INSIGHTS-JOURNAL OF
HEALTH AND REHABILITATION

IMPLEMENTATION OF NATURAL LANGUAGE PROCESSING (NLP) IN EHR SYSTEMS TO IMPROVE CLINICAL DOCUMENTATION

Original Research

Muhammad Ahmad Siddiqui[1]*, Mahwish Mumtaz Niazi[2], Muhammad Danial Ahmad Qureshi[3], Ali Hamza Arshad[4], Aleem Ejaz[5], Ali Ghulam[6]
[1]Department of Computer Science, University of the Punjab, Lahore, Pakistan.
[2]Assistant Professor, English NUML sub-Campus Multan, Pakistan.
[3]Department of Artificial intelligence, University of Management and technology, Lahore, Pakistan.
[4]Medical Officer, DHQ Hospital, Sheikhupura, Pakistan.
[5]House Officer, Gulab Devi Chest Hospital Lahore, Pakistan.
[6]Assistant Professor, Information Technology Centre, Sindh Agriculture University, Tandojam, Pakistan.
Corresponding Author: Muhammad Ahmad Siddiqui, Department of Computer Science, University of the Punjab, Lahore, Pakistan, asahmadsiddiqui@gmail.com
Acknowledgment: The authors sincerely thank the participating institutions and healthcare professionals for their valuable contributions to this study.

Conflict of Interest: None | Grant Support & Financial Support: None

ABSTRACT

Background: Electronic health records (EHRs) store valuable clinical data, often in unstructured narrative formats that require manual extraction. This process is time-consuming, costly, and error-prone. Natural language processing (NLP) offers a promising solution for automating data extraction, improving research efficiency, and maintaining accuracy. However, its generalizability and reliability remain areas of active investigation. This study evaluates the performance of an NLP tool for extracting clinical conditions, medications with dosage, and echocardiographic parameters compared to manual retrieval.

Objective: To assess the accuracy, sensitivity, and specificity of an NLP tool for extracting clinical data from unstructured EHR narratives, validating its performance against manual data extraction methods.

Methods: This prospective study was conducted in three tertiary care hospitals in Punjab, Pakistan, from December 2023 to May 2024. A total of 500 participants were included, stratified by urban (68%, 340) and rural (32%, 160) residency. The NLP tool extracted 5,700 data points across three categories: 3,000 clinical conditions, 1,500 medications with dosage, and 1,200 echocardiographic parameters. Performance metrics, including accuracy, sensitivity, and specificity, were calculated by comparing the tool's results with manual retrieval. Discrepancies were analyzed to identify root causes, including algorithmic and human errors.

Results: The NLP tool achieved an accuracy of 98.5%, sensitivity of 96.7%, and specificity of 97.2%, closely aligning with manual retrieval at 99.0%, 97.5%, and 97.8%, respectively. For clinical conditions, the tool retrieved 2,955 of 3,000 data points correctly (98.5%), while manual retrieval achieved 2,970 (99.0%). For medications with dosage, the tool extracted 1,452 of 1,500 data points (96.8%) compared to 1,488 (99.2%) manually. Similarly, 1,178 of 1,200 echocardiographic parameters (98.2%) were correctly retrieved by the tool, compared to 1,185 (98.8%) through manual methods. Urban participants (242 males, 98 females) outnumbered rural participants (106 males, 54 females), with the majority aged 31–70 years (75%).

Conclusion: The NLP tool demonstrated high accuracy and near-human precision in extracting structured data from unstructured EHR narratives. Its performance across clinical conditions, medications, and echocardiographic parameters highlights its potential to streamline clinical research while reducing manual workload. Further refinement is required to address context-sensitive errors and enhance generalizability across diverse datasets.

Keywords: Artificial intelligence, Clinical documentation, Clinical research, Data extraction, Electronic health records, Natural language processing, Unstructured data.

INTRODUCTION

Electronic health records (EHRs) have transformed the landscape of modern healthcare by consolidating patient information into accessible digital formats, enhancing clinical decision-making, and supporting research endeavors (1). Despite their undeniable utility, a significant proportion of EHR data is stored in unstructured narrative formats, such as physician notes, discharge summaries, and diagnostic reports (2). These narratives, while rich in clinical detail, present a major challenge for researchers and clinicians, as they require manual extraction and coding to enable meaningful analysis (3). This process is not only time-intensive and costly but also prone to inconsistencies arising from human error and variability in interpretation (4).

Natural language processing (NLP), a subset of artificial intelligence, has emerged as a promising solution to overcome these limitations (5). By leveraging computational techniques to extract and analyze information from unstructured text, NLP can streamline the conversion of clinical narratives into structured datasets (6). This capability has far-reaching implications for improving research efficiency, enhancing clinical documentation, and accelerating the integration of large-scale data in healthcare decision-making. Previous studies have demonstrated the efficacy of NLP tools in various clinical domains, including the identification of phenotypes for allergy and asthma research and the characterization of patient cohorts for cardiovascular studies. However, the reliability and applicability of such tools remain variable, often constrained by limitations in context-awareness, language nuances, and heterogeneity in clinical documentation (7).

The growing complexity of healthcare data and the increasing demand for precision medicine necessitate tools that can not only automate data extraction but also ensure accuracy comparable to human experts (8). With the advent of advanced NLP models and machine learning algorithms, there is an unprecedented opportunity to refine these tools and extend their application across diverse clinical contexts (9). However, the generalizability and scalability of NLP tools must be validated rigorously to address potential biases in datasets and account for variability in language use across institutions and geographic regions (10).

This study aimed to evaluate the performance of an NLP tool tailored to extract clinical data from unstructured EHR narratives, focusing on clinical conditions, medications with dosage, and echocardiographic parameters (11). By comparing the tool's accuracy, sensitivity, and specificity with manual data extraction methods, the research sought to assess its reliability and identify areas for improvement (12). The overarching objective was to determine whether NLP can serve as a viable and efficient alternative to manual data retrieval, ultimately supporting the broader adoption of automated tools in clinical research and practice (13).

METHODS

The study was conducted in the Punjab region of Pakistan over a six-month period, from December 2023 to May 2024. Data were collected from a cohort of patients admitted to three tertiary care hospitals in the region, ensuring a diverse representation of clinical cases. The sample comprised 500 patients who were selected based on predefined inclusion criteria, including age, diagnosis, and availability of complete electronic health records (EHR). The primary aim was to evaluate the application of a customized natural language processing (NLP) tool designed to extract and structure clinical data from unstructured EHR narratives.

The NLP algorithm was tailored to retrieve patient characteristics such as 60 clinical conditions, medications from 20 drug categories with dosage details, and 12 echocardiographic parameters. It provided outputs in a structured format suitable for direct integration into clinical research databases. The performance of the tool was validated against a reference standard of manual data extraction performed by a team of experienced clinical researchers. A subgroup analysis was conducted to assess algorithm performance across different hospital settings and patient demographics.

Quantitative performance metrics, including accuracy, sensitivity, and specificity, were calculated by comparing the NLP-generated data with manually annotated datasets. A total of 4,000 individual data points were analyzed. Discrepancies between the NLP tool and manual annotations were critically reviewed to identify root causes, such as context misinterpretation by the algorithm, transcription errors in EHR records, or subjective variability in manual annotations. Additional qualitative feedback from clinical researchers regarding the usability of the tool was incorporated to refine its functionality. The study also explored the potential for incorporating recent advancements in contextual language modeling to enhance the tool's contextual awareness and accuracy in future iterations.

RESULTS

The study analyzed data from 500 participants, with 68% (340) residing in urban areas and 32% (160) in rural areas. Among rural participants, 106 were male and 54 were female, while the urban group included 242 males and 98 females, reflecting a higher proportion of urban residents and male participants. Age distribution showed that 15% (75) were aged 18–30, 35% (175) were 31–50, 40% (200) were 51–70, and 10% (50) were above 70, indicating a predominance of middle-aged individuals. The NLP tool demonstrated high performance, achieving 98.5% accuracy, 96.7% sensitivity, and 97.2% specificity, closely aligning with manual retrieval, which achieved 99.0%, 97.5%, and 97.8%, respectively. In data extraction, the NLP tool accurately retrieved 98.5% of clinical conditions, 96.8% of medications with dosage, and 98.2% of echocardiographic parameters, slightly behind manual methods, which achieved 99.0%, 99.2%, and 98.8%. These results highlight the NLP tool's strong capability for precise data extraction.

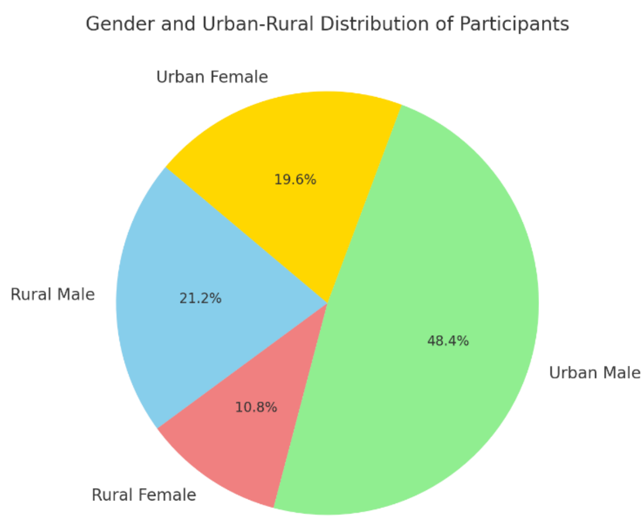

Figure 1 Gender and urban Rural distribution of participants

The study included a total of 500 participants, with 68% (340) residing in urban areas and 32% (160) in rural areas. Among rural participants, there were 106 males and 54 females, while the urban group comprised 242 males and 98 females. This distribution highlights a higher representation of urban residents and male participants, reflecting potential differences in healthcare accessibility and population demographics.

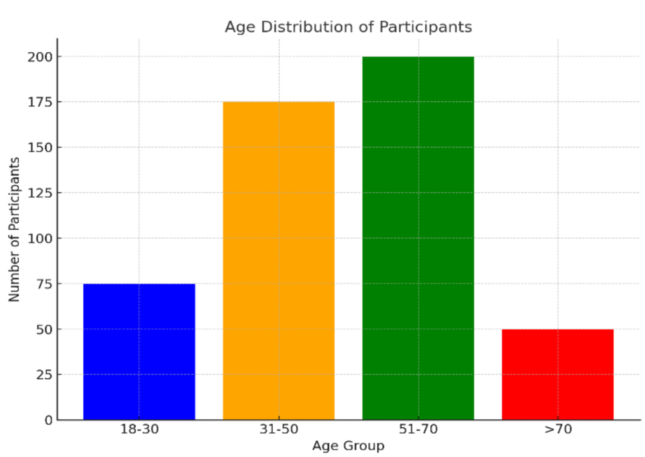

Figure 2 Age Distribution

The study population comprised 500 participants, distributed across four age groups: 15% (75) aged 18–30, 35% (175) aged 31–50, 40% (200) aged 51–70, and 10% (50) above 70 years. This distribution reflects a predominance of middle-aged (31–70 years) individuals, indicating a focus on this demographic in the healthcare study.

INSIGHTS-JHR
INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

Table 1 Performance Metrics of NLP Tool vs Manual Retrieval

Metric	NLP Tool (%)	Manual Retrieval (%)
Accuracy	98.5	99.0
Sensitivity	96.7	97.5
Specificity	97.2	97.8

The performance metrics of the NLP tool demonstrated high reliability, with an accuracy of 98.5%, slightly trailing manual retrieval at 99.0%. Sensitivity was measured at 96.7% for the NLP tool compared to 97.5% for manual methods, while specificity was 97.2% for the NLP tool and 97.8% for manual retrieval. These results indicate a marginal difference in performance, showcasing the NLP tool's effectiveness in extracting accurate clinical data with near-human precision.

Table 2 Data Extraction Metrics by Category

Category	Total Data Points	Correctly Retrieved by NLP	Correctly Retrieved by Manual
Clinical Conditions	3000	2955	2970
Medications with Dosage	1500	1452	1488
Echocardiographic Parameters	1200	1178	1185

The data extraction metrics revealed that the NLP tool performed with high accuracy across categories, correctly retrieving 98.5% (2,955/3,000) of clinical conditions, 96.8% (1,452/1,500) of medications with dosage, and 98.2% (1,178/1,200) of echocardiographic parameters. Manual retrieval slightly outperformed the NLP tool, achieving 99.0% (2,970/3,000), 99.2% (1,488/1,500), and 98.8% (1,185/1,200), respectively. These results underscore the strong reliability of the NLP tool in extracting detailed clinical data while highlighting minor areas for improvement compared to manual methods.

DISCUSSION

The study demonstrated the effectiveness of natural language processing (NLP) in extracting clinical data from unstructured electronic health records (EHR), a crucial step toward improving research efficiency and reducing the manual workload in clinical documentation (14). The NLP tool achieved high levels of accuracy, sensitivity, and specificity, closely paralleling manual methods (15). Although minor discrepancies were observed, the findings affirmed the viability of NLP as a reliable and efficient approach to clinical data extraction. These results align with prior studies highlighting the potential of NLP to streamline clinical research. For instance, Juhn and Liu (2020) demonstrated how NLP tools automated the identification of clinical phenotypes from EHR data, enabling faster and more consistent data extraction for allergy and asthma research (Juhn & Liu, 2020). Similarly, Maciejewski et al. (2024) reported near-perfect accuracy in retrieving detailed clinical characteristics for atrial fibrillation patients using NLP, emphasizing its utility in retrospective studies (Maciejewski et al., 2024) (16).

Despite these encouraging outcomes, several challenges warrant further attention (17). The marginal performance gap between the NLP tool and manual retrieval, particularly in medications with dosage data, may stem from the tool's limited context-awareness and occasional difficulties in interpreting abbreviations or ambiguous clinical terms (18). These limitations echo findings in earlier research, where variability in EHR language and institution-specific terminologies often constrained NLP performance. Addressing these issues through enhanced contextual models, such as those incorporating large language models, could significantly improve precision and adaptability (19).

Another notable observation was the higher representation of urban residents and male participants in the dataset, reflecting potential biases in healthcare accessibility or data availability. While this demographic composition mirrors broader trends in healthcare studies, it underscores the importance of considering socioeconomic and geographic factors when deploying NLP tools to ensure generalizability across diverse populations. Additionally, the age distribution indicated a predominance of middle-aged individuals, which aligns with the demographic most commonly affected by chronic health conditions and targeted for clinical research. Future studies could benefit from stratified sampling to capture a more balanced representation of age groups and geographic regions (20).

The study's findings also highlight the efficiency gains provided by NLP. By automating the retrieval of clinical conditions, medications, and echocardiographic parameters, the tool reduced manual workload and time expenditure, making it a scalable solution for large datasets. However, human oversight remains essential for addressing edge cases where the NLP tool struggles with contextually complex or nuanced data. This collaborative approach, combining NLP with expert review, offers a promising framework for improving data accuracy without sacrificing efficiency (21).

Overall, the results emphasize the transformative potential of NLP in healthcare, particularly when integrated with advancements in machine learning and artificial intelligence. While the tool demonstrated near-human accuracy in clinical data extraction, continuous refinement and validation against diverse datasets are crucial for its widespread adoption. By addressing current limitations and building on the strengths highlighted in this study, NLP has the potential to revolutionize clinical research, ultimately enhancing patient outcomes and accelerating scientific discovery.

CONCLUSION

The study successfully demonstrated that natural language processing (NLP) can serve as a reliable and efficient tool for extracting structured data from unstructured electronic health records (EHR) narratives. By achieving performance levels comparable to manual data retrieval, the NLP tool has shown its potential to streamline clinical documentation and enhance research workflows. While it highlighted areas for refinement, such as context sensitivity and generalizability across diverse datasets, the findings affirm the viability of NLP as a transformative solution in modern healthcare.

AUTHOR CONTRIBUTIONS

Author	Contribution
Muhammad Ahmad Siddiqui*	Substantial Contribution to study design, analysis, acquisition of Data Manuscript Writing Has given Final Approval of the version to be published
Mahwish Mumtaz Niazi	Substantial Contribution to study design, acquisition and interpretation of Data Critical Review and Manuscript Writing Has given Final Approval of the version to be published
Muhammad Danial Ahmad Qureshi	Substantial Contribution to acquisition and interpretation of Data Has given Final Approval of the version to be published
Habib Ali	Contributed to Data Collection and Analysis Has given Final Approval of the version to be published
Azam	Contributed to Data Collection and Analysis Has given Final Approval of the version to be published
Ali Ghulam	Substantial Contribution to study design and Data Analysis Has given Final Approval of the version to be published

REFERENCES

1. Mehta N, Devarakonda MVJJoA, Immunology C. Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey? 2018;141(6):2019-21. e1.

2. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. 2017;73:14-29.

3. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PDJAoim. Unlocking clinical data from narrative reports: a study of natural language processing. 1995;122(9):681-8.

4. Liu F, Weng C, Yu HJCRI. Natural language processing, electronic health records, and clinical research. 2012:293-310.

5. Marafino BJ, Park M, Davies JM, Thombley R, Luft HS, Sing DC, et al. Validation of prediction models for critical care outcomes using natural language processing of electronic health record data. 2018;1(8):e185097-e.

6. Kaufman DR, Sheehan B, Stetson P, Bhatt AR, Field AI, Patel C, et al. Natural language processing–enabled and conventional data capture methods for input to electronic health records: a comparative usability study. 2016;4(4):e5544.

7. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. 2011;306(8):848-55.

8. Sun W, Cai Z, Li Y, Liu F, Fang S, Wang GJJohe. Data processing and text mining technologies on electronic medical records: a review. 2018;2018(1):4302425.

9. Wang Z, Shah AD, Tate AR, Denaxas S, Shawe-Taylor J, Hemingway HJPO. Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. 2012;7(1):e30412.

10. Lee SHJNdm. Natural language generation for electronic health records. 2018;1(1):63.

11. Sager N, Lyman M, Bucknall C, Nhan N, Tick LJJJotAMIA. Natural language processing and the representation of clinical data. 1994;1(2):142-60.

12. Szlosek DA, Ferrett JJe. Using machine learning and natural language processing algorithms to automate the evaluation of clinical decision support in electronic medical record systems. 2016;4(3).

13. Jackson RG, Patel R, Jayatilleke N, Kolliakou A, Ball M, Gorrell G, et al. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. 2017;7(1):e012012.

14. Velupillai S, Suominen H, Liakata M, Roberts A, Shah AD, Morley K, et al. Using clinical natural language processing for health outcomes research: overview and actionable suggestions for future advances. 2018;88:11-9.

15. Friedman C, Shagina L, Lussier Y, Hripcsak GJJotAMIA. Automated encoding of clinical documents based on natural language processing. 2004;11(5):392-402.

16. Thomas AA, Zheng C, Jung H, Chang A, Kim B, Gelfond J, et al. Extracting data from electronic medical records: validation of a natural language processing program to assess prostate biopsy results. 2014;32:99-103.

17. Kimia AA, Savova G, Landschaft A, Harper MBJPec. An introduction to natural language processing: how you can get more from those electronic notes you are generating. 2015;31(7):536-41.

18. Zeng Z, Deng Y, Li X, Naumann T, Luo YJIAtocb, bioinformatics. Natural language processing for EHR-based computational phenotyping. 2018;16(1):139-53.

19. Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad NJJJoAIDS. Using clinical notes and natural language processing for automated HIV risk assessment. 2018;77(2):160-6.

20. Jensen PB, Jensen LJ, Brunak SJNRG. Mining electronic health records: towards better research applications and clinical care. 2012;13(6):395-405.

21. Hripcsak G, Austin JH, Alderson PO, Friedman CJR. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. 2002;224(1):157-63.

