al., 2023	97%	TransformerCRF	TransformerCRF is slightly lower than the specialized models by .03%. However, it lacks specialized training and has fewer parameters, making it much more efficient.	Ethical concerns and bias in decision making.
Kalavathi et al.	92%	PubMedBERT, BioBERT, DeBERTa	PubMedBERT and BioBERT were effective for extractive summarization of clinical text, outperforming DeBERTa.	Requires further fine-tuning on specific healthcare domains. Exploration of other summarization approaches is needed. Hybrid and graph-based methods are recommended.
A.A. Funkner et al.	91%	RuBERT, SBERT (Sentence-BERT)	RuBERT achieved the highest accuracy for symptom extraction from Russian medical notes. SBERT performed best for precision and F1-score.	No limitations mentioned

After a thorough literature review, it's evident that NLP employs various techniques for automated clinical note generation and summarization. A key technique is named entity recognition (NER), which identifies and extracts entities such as medical conditions, medications, and procedures (28–30). NER streamlines information extraction from unstructured clinical notes, highlighting critical data points like symptoms and treatments. Another method is Term Frequency-Inverse Document Frequency (TF-IDF), which extracts key terms based on their frequency and significance (16). Authors Pal et al. (31) and Sreenivasgoud et al. (16) emphasized relationship extraction, identifying connections between medical entities. Techniques like word embeddings capture semantic relationships, enhancing information accuracy (16). Researchers like Ananthajothi et al. (32) have also noted temporal analysis for its role in analyzing disease progression and treatment timelines. Summarization techniques have been widely discussed by Ji et al. (2024) and others to condense lengthy clinical notes into concise summaries. Recently, large language models such as GPT-4 and others have gained attention for generating clinical notes from doctor-patient conversations. Researchers including J. Li et al. Ji et al. (2024) and Mahalaskhmi (27) have explored these models for their effectiveness in text generation.

Table 4: NLP Techniques utilized by different studies

NLP Techniques	Count
NER	70
BERT	50
Deep Learning	43
Transformer	39
BioBERT	32
GPT	12
LLM	11
Clinical BERT	6
Word2Vec	3

- **RQ2:** What are the transparency and trust challenges and concerns of using NLP in clinical decision-making?

The study's findings highlight transparency, trust challenges, and concerns in using NLP in clinical decision-making, especially with deep learning models. Due to the nature of NLP models and the complexity involved, they operate as a black box where it is possible to have a higher accuracy score in decision-making (25,34). However, the decision-making process and criteria to arrive at a decision are difficult for healthcare professionals to interpret, causing a trust deficit in the recommendations (25). Authors Zhou et al. (2024) have highlighted the interpretability and explainability concerns where the model makes certain decisions based on patterns and associations of different data points from the clinical notes. Several authors have discussed data privacy and security concerns while using patient data for training purposes (3,25,35,36). The data privacy and security challenges could aid in data breaches and privacy violations.

Table 5: Transparency and Trust challenges highlighted by various studies

Authors	Year	Transparency, Trust Challenges, and Concerns
Jennifer A. Salvi et al.	2023	Limited transparency in how NLP/ML systems identify and extract PRO data from EHRs can impact trust in the findings. The lack of access to the extracted data further limits transparency and the ability to independently verify results.
Kalavathi et al.	2023	While effective, the complexity of BERT-based models can make it challenging to understand their decision-making processes, hindering transparency and trust in the generated summaries.
A.A. Funkner et al.	2023	While RuBERT achieved high accuracy in symptom extraction, its inner workings remain opaque, potentially limiting trust in its outputs for clinical decision-making.
Ahne, A et al.	2022	The "black box" nature of BERT models necessitates developing techniques for interpretability and explainability to foster trust in their applications for clinical decision support. Data privacy and security concerns, as well as potential biases, need careful consideration.
Hankook Lee et al.	2023	The study highlights the need to integrate transparency and trust mechanisms to ensure the responsible use of the extracted alcohol-related information in clinical settings.
Amna Iqbal et al.	2023	Combining BERT and CHARACTER-BERT embeddings, while achieving high accuracy, may further obscure the model's reasoning, potentially hindering trust and transparency in cardiovascular risk assessment.

It goes without mentioning that addressing the challenges of transparency and trust is critical for the successful adoption of NLP in clinical decision-making. By developing more interpretable models, ensuring data privacy and security, mitigating biases, and improving generalizability, researchers, and developers can pave the path for trustworthy and reliable NLP-driven healthcare solutions that enhance patient care and clinical outcomes. There is a greater need for researchers and developers to prioritize transparency and trust when building these NLP models for clinical insights and decision-making. The next section discusses the various approaches to address the block box nature and enhance interpretability and explainability for responsible and trustworthy clinical decision support.

- **RQ3:** What are the novel approaches in NLP to improve transparency and trust in clinical decision-making?

Maintaining transparency and trust in NLP-powered clinical decision-making is imperative for ensuring patient safety. Transparency and trust are needed to uphold ethical standards, comply with regulatory requirements, and facilitate ongoing improvement. Several novel approaches have been proposed to improve transparency and trust in clinical decision-making. The approaches are categorized into two main types: 1) Improve the interpretability of models and 2) Enhance the trustworthiness of models.

One of the approaches proposed to improve interpretability is combining expert-based systems in machine learning (16,37). This hybrid approach allows for the development of models that leverage expert knowledge and clearly defined rules while incorporating machine learning's ability to learn from data. Another emerging approach to trustworthiness is explainable AI, which combines multiple disciplines of machine learning, such as attention, feature importance, and scenario simulations, to explain the decision-making process (3). Both of these approaches allow for transparency in learning and model reasoning to enhance trust in NLP systems, enabling clinical decision-making with confidence (37).

Healthcare is a unique area with complex terminology and concepts. By grounding the model's understanding in established medical concepts, terminology, and ontologies, it is possible to instill more confidence in clinicians during decision-making (34). The inclusion of prompt engineering to understand reasoning and evaluation of trustworthiness can improve reliability, fairness, and explainability. Ultimately, the choice of approach depends on the specific NLP application and the clinical context. Combining multiple approaches leads to the most robust and trustworthy NLP systems for clinical decision-making.

Table 6: Approach to improve transparency and trust by different studies

Authors	Approach to Improve Transparency and Trust
Sharaf & Anoop, 2023 (38)	Domain Knowledge Integration: BioBERT is pre-trained on a vast biomedical corpus, enhancing its understanding of medical language and improving trust by grounding decisions in established medical concepts.
Berloco et al., 2023(39)	Hybrid Model: Combines rule-based systems and machine learning, leveraging expert knowledge and data-driven insights for increased transparency and accuracy.
Srivastava et al., 2023 (20)	Comprehensive Evaluation and Multi-faceted Approach: Emphasizes a holistic approach including NLP, continuous monitoring, and data integration to provide a more complete and transparent view of patient data for decision-making.
Zhou et al., 2024(25)	Clinical Decision Support Systems based on Question Answering: Focuses on systems that mimic natural patient-provider communication, leading to improved interpretability and trust due to their familiarity and alignment with traditional clinical workflows.
Sreenivasgoud et al., 2023(16)	Combining Multiple NLP Techniques: Explores the use of TF-IDF, Word2Vec, NER, and LDA to provide a more comprehensive understanding of patient data, enhancing transparency by highlighting the strengths and limitations of each method in the decision-making process.
Haddad et al., 2021(26)	Automated Prescription Evaluation System: Utilizes text analysis algorithms to assess the compliance of doctor prescriptions with established standards, improving transparency by providing clear visualizations and tables comparing predicted outputs with standard values.

Explainable AI (XAI) also aims to make the operations of AI systems transparent, interpretable, and comprehensible to humans. XAI in Natural Language Processing (NLP) aims to make model decisions more transparent and interpretable. By using techniques like attention mechanisms and rule-based systems, XAI highlights the key input elements that impact a model's outcomes, enhancing our understanding of AI behavior and decision-making. It is crucial for grasping a model's reasoning and building trust in AI applications. XAI can also reveal biases in AI systems, shedding light on their workings.

5.3 Challenges and Limitations

Most literature has discussed the importance of large, high-quality datasets for effective training of the NLP models. Though training is crucial, obtaining labeled data in the healthcare domain is critical for higher accuracy. Labeled data has several challenges due to concerns around privacy, the complexity of medical information, and the massive effort required for manual annotation (3,17,20,25,26,40). Another limitation that emerged from the research was medical language complexity and ambiguity. There is a consensus among researchers that medicine is highly specialized and nuanced, making it difficult for NLP models to accurately interpret clinical text (3,25). Other researchers have highlighted the difficulty in handling the inconsistent formats and structure that make it hard for NLP models to identify similar information consistently.

Modern clinical documentation has undergone significant advancements, now incorporating not only patient narratives but also an extensive array of diagnostic tests and medical data. This growing complexity highlights the necessity for robust Electronic Medical Records (EMRs) and Electronic Health Records (EHRs) (3,5,26). However, many of these systems face challenges in effectively managing and integrating the increasing volume of patient information. Issues related to interoperability, usability, and data overload frequently impede their capacity to facilitate clinical decision-making and optimize workflow processes. Addressing these challenges is essential to fully leverage the potential of these systems in enhancing patient care and administrative efficiency (13).

Ethical considerations and patient privacy are significant concerns when using NLP models to analyze patient data (3). While NLP models have demonstrated their effectiveness in extracting and summarizing standard clinical documentation and serving specialized purposes, there are still worries about their generalizability and transferability across different healthcare settings and patient populations.

5.4 Current Gaps and Future Research Directions

Several gaps have been identified in previous research based on the points discussed in the results and discussion section. These include addressing issues such as data sparsity and domain-specific language variations to enhance the reliability and generalizability of NLP models (3). Additionally, the availability of limited data due to privacy concerns and the high costs of annotation, along with the nuanced medical language, specific domain terminologies, and context sensitivity, are areas that require further investigation (17,25,26). As with other machine learning and artificial intelligence applications, future research has significant potential to address ethical considerations in healthcare, where data privacy is critical and biases towards vulnerable populations are unacceptable. Future studies should focus on developing a standardized ethical framework for NLP usage in healthcare. Such a framework would promote transparent data handling, reduce bias in models, and minimize disparities in healthcare outcomes. Another important area for future research is using multimodal and longitudinal data to provide holistic and accurate insights into healthcare.

In clinical decision-making the success of ML systems hinges on the quality of their training data. Addressing biases and gaps in healthcare data is critical for improving the reliability and accuracy of predictive models. Current limitations in data representativeness, such as underrepresentation of certain populations or incomplete patient histories, can lead to skewed predictions and inequitable outcomes. Bridging these gaps is essential to ensure that AI systems deliver fair and accurate insights. (Antoniadi et al., 2021).

6. Conclusion

This research examines the transformative impact of Natural Language Processing (NLP) on improving clinical documentation and automating the generation of clinical notes. The findings show that NLP has significantly enhanced the efficiency and accuracy of clinical documentation by reducing manual data entry, improving data completeness, and standardizing medical information. By automating tasks such as clinical note summarization and transcription, NLP has lessened the administrative workload on clinicians, allowing them to concentrate more on patient care. Additionally, the study emphasizes important NLP techniques, such as deep learning models and transformer architectures, that facilitate automation in clinical documentation. These techniques enable the extraction of crucial information from unstructured clinical texts, supporting real-time decision-making and personalized patient care.

Despite recent advancements, challenges remain regarding transparency and trust in NLP-based decision-making. This study has explored various approaches to improve the understanding of how NLP models make decisions, such as explainable AI (XAI) and the integration of expert systems within these models. Enhancing transparency and trust in AI-based decisions is crucial. Future research should focus on addressing ethical concerns, including data privacy and bias. NLP has significant potential to transform healthcare documentation by providing more efficient, accurate, and context-aware solutions that improve clinical outcomes and streamline healthcare delivery.

References

[1] D. Katyan, G. Gulati, and G. Upreti, "Utilising NLP for Enhanced Clinical Text Mining," in 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), IEEE, Jun. 2024, pp. 883–889. doi:

10.1109/ICAAIC60222.2024.10575794.

[2] J. Wu, X. Ye, C. Mou, and W. Dai, "FineEHR: Refine Clinical Note Representations to Improve Mortality Prediction," in 2023 11th International Symposium on Digital Forensics and Security (ISDFS), IEEE, May 2023, pp. 1–6. doi: 10.1109/ISDFS58141.2023.10131726.

[3] K. Yadav, M. Sharma, J. Michaelson, K. Laxminarayanamma, V. K. Yadav, and O. Krishna, "BERT-Driven Natural Language Processing in Electronic Health Records for Improved Clinical Insights," in 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICAIIHI57871.2023.10489730.

[4] M. Ehghaghi, P.-Y. Zhou, W. Y. Cheng, S. Rajabi, C.-H. Kuo, and E.-S. A. Lee, "Interpretable Disease Prediction from Clinical Text by Leveraging Pattern Disentanglement," in 2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), IEEE, Oct. 2023, pp. 1–7. doi: 10.1109/BHI58575.2023.10313408.

[5] T. I. Amosa, L. I. B. Izhar, P. Sebastian, I. B. Ismail, O. Ibrahim, and S. L. Ayinla, "Clinical Errors From Acronym Use in Electronic Health Record: A Review of NLP-Based Disambiguation Techniques," IEEE Access, vol. 11, pp. 59297–59316, 2023, doi: 10.1109/ACCESS.2023.3284682.

[6] B. Javagal and S. Sharma, "Performance Evaluation of Clinical Models on Sequential Clinical Text for AI Powered Medical Applications," in 2024 IEEE International Conference on Contemporary Computing and Communications (InC4), IEEE, Mar. 2024, pp. 1–4. doi: 10.1109/InC460750.2024.10649316.

[7] Z. Li, K. Thaker, and D. He, "SiaKey: A Method for Improving Few-shot Learning with Clinical Domain Information," in 2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), IEEE, Oct. 2023, pp. 1–4. doi: 10.1109/BHI58575.2023.10313362.

[8] S. Sivarajkumar and Y. Wang, "Evaluation of Healthprompt for Zero-shot Clinical Text Classification," in 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI), IEEE, Jun. 2023, pp. 492–494. doi: 10.1109/ICHI57859.2023.00081.

[9] A. K. Rai, U. S. Aswal, S. K. Muthuvel, A. Sankhyan, S. L. Chari, and A. K. Rao, "Clinical Text Classification in Healthcare: Leveraging BERT for NLP," in 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), IEEE, Dec. 2023, pp. 1–7. doi: 10.1109/ICAIIHI57871.2023.10489434.

[10] K. P. P, R. Roj, J. Bhadra, and S. Dutta, "The Detection of Adverse Drug Reactions in Clinical Text Data Using Transformer Models," in 2023 Fourth International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICSTCEE60504.2023.10584859.

[11] S. S. Tiwari, H. R. Goyal, M. M. adnan, A. Shrivastava, K. Yadav, and A. L. N. Rao, "Clinical Text Analysis with Natural Language Processing: A BERT-based Approach," in 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), IEEE, May 2024, pp. 1382–1388. doi: 10.1109/IC3SE62002.2024.10593504.

[12] A. Khanday, S. Bouktif, and A. Ouni, "RNN-based Model for an Optimal COVID-19 Cases Detection using Clinical Reports," in 2023 9th International Conference on Optimization and Applications (ICOA), IEEE, Oct. 2023, pp. 1–6. doi: 10.1109/ICOA58279.2023.10308812.

[13] D. F. Navarro, "Natural Language Processing Models in Clinical Medicine: development, validation and implementation of automated text analysis in electronic medical records," Dec. 2023, doi: 10.25949/24330943.v1.

[14] J. Lenin, A. Komathi, H. Vijayan, A. R. Rathinam, A. Kasthuri, and C. Srinivasan, "Revolutionizing Healthcare With Cloud Computing: The Impact of Clinical Decision Support Algorithm," in 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICAIIHI57871.2023.10489022.

[15] N. Tyagi and B. Bhushan, "Demystifying the Role of Natural Language Processing (NLP) in Smart City Applications: Background, Motivation, Recent Advances, and Future Research Directions," Wirel Pers Commun, vol. 130, no. 2, pp. 857–908, May 2023, doi: 10.1007/s11277-023-10312-8.

[16] P. Sreenivasgoud, M. K. Sharma, B. S. Kumar, G. Nijhawan, H. M. Al-Jawahry, and R. Udhayakumar, "Natural Language Processing in Electronic Health Record Mining for Clinical Decision Support," in 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), IEEE, Dec. 2023, pp. 1–8. doi: 10.1109/ICAIIHI57871.2023.10489801.

[17] M. Arqam, M. Hussain, H. Zafar, A. Iqbal, and M. Liaqat, "Natural Language Processing for Diagnosis and Risk Assessment of Cardiovascular Disease," in 2023 International Conference on IT and Industrial Technologies (ICIT), IEEE, Oct. 2023, pp. 1–6. doi: 10.1109/ICIT59216.2023.10335838.

[18] Y. Ji, Z. Yu, and Y. Wang, "Assertion Detection in Clinical Natural Language Processing Using Large Language Models," in 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI), IEEE, Jun. 2024, pp. 242–247. doi: 10.1109/ICHI61247.2024.00039.

[19] A. M. Antoniadi et al., "Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review," Applied Sciences, vol. 11, no. 11, p. 5088, May 2021, doi: 10.3390/app11115088.

[20] D. K. Srivastava, S. Parveen, and D. K. Yadav, "Advanced AI Approaches for Detailed Examination of Individuals Prone to Nightly Respiratory Challenges Using Medical Records," in 2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG), IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICTBIG59752.2023.10455993.

[21] T. C. Kalaiselvi, C. N. Vanitha, and R. Vinodavarshini, "NLP-Powered Oncology Patient Summary," in 2023

7th International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, Nov. 2023, pp. 253–258. doi: 10.1109/ICECA58529.2023.10395425.
[22] D. Tranfield, D. Denyer, and P. Smart, "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review," British Journal of Management, vol. 14, no. 3, pp. 207–222, Sep. 2003, doi: 10.1111/1467-8551.00375.
[23] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," BMJ, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.
[24] N. Shaheen et al., "Appraising systematic reviews: a comprehensive guide to ensuring validity and reliability," Front Res Metr Anal, vol. 8, Dec. 2023, doi: 10.3389/frma.2023.1268045.
[25] B. Zhou, G. Yang, Z. Shi, and S. Ma, "Natural Language Processing for Smart Healthcare," IEEE Rev Biomed Eng, vol. 17, pp. 4–18, 2024, doi: 10.1109/RBME.2022.3210270.
[26] N. Haddad, K. S. Myshenkov, and G. I. Afanasiev, "Introducing Text Analysis Algorithms in Decision Support Systems for Automated Evaluation of the Doctor Prescriptions," in 2024 6th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), IEEE, Feb. 2024, pp. 1–5. doi: 10.1109/REEPE60449.2024.10479914.
[27] M. Mahalakshmi, S. Bharadwaj, and A. N. Bhuyan, "A Real-Time Medical Report Analysis and AI-Powered Diagnosis: A Cloud-Based Solution for Improved Patient Care," in 2024 Second International Conference on Advances in Information Technology (ICAIT), IEEE, Jul. 2024, pp. 1–6. doi: 10.1109/ICAIT61638.2024.10690733.
[28] Z. Sun and C. Tao, "Named Entity Recognition and Normalization for Alzheimer's Disease Eligibility Criteria," in 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI), IEEE, Jun. 2023, pp. 558–564. doi: 10.1109/ICHI57859.2023.00100.
[29] A. Chen et al., "Identifying Symptoms of Delirium from Clinical Narratives Using Natural Language Processing," in 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI), IEEE, Jun. 2024, pp. 305–311. doi: 10.1109/ICHI61247.2024.00046.
[30] A. Sankaranarayanan, D. Shetty, K. Chetwani, and B. R. Shambhavi, "Exploring BioClinical BERT's NLP Capabilities with Explainability Techniques," in 2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS), IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/ICETCS61022.2024.10544307.
[31] S. Pal et al., "Recommendation System for Clinical Concept Mapping," in 2023 International Conference on Computational Intelligence, Networks and Security (ICCINS), IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICCINS58907.2023.10450023.
[32] K. Ananthajothi, J. David, and A. Kavin, "Cardiovascular Disease Prediction Using Langchain," in 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), IEEE, May 2024, pp. 1–6. doi: 10.1109/ACCAI61061.2024.10601906.
[33] Y. Ji, Z. Yu, and Y. Wang, "Assertion Detection in Clinical Natural Language Processing Using Large Language Models," in 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI), IEEE, Jun. 2024, pp. 242–247. doi: 10.1109/ICHI61247.2024.00039.
[34] P. K. Nag, A. Bhagat, R. Vishnu Priya, and D. K. Khare, "Emotional Intelligence Through Artificial Intelligence: NLP and Deep Learning in the Analysis of Healthcare Texts," in 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), IEEE, Dec. 2023, pp. 1–7. doi: 10.1109/ICAIIHI57871.2023.10489117.
[35] J. Jeon, S. Cho, D. Lee, C. Lee, and J. Kim, "BioBridge: Unified Bio-Embedding With Bridging Modality in Code-Switched EMR," IEEE Access, vol. 12, pp. 141866–141877, 2024, doi: 10.1109/ACCESS.2024.3467251.
[36] A. Bhandari, "Revolutionizing Radiology With Artificial Intelligence," Cureus, Oct. 2024, doi: 10.7759/cureus.72646.
[37] J. Sim et al., "Natural language processing with machine learning methods to analyze unstructured patient-reported outcomes derived from electronic health records: A systematic review," Artif Intell Med, vol. 146, p. 102701, Dec. 2023, doi: 10.1016/j.artmed.2023.102701.
[38] S. Sharaf and V. S. Anoop, "An Analysis on Large Language Models in Healthcare: A Case Study of BioBERT," Oct. 2023.
[39] F. Berloco, S. Ciavarella, S. Colucci, L. A. Grieco, A. Guarini, and G. M. Zaccaria, "ARGO 2.0: a Hybrid NLP/ML Framework for Diagnosis Standardization*," in 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, Jul. 2023, pp. 1–4. doi: 10.1109/EMBC40787.2023.10340022.
[40] S. K. Nayak, M. Garanayak, S. K. Swain, S. K. Panda, and D. Godavarthi, "An Intelligent Disease Prediction and Drug Recommendation Prototype by Using Multiple Approaches of Machine Learning Algorithms," IEEE Access, vol. 11, pp. 99304–99318, 2023, doi: 10.1109/ACCESS.2023.3314332

Author Profile

Chandra Prakash earned his Master's Diploma in 2004, MS in 2021, and a Ph.D. in IT from the University of the Cumberlands in 2024. A distinguished contributor to the ASP.NET community, he has authored 3,400+ posts on the ASP.NET Forum with an All-Star contributor ranking. In 2012, he developed the Sitecore development framework to streamline Sitecore development ahead of the official framework released in 2015. In his current role, Chandra's work has substantially impacted healthcare delivery, brought operational efficiency, and improved patient experience. His research interests focus on advancing software architecture, microservices, zero-trust security frameworks, and the application of machine learning in healthcare.

Musa Shaikh holds a Master's degree in Computer Science from Pune University, India, awarded in 2002, and a Master of Science in Artificial Intelligence from the University of the Cumberlands, USA, conferred in 2025. Currently employed as an Architect specializing in an integration platform and possesses extensive technical expertise in the IBM Integration products suite. Additionally, I am proficient in Azure Cloud, Node.js, StreamSets, and Kafka.

Pavan Mutha is pursuing a Master of Science in Artificial Intelligence from the University of the Cumberlands. Before this, Mr. Mutha completed a Master of Science in Information Technology from the University of the Cumberlands, a Bachelor of Mechanical Engineering from Nagpur University, and a Diploma in Automobile Engineering with around 20 years of experience in the field of information technology, business management, supply chain, and logistics domain. Mr. Mutha is an avid follower and learner of transition in the technology space and exploring AI/ML and its interaction with existing business processes.

Praveen Kumar Parimi is currently pursuing a Master of Science in Artificial Intelligence from the University of the Cumberlands. Before this, Mr. Parimi completed a Bachelor of Technology in Computer Science and Engineering from Vellore Institute of Technology. Mr. Parimi is a certified Annuities and Life Insurance business domain specialist with AAPA and LOMA certifications. He is also a certified Salesforce Architect, helping clients meet their business needs by optimizing, designing, and providing robust solutions on the Salesforce ecosystem and integrating with other cloud platforms. With 18 years of experience in the IT industry, he has worked across diverse domains, including Banking and Financial Services, Healthcare, and Life P&C Insurance. Currently, Praveen serves as a Salesforce Architect, specializing in designing and developing innovative solutions. As a passionate learner and explorer, he actively explores advancements in technology, focusing on various subfields of Artificial Intelligence, including Machine Learning, Deep Learning, and Natural Language Processing, and how these subdomains can help businesses build more optimized and robust processes for better customer satisfaction.