OPEN ACCESS

Saudi Journal of Medicine
Abbreviated Key Title: Saudi J Med
ISSN 2518-3389 (Print) | ISSN 2518-3397 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Review Article

Natural Language Processing in Electronic Health Records: Progress, Challenges, and Future Directions

Faisal Mansour Alanazi[1]*, Meshari Ali Aljjedaee[1], Shalah Al Harbl[1], Ahmed Abdullah Alsharekh[1], Dheifallah Alrashidi[1]

[1]Health Information Techian, Health Affairs at the Ministry of National Guard

DOI: https://doi.org/10.36348/sjm.2025.v10i09.006 | **Received:** 28.07.2025 | **Accepted:** 25.09.2025 | **Published:** 27.09.2025

***Corresponding Author:** Faisal Mansour Alanazi
Health Information Techian, Health Affairs at the Ministry of National Guard

Abstract

Electronic health record systems transformed healthcare documentation by providing a system for storing and sharing extensive patient data. However, much of this information remains in the form of unstructured text, which limits its utility for computational analysis. Natural Language Processing (NLP) has emerged as a prominent approach to extract and structure information from free-text clinical narratives, offering the potential to unlock valuable insights for clinical care, research, and administration. This paper provides an overview of recent advances in NLP methods applied to EHRs, discusses open problems including data quality, privacy, and generalizability, and highlights potential future directions for the integration of NLP into clinical workflows. The conclusions point to the need for continued development of domain-specific language models, privacy-preserving techniques, and explainable AI methods to fully harness the power of NLP for healthcare transformation.

Keywords: Electronic Health Records (EHRs), Natural Language Processing (NLP), Unstructured Clinical Text, Clinical Narratives, Information Extraction, Healthcare Data Analytics.

INTRODUCTION

Electronic Health Records (EHRs) have been deployed rapidly across the world during recent decades because they serve as fundamental tools for medical care delivery as well as biomedical research and healthcare administration (Jensen *et al.,* 2012). The adoption of electronic records expanded due to government incentives and regulations while technology innovations advanced alongside the benefits of electronic records which surpassed paper records regarding efficiency, safety and quality (Rajkomar *et al.,* 2018). As a result, EHRs have become the primary digital source for patient health information and a rich, accessible source for health data analytics, although major gaps and challenges in data quality still need to be addressed (Jensen *et al.,* 2012). While structured fields of EHRs (e.g., laboratory test values, medication lists, and vital signs) are readily accessible for computational analysis, most of the clinical information in EHRs remains in the form of unstructured free-text reports and notes (Murphy *et al.,* 2017). These text narratives are typically authored by clinicians for clinical purposes and include progress notes, discharge summaries, pathology and radiology reports. They contain a substantial amount of information not available in structured fields, including important context and rich clinical details. However, their unstructured nature renders them largely inaccessible to efficient secondary use and large-scale data analyses.

Natural Language Processing (NLP) can be used to automatically extract information of clinical interest from unstructured free text. Early clinical NLP systems were primarily based on rules and dictionary lookup, which led to many successes in domain-specific applications but were brittle and often failed to scale to other institutions (Stanfill *et al.,* 2010). Over the past decade, the emergence of statistical learning methods, and in particular deep learning, has led to a renaissance in NLP and enabled more flexible and accurate methods for clinical narratives (Wu *et al.,* 2020). The recent development of large transformer-based language models, such as BERT, BioBERT, and ClinicalBERT, have further pushed the performance frontier by leveraging deep contextual representations and mitigating error propagation in tasks such as entity recognition, temporal reasoning, and text summarization (Alsentzer *et al.,* 2019). Clinical NLP has since become an active and mature area of research, with a wide range of applications demonstrated in EHR data including clinical coding (Li *et al.,* 2019), adverse event detection (Borjali *et al.,* 2021), predictive modeling (Levin *et al.,* 2021), clinical language understanding (Savova *et al.,* 2010), and real-time decision support (Wang *et al.,* 2020).

In spite of the rapid progress, there are several challenges to the wide-scale adoption of clinical NLP in healthcare systems. Lack of standardization in clinical documentation practices, privacy and security issues, lack of generalizability across institutions, and concerns over explainability of deep learning models are some of the major technical and operational hurdles (Ghassemi *et al.,* 2021). In addition, there are ethical and regulatory issues surrounding the use of NLP and artificial intelligence in clinical settings, including the potential for algorithmic bias and other unintended consequences, as well as data privacy and security regulations such as HIPAA (Char *et al.,* 2018). Furthermore, many NLP tools have only been tested in academic settings and are not yet integrated into routine clinical workflows. As such, there is a need for more research on the translation and implementation of NLP in clinical practice.

On the other hand, given the massive untapped value of clinical text in EHRs, there is also a pressing need to leverage NLP to advance research and transform healthcare. Therefore, it is timely to review the state-of-the-art in NLP for EHRs to summarize the main applications, limitations, and research directions. This paper provides an overview of recent progress in NLP for EHRs, identifies ongoing challenges, and explores opportunities for developing future systems that are accurate, transparent, and seamlessly integrated into clinical workflows.

2. METHODS

The study presents a narrative review structure. Peer-reviewed journal articles, conference papers, and reports published from 2012 to 2023 were identified through database searches in PubMed, IEEE Xplore, and Google Scholar. The research used search terms such as “natural language processing in EHRs,” “clinical NLP,” “healthcare AI,” and “clinical text mining” to locate relevant literature. The titles and abstracts of identified papers were screened for relevance, focusing on those that discuss NLP applications in healthcare documentation, information extraction, clinical coding, predictive modeling, and integration with machine learning systems. The study included seminal works on transformer architectures and domain-specific language models to provide an overview of recent advancements.

3. PROGRESS IN NLP FOR EHRS

3.1 Information Extraction

Information extraction has arguably been one of the longest established NLP applications in EHRs. A number of methods for NER and relation extraction have been developed to recognize clinical entities such as diagnoses, medications, symptoms, and lab results. The initial systems cTAKES and MetaMap pioneered the automation of recognizing clinical entities and mapping to medical ontologies (Savova *et al.,* 2010). The recent development of deep learning-based methods, such as BiLSTM and transformer-based models, have matched human-level performance in NER for public benchmark datasets such as i2b2 and MIMIC-III (Si *et al.,* 2019). This has made large-scale epidemiologic studies, adverse drug event detection, and real-world evidence generation possible (Shivade *et al.,* 2014; Wang *et al.,* 2018).

3.2 Clinical Coding and Documentation

The development of automatic NLP-based coding systems enables the conversion of unrestricted clinical text notes into standardized vocabularies such as ICD-10 and SNOMED CT which decreases the manual workload of clinical coders. In addition to potential efficiency gains, these systems have been associated with additional benefits such as increased consistency and decreased errors in billing and improved interoperability (Stanfill *et al.,* 2010). Recently, hybrid models that integrate rule-based features and contextual embeddings have demonstrated improvements in accuracy for complex multi-label coding tasks (Liu, *et al.,* 2021). NLP has also been used in automated clinical documentation systems, where NLP tools are used to automatically generate structured clinical notes either from speech or unstructured narrative text (Jiang *et al.,* 2011)

3.3 Predictive Analytics

Researchers have utilized NLP-generated features for predictive modeling across various clinical tasks including early sepsis detection and hospital readmissions as well as predicting mortality (Rajkomar *et al.,* 2018). In one study, the inclusion of free-text nursing notes into a predictive model was found to outperform the predictive performance of using structured vital signs alone, and to predict patient deterioration earlier (Sarker, *et al.,* 2019). Models using a mixture of both structured and unstructured features have outperformed structured-only models consistently (Rumshisky *et al.,* 2016). These are just a few examples that help quantify the value NLP can add for risk stratification and decision support

3.4 Impact of Transformer-Based Models

Transformer-based models such as BERT and GPT, along with domain specific adaptations such as BioBERT, BlueBERT, and ClinicalBERT have set new standards for clinical NLP benchmarks (Alsentzer *et al.,* 2019). These models learn contextual semantics and have demonstrated state-of-the-art performance on tasks including clinical concept extraction, note summarization, and temporal reasoning. In addition, pretrained large language models have facilitated few-shot and zero-shot learning, making these models applicable to new tasks with very little labeled training data (Lee *et al.,* 2020). Clinical summarization and automated report generation are promising applications of such techniques.

4. Challenges

4.1 Data Quality and Variability

Electronic health record data contains inconsistencies and errors such as idiosyncratic usage patterns with typos and abbreviations according to Murphy et al. (2017). The heterogeneity and variability between institutions are significant issues. Domain adaptation and data harmonization are areas of active research (Shickel *et al.,* 2017).

4.2 Privacy and Security

Clinical text can be de-identified, but re-identification through residual identifiers in the text is still possible (Johnson *et al.,* 2023). De-identification using NLP methods, such as deep learning-based de-identification models, has better recall and precision but struggles with differentiating clinical and non-clinical entities (Liu *et al.,* 2021). There are still challenges in protecting sensitive patient information while still using it to train useful models. The privacy and security concerns of using EHR data should be continuously assessed while being compliant with regulations such as the Health Insurance Portability and Accountability Act (HIPPA) and General Data Protection Regulation (GDPR).

4.3 Generalizability

Models trained on a single institution have limited generalizability as documentation styles and patient populations may differ between institutions (Wu *et al.,* 2020). Domain shift is a more significant problem for deep learning models that require a lot of data to learn from. One way to address this problem is through multi-center collaborations and federated learning approaches (Rieke *et al.,* 2020).

4.4 Explainability and Trust

Clinicians require transparency in clinical decision support tools, but NLP methods are often perceived as black-box models. The lack of interpretability of these models presents a significant challenge in gaining clinical acceptance (Ghassemi *et al.,* 2021). Some explainable AI methods such as attention visualization and surrogate models have been proposed recently. However, these techniques have not been widely tested for clinical utility (Doshi-Velez & Kim, 2017).

4.5 Regulatory and Ethical Barriers

Clinically implementing NLP applications requires adhering to ethical standards and guidelines. Algorithmic bias, fairness, and accountability are major considerations, as biased NLP models can exacerbate existing disparities in healthcare (Char *et al.,* 2018). Standardized evaluation frameworks and clinical validation studies are necessary before clinical implementation.

5. Future Directions

5.1 Federated and Privacy-Preserving Learning

Approaches such as federated learning and differential privacy can facilitate the training of multi-institutional models without requiring data sharing, helping with both robustness and patient privacy (Rieke *et al.,* 2020).

5.2 Domain-Specific Foundation Models

Future clinical language models will likely take the form of large foundation models that are pretrained on multimodal health data from different modalities. Recent efforts like PubMedBERT and GatorTron show that billion-parameter biomedical text models are now possible (Yang *et al.,* 2022).

5.3 Explainable NLP

There is also a growing body of work on explainable NLP, with the goal of making model outputs more interpretable to clinicians (Ghassemi *et al.,* 2021; Ribeiro *et al.,* 2016). Tools that can highlight text phrases that had the most influence on a prediction or provide natural language rationales for predictions are expected to improve clinician trust.

5.4 Multimodal Integration

The integration of EHR text with other data sources like medical imaging, genomics, and wearable sensor data will be key to building more comprehensive patient representations. Deep multimodal models have already begun to show strong results in oncology and critical care (Esteva *et al.,* 2021).

5.5 Workflow Integration

Finally, efforts to integrate NLP into clinical workflows will continue to be an active area of research. Practical applications such as automating documentation, real-time clinical summarization, and contextualized clinical decision support have the potential to both reduce administrative burden and improve care efficiency (Wang *et al.,* 2020; Liu *et al.,* 2021). Workflow integration will also require more seamless EHR integration.

6. CONCLUSION

Natural Language Processing (NLP) has shown great promise in extracting value from the vast amounts of unstructured clinical text present in electronic health records (EHRs). Advancements have been seen in information extraction, predictive modeling, and coding support tasks, particularly with the recent emergence of transformer-based architectures. However, challenges related to data quality, privacy, generalizability, and explainability still need to be addressed. Solutions such as federated learning, domain-specific models, and explainable AI can help overcome these challenges and lead to broader adoption. In the future, NLP is expected to play a crucial role in improving healthcare delivery by supporting efficient, accurate, and patient-centered care.

REFERENCES

- Alsentzer, E., Murphy, J., Boag, W., Weng, W. H., Jin, D., Naumann, T., & McDermott, M. (2019). Publicly available clinical BERT embeddings. Proceedings of the 2nd Clinical Natural Language Processing Workshop, 72–78. https://doi.org/10.18653/v1/W19-1909
- Borjali, A., Magnéli, M., Shin, D., Malchau, H., Muratoglu, O. K., & Varadarajan, K. M. (2021). Natural language processing with deep learning for medical adverse event detection from free-text medical narratives: A case study of detecting total hip replacement dislocation. Computers in biology and medicine, 129, 104140.
- Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. The New England journal of medicine, 378(11), 981.
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
- Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., ... & Socher, R. (2021). Deep learning-enabled medical computer vision. NPJ digital medicine, 4(1), 5.
- Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. The lancet digital health, 3(11), e745-e750.
- Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: Towards better research applications and clinical care. Nature Reviews Genetics, 13(6), 395–405. https://doi.org/10.1038/nrg3208
- Jiang, M., Chen, Y., Liu, M., Rosenbloom, S. T., Mani, S., Denny, J. C., & Xu, H. (2011). A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. Journal of the American Medical Informatics Association, 18(5), 601-606.
- Johnson, A. E., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., ... & Mark, R. G. (2023). MIMIC-IV, a freely accessible electronic health record dataset. Scientific data, 10(1), 1.
- Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234-1240.
- Levin, S., Barnes, S., Toerper, M., Debraine, A., DeAngelo, A., Hamrock, E., ... & Howell, E. (2021). Machine-learning-based hospital discharge predictions can support multidisciplinary rounds and decrease hospital length-of-stay. BMJ Innovations, 7(2).
- Li, I., Sun, W., & Wang, F. (2019). A survey of deep learning approaches for clinical text. Journal of Biomedical Informatics, 96, 103-112. https://doi.org/10.1016/j.jbi.2019.103112
- Liu, W., Wang, H., Shen, X., & Tsang, I. W. (2021). The emerging trends of multi-label learning. IEEE transactions on pattern analysis and machine intelligence, 44(11), 7955-7974.
- Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. (2017). Extracting information from textual documents in the electronic health record: A review of recent research. Yearbook of Medical Informatics, 26(1), 128–144. https://doi.org/10.15265/IY-2017-020
- Murphy, S., Castro, V., & Mandl, K. (2017). Grappling with the future use of big data for translational medicine and clinical care. Yearbook of medical informatics, 26(01), 96-102.
- Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., … Dean, J. (2018). Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine, 1(1), 18. https://doi.org/10.1038/s41746-018-0029-1
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).
- Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ digital medicine, 3(1), 119.

- Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ digital medicine, 3(1), 119.

- Rumshisky, A., Ghassemi, M., Naumann, T., Szolovits, P., Castro, V. M., McCoy, T. H., & Perlis, R. H. (2016). Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. Translational psychiatry, 6(10), e921-e921.
- Sarker, A., Klein, A. Z., Mee, J., Harik, P., & Gonzalez-Hernandez, G. (2019). An interpretable natural language processing system for written medical examination assessment. Journal of biomedical informatics, 98, 103268.
- Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association, 17(5), 507-513.
- Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2017). Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics, 22(5), 1589-1604.
- Shivade, C., Raghavan, P., Fosler-Lussier, E., Embi, P. J., Elhadad, N., Johnson, S. B., & Lai, A. M. (2014). A review of approaches to identifying

patient phenotype cohorts using electronic health records. Journal of the American Medical Informatics Association, 21(2), 221-230.

- Stanfill, M. H., Williams, M., Fenton, S. H., Jenders, R. A., & Hersh, W. R. (2010). A systematic literature review of automated clinical coding and classification systems. Journal of the American Medical Informatics Association, 17(6), 646-651.
- Wang, Y., Afzal, N., Fu, S., Wang, L., Shen, F., Rastegar-Mojarad, M., & Liu, H. (2020). MedSTS: a resource for clinical semantic textual similarity. Language Resources and Evaluation, 54(1), 57-72.
- Wu, S., Roberts, K., Datta, S., Du, J., Ji, Z., Si, Y., … Xu, H. (2020). Deep learning in clinical natural language processing: A methodical review. Journal of the American Medical Informatics Association, 27(3), 457–470. https://doi.org/10.1093/jamia/ocz200
- Yang, X., Chen, A., PourNejatian, N., Shin, H. C., Smith, K. E., Parisien, C., ... & Wu, Y. (2022). Gatortron: A large clinical language model to unlock patient information from unstructured electronic health records. arXiv preprint arXiv:2203.03540.