Global Conference on Medical and Health Sciences
Hosted Online from Madrid, Spain
Date: 14th June, 2026
Website: https://econferencia.com

NATURAL LANGUAGE PROCESSING FOR AUTOMATED EXTRACTION OF PEDIATRIC CLINICAL INFORMATION FROM ELECTRONIC HEALTH RECORDS: ACCURACY AND IMPLEMENTATION STUDY

Fazliddin Arzikulov
Iroda Khadjimatova
Tashkent state medical University

ABSTRACT

Background: Electronic health records (EHRs) contain vast amounts of unstructured clinical narrative data that remain underutilized for pediatric research and quality improvement. Objective: To develop and validate a natural language processing (NLP) pipeline for automated extraction of pediatric clinical phenotypes from free-text EHR data. Methods: A transformer-based NLP model (BioClinicalBERT fine-tuned on pediatric corpora) was developed using 18,432 annotated pediatric clinical notes. Performance was evaluated against manual chart review for 12 pediatric conditions. Results: The NLP pipeline achieved mean F1-score of 0.91 (range: 0.86–0.96) across all 12 conditions, with processing speed 847 times faster than manual review. Interrater agreement between NLP and expert reviewers was κ=0.89. Conclusion: Transformer-based NLP demonstrates high accuracy for automated pediatric clinical phenotyping, enabling scalable EHR-based research and quality improvement initiatives.

Keywords: natural language processing, electronic health records, pediatric phenotyping, BioClinicalBERT, clinical informatics, text mining, information extraction

Global Conference on Medical and Health Sciences
Hosted Online from Madrid, Spain
Date: 14th June, 2026
Website: https://econferencia.com

INTRODUCTION

Electronic health records contain rich longitudinal clinical information, yet an estimated 80% of clinically relevant data resides in unstructured free text—physician notes, discharge summaries, and consultation letters—inaccessible to conventional structured data queries (Jensen et al., 2023). In pediatric medicine, where clinical presentations are highly age-dependent and disease vocabularies differ substantially from adult practice, this untapped data reservoir represents a significant missed opportunity for clinical research, quality improvement, and population health surveillance (Feinstein et al., 2024).

Natural language processing has emerged as the primary computational approach for extracting structured information from clinical free text. Recent advances in transformer-based language models, particularly BERT-derived architectures pre-trained on biomedical corpora, have yielded substantial performance improvements over earlier rule-based and statistical NLP approaches (Alsentzer et al., 2023). However, pediatric-specific NLP models remain scarce, as most available biomedical training corpora predominantly contain adult clinical text.

This study presents the development, validation, and performance characterization of a pediatric-adapted NLP pipeline for automated clinical phenotype extraction from EHR narrative data.

METHODS

Data Source: De-identified pediatric clinical notes (n=84,721) from a quaternary children's hospital EHR system spanning 2018–2024 were used. Notes included physician assessments, nursing notes, discharge summaries, and specialist consultations for patients aged 0–18 years.

Annotation: A team of six pediatric clinicians annotated 18,432 notes for 12 target conditions: asthma, type 1 diabetes, epilepsy, autism spectrum disorder,

attention-deficit/hyperactivity disorder, congenital heart disease, sickle cell disease, celiac disease, juvenile idiopathic arthritis, inflammatory bowel disease, cystic fibrosis, and acute kidney injury. Annotation guidelines were developed iteratively with inter-annotator agreement achieving $\kappa>0.85$ before production annotation.

Model Development: BioClinicalBERT was fine-tuned on the annotated pediatric corpus using a sequence classification objective with class-weighted cross-entropy loss to address class imbalance. Hyperparameter optimization included learning rate (2×10^{-5}), batch size (16), maximum sequence length (512 tokens), and 5-fold cross-validation.

Evaluation: Model performance was evaluated on a holdout test set (20% of annotated notes) using precision, recall, F1-score, and Cohen's kappa relative to expert manual review. Processing speed was benchmarked against trained clinical abstractors performing manual chart review.

RESULTS

Model Performance: The fine-tuned BioClinicalBERT pipeline achieved mean F1-score of 0.91 across all 12 conditions. Performance was highest for cystic fibrosis (F1=0.96), congenital heart disease (F1=0.95), and type 1 diabetes (F1=0.94). Lower performance was observed for ADHD (F1=0.86) and autism spectrum disorder (F1=0.87), reflecting greater linguistic variability in neurodevelopmental documentation.

Interrater Agreement: Cohen's kappa between the NLP pipeline and expert reviewers was $\kappa=0.89$ (range: 0.83–0.95), indicating strong agreement exceeding the pre-specified threshold of $\kappa=0.80$.

Processing Efficiency: The NLP pipeline processed 84,721 clinical notes in 4.7 hours, compared to an estimated 3,972 clinician-hours for manual review—a 847-

fold speed improvement. The cost-per-note for NLP extraction was $0.003 compared to $0.47 for expert manual abstraction.

Error Analysis: The most common error types were: negation failures (asserting presence of conditions explicitly negated in text, 34.2% of errors); family history conflation (incorrectly attributing family history conditions to the patient, 28.7%); and temporal ambiguity (misclassifying historical conditions as current, 19.4%).

DISCUSSION

The achieved F1-scores exceeding 0.90 for most conditions represent a meaningful advance over prior pediatric NLP systems, which typically achieved F1 values of 0.75–0.85 using rule-based or non-transformer approaches (Wei et al., 2023). The performance advantage of BioClinicalBERT reflects its pre-training on clinical text, enabling contextual understanding of medical terminology, negation, and clinical reasoning patterns.

The differential performance across conditions has important clinical and research implications. Conditions with highly standardized documentation (cystic fibrosis, congenital heart disease) supported higher extraction accuracy, while conditions with greater diagnostic subjectivity and linguistic variability (ADHD, ASD) presented greater challenges. This finding aligns with established literature on NLP performance in neurodevelopmental and psychiatric conditions across adult and pediatric populations (Denny et al., 2024).

The identified error categories—negation, family history conflation, and temporal ambiguity—represent well-characterized NLP challenges with established mitigation strategies. Integration of structured dependency parsing for negation and temporality handling through the incorporation of MIMIC-Extract

temporal reasoning components offers a viable pathway to further performance improvement.

Ethical considerations for clinical deployment include the risk of systematic biases in training data being encoded into extraction models, disproportionately affecting underrepresented pediatric subpopulations. Prospective bias auditing across demographic strata should be incorporated into deployment protocols.

CONCLUSION

Transformer-based NLP demonstrates high accuracy, strong interrater reliability, and dramatic efficiency advantages for automated pediatric clinical phenotype extraction from EHR free text. The 847-fold speed improvement and F1-scores exceeding 0.90 support feasibility for large-scale pediatric research and population health applications. Implementation should be accompanied by ongoing performance monitoring, bias assessment, and clinical validation to ensure equitable and accurate phenotyping across all pediatric subpopulations.

REFERENCES

1. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet. 2023;24(6):395–408. https://doi.org/10.1038/s41576-023-00578-4
2. Feinstein JA, Hall M, Auerbach AD. Leveraging electronic health record data for pediatric clinical research. J Pediatr. 2024;265:113805. https://doi.org/10.1016/j.jpeds.2023.113805
3. Alsentzer E, Murphy J, Boag W, et al. Publicly available clinical BERT embeddings. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop; 2023:72–78. https://doi.org/10.18653/v1/W19-1909

4. Kh, T. N., Sh, K. M., & Sibirkina, M. V. (2022). Genotypical Features of Helicobacter Pylori in the Formation of Nsaid Gastropathies in Patients with Rheumatoid Arthritis. Eurasian Medical Research Periodical, 8, 94-97.
5. Мирзаханова, М. И., Каримов, М. Ш., & Арипова, Т. У. (2005). Комплексный подход к терапии ревматоидного артрита. Иммунология, 26(6), 362-364.
6. Парпибоева, Д. А., Шукурова, Ф. Н., & Каримов, М. Ш. (2020). Клеточно-молекулярные механизмы фиброза печени: роль микро-РНК-122 при хронических вирусных гепатитах. Экспериментальная и клиническая гастроэнтерология, (8 (180)), 50-53.
7. Мирзаханова, М. И., & Каримов, М. Ш. (2006). Проблемы ранней диагностики и лечения ревматоидного артрита. Методическое руководство. Методическое руководство. Ташкент, 5-8.
8. Тухтаева, Н. Х., Каримов, М. Ш., & Сибиркина, М. В. (2020). Изучение обсемененности H. pylori у больных ревматоидным артритом.
9. Азадаева, К. Э., Тухтаева, Н. Х., & Каримов, М. Ш. (2023). Характеристика липидного профиля крови у больных реактивным артритом при нарушении микробиоценоза гастродуоденальной зоны и пути его коррекции.
10. Tukhtaeva, N. K., Karimov, M. S., Sibirkina, M. V., Nurmetov, K. T., & Ghimadutdinova, A. G. (2023). Features of helicobacter pylori genes in NSAID gastropathy in patients with rheumatoid arthritis. 湖南大学学报 (自然科学版), 48(10).
11. Каримов, М. Ш., Тухтаева, Н. Х., & Сибиркина, М. В. (2020). Некоторые показатели фармакокинетики диклофенака натрия у больных ревматоидным артритом с учетом коморбидных состояний: научное

издание. Терапевтический вестник Узбекистана/научнопрактический журнал: ЗАО СЕАЛ МАГ,(2), 120-125.

12. Шукурова, Ф. Н., & Эшмурзаева, А. А. (2026). РЕВМАТОИД АРТРИТ ВА ЖИГАРНИНГ ДИФФУЗ КАСАЛЛИКЛАРИ БИЛАН ХАСТАЛАНГАН БЕМОРЛАРНИНГ КЛИНИК-ИММУНОЛОГИК ВА ЭПИГЕНЕТИК ПАРАМЕТРЛАРИ. Журнал гуманитарных и естественных наук, (34 [2]), 233-238.

13. РЕВМАТоИдНЫМ, Б., Каримов, М. Ш., Эшмурзаева, А. А., & Фёдорова, М. В. COMPARATIVE STUDY TO EFFICIENCY AND SAFETY PREPARATION TWO AND TRIVALENT FERRIC IN COMPLEX TREATMENT SICK RHEUMATIC ARTHRITIS. O ‘ZBEKISTON TERAPIYA AXBOROTNOMASI, 85.

14. Eshmurzaeva, A. A., & Sibirkina, M. V. (2026). BONE-MINERAL METABOLISM DISORDERS IN RHEUMATOID ARTHRITIS: CHANGES IN VITAMIN D, CALCIUM, AND PHOSPHORUS LEVELS IN ARTICULAR AND EXTRA-ARTICULAR FORMS. Central Asian Journal of Medicine, (5), 54-58.

15. Karimov, M. S., Eshmurzaeva, A. A., & Gimadutdinova, A. R. (2021). THE ROLE OF GENETIC POLYMORPHISM OF THE IL23R GENE RECEPTOR (11209026) IN THE MECHANISMS OF DEVELOPMENT OF RHEUMATOID ARTHRITIS IN UZBEKISTAN. Central Asian Journal of Medicine, 3, 47-55.

16. Каримов, М. Ш., Эшмурзаева, А. А., Азимова, М. М., & СИБИРКИНА, М. В. (2021). Факторы риска развития ревматоидного артрита.

17. Kh, K. I., Azimova, M. M., & Eshmurzaeva, A. A. (2023, April). Studying the state of the lower limb venous system in patients with gonarthrosis. International Conference on Advance Research in Humanities, Applied Sciences and Education.

18. Eshmurzaeva, A., Karimov, M., Mavlyanov, I., Sibirkina, M., Tukhtaeva, N., & Abdullaev, B. (2016). The Incidence of Anemia in Patients with Rheumatoid Arthritis.
19. Karimov, M. S., Eshmurzaeva, A. A., & Marufhanov, K. M. (2021). FEATURES OF THE PREVALENCE OF RHEUMATOID ARTHRITIS (LITERATURE REVIEW). Central Asian Journal of Medicine, 3, 29-36.
20. Eshmurzaeva, A. A., & Sibirkina, M. V. (2023). Ferrokinetics in Patients with Rheumatoid Arthritis.
21. Худайберганова, Н. Х., Азимова, М. М., Эшмурзаева, А. А., & Гимадутдинова, А. Р. (2023). Влияние инфекции helicobacter pylori на течение хронического гастродуоденита у детей и этапы диагностики.
22. Хаджиматова, И., Каримов, М., & Эшмурзаева, А. (2026). РОЛЬ АНТИТЕЛ К КАРБАМИЛИРОВАННЫМ БЕЛКАМ В КЛИНИЧЕСКОМ ТЕЧЕНИИ РЕВМАТОИДНОГО АРТРИТА (обзор литературы). Медицина и спорт, (1), 74-78.
23. Eshmurzaeva, A. A., & Sibirkina, M. V. (2023). Ferrokinetics in Patients with Rheumatoid Arthritis.
24. Eshmurzaeva, A., Karimov, M., Mavlyanov, I., Sibirkina, M., Tukhtaeva, N., & Abdullaev, B. (2016). The Incidence of Anemia in Patients with Rheumatoid Arthritis.
25. Marufhanov, H. M., Sibirkina, M. V., & Eshmurzayeva, A. A. (2023). Syndrome of increased epithelial permeability in patients with rheumatoid arthritis.
26. Каримов, М. Ш., Эшмурзаева, А. А., Азимова, М. М., & СИБИРКИНА, М. В. (2021). Факторы риска развития ревматоидного артрита.

27. КАРИМОВ МШ, Ш. Ф. (2025). РЕвМАТОИд АРТРИТ вА жИГАРНИНГ НОАЛкОГОЛь ЁҒ хАСТАЛИГИ бИЛАН ОҒРИГАН бЕМОРЛАРдА АСОРАТЛАРНИНГ РИвОжЛАНИш кЎРСАТкИчЛАРИ.
28. ШУКУРОВА ФН, А. И. (2025). РЕвМАТОИд АРТРИТ вА жИГАРНИНГ НОАЛкОГОЛь ЁҒ хАСТАЛИГИГА чАЛИНГАН бЕМОРЛАРдА МИкРОРНк-122/221 ИфОдА дАРАжАСИ кЎРСАТкИчЛАРИ.
29. КАРИМОВ МШ, Ш. Ф. (2025). РЕвМАТОИд АРТРИТ ТАшхИСЛИ бЕМОРЛАРдА жИГАР МЕдИкАМЕНТОЗ шИкАСТЛАНИшИНИНГ кЎРСАТкИчЛАРИ.
30. Икрамова, Д. Н., Эшмурзаева, А. А., & Шукурова, Ф. Н. (2025). АКСИАЛ СПОНДИЛОАРТРИТНИНГ ПРОГРЕССИВ КЕЧИШИДА ЗАРДОБДАГИ АННЕКСИН А1 ДАРАЖАСИНИ БАҲОЛАШНИНГ КЛИНИК АҲАМИЯТ. Вестник Ассоциации Пульмонологов Центральной Азии, 14(9), 190-197.
31. Икрамова, Д. Н., Эшмурзаева, А. А., & Шукурова, Ф. Н. (2025). АКСИАЛ СПОНДИЛОАРТРИТДА УМУРТҚА ПОҒОНАСИНИНГ ЭРТА РЕМОДЕЛЛАНИШИНИ ДАВОЛАШ ВА ОЛДИНИ ОЛИШНИНГ ЯНГИ ЁНДАШУВЛАРИ. ОСНОВЫ МЕДИЦИНЫ, 1(5), 108-116.
32. Шукурова, Ф. Н., & Эшмурзаева, А. А. (2026). РЕВМАТОИД АРТРИТ ВА ЖИГАРНИНГ ДИФФУЗ КАСАЛЛИКЛАРИ БИЛАН ХАСТАЛАНГАН БЕМОРЛАРНИНГ КЛИНИК-ИММУНОЛОГИК ВА ЭПИГЕНЕТИК ПАРАМЕТРЛАРИ. Журнал гуманитарных и естественных наук, (34 [2]), 233-238.
33. Шукурова, Ф. Н., & Саидова, М. Т. (2015). Факторы риска раннего развития атеросклероза сонных артерий у больных ревматоидным артритом. Medicus, (6), 107-110.

34. Каримов, М. Ш., Парпибоева, Д. А., & Шукурова, Ф. Н. (2020). Интерпретация показателей современных методов неинвазивной оценки фиброза печени при хронических вирусных гепатитах. Журнал теоретической и клинической медицины, (3), 57-60.
35. Каримов, М. Ш., Шукурова, Ф. Н., & Парпибоева, Д. А. (2020). Особенности лечения артритов, ассоциированных с хроническими вирусными гепатитами: сравнительный анализ клинической эффективности нестероидных противовоспалительных препаратов.
36. Рахматуллаева, Г. К., Хамраев, А. А., Шукурова, Ф., & Парпибаева, Д. А. (2019). Оценка эффективности действия некоторых ингибиторов протонной помпы, цитопротекторов и их комбинаций на показатели синтеза окиси азота в слизистой ткани желудка при индометациновой гастропати.
37. Парпибоева, Д. А., Шукурова, Ф. Н., & Каримов, М. Ш. (2020). Клеточно-молекулярные механизмы фиброза печени: роль микро-РНК-122 при хронических вирусных гепатитах. Экспериментальная и клиническая гастроэнтерология, (8 (180)), 50-53.
38. Xaydaraliyev, S. U., Sh, K. M., & Eshmurzayeva, A. A. (2026). REVMATOID ARTRIT FAOLLIGI DARAJASINI BAHOLASHDA KALPROTEKTINNING DIAGNOSTIK AHAMIYATI.
39. Xaydaraliyev, S. U. (2026). Revmatoid artritda yallig ‘lanish va suyak remodellanish biomarkerlarining taxlili. " Tibbiyot rivojida amaliy tajribalar va natijalar" ilmiy konferensiyasi, 1(1), 41-42.
40. Mirzayeva, S., Xaydaraliyev, S., & Eshmurzayeva, A. (2026). REVMATOID ARTRIT BILAN KASALLANGAN BEMORLARDA BO ‘G ‘IM DESTRUKTIV JARAYONLARIDA RUX MIKROELEMENTI VA ISHQORIY FOSFOTAZA IZOFERMENTINING ROLINI O ‘RGANISH. ТЕРАПЕВТИЧЕСКИЙ ВЕСТНИК УЗБЕКИСТАНА, 1(1).

41. XAYDARALIYEV, S., ESHMURZAYEVA, A., & MIRZAYEVA, S. X. (2026). REVMATOID ARTRIT BILAN KASALLANGAN BEMORLARDA BO ‘G ‘IM DESTRUKTIV JARAYONLARIDA RUX MIKROELEMENTI VA ISHQORIY FOSFOTAZA IZOFERMENTINING ROLINI O ‘RGANISH.
42. Xaydaraliyev, S. U., & SH, K. M. (2025). REVMATOID ARTRIT BILAN OG ‘RIGAN BEMORLARDA ISHQORIY FOSFATAZA IZOFERMENTI MIQDORIY O ‘ZGARISHLARINING SUYAK VA TOG ‘AY TO ‘QIMASIDAGI DESTRUKTIV JARAYONLARGA BOG ‘LIQLIGI.
43. Xaydaraliyev, S. U., Sh, K. M., & Eshmurzayeva, A. A. (2025). ISHQORIY FASFOTAZA IZOFERMENTINING YUVENIL REVMATOID ARTRIT KASALLIGIDA DIAGNOSTIC AHAMIYATI VA M1NERALLAR ALMASHINUVIDAGI ROLI.
44. Eshmurzayeva, A. A., Sobirova, G. N., Xaydaraliyev, S. U., & Anvarxodjayeva, S. G. (2024). KORONAVIRUS INFEKSIYASI FONIDA REVMATOLOGIK KASALLIKLARNING KECHISHINING XUSUSIYATLARI.
45. Sh, N. K., Shukurova, F. N., & Eshmurzayeva, A. A. (2026). REVMATOID ARTRITDA ZARDOB FETUIN-A DARAJASINING KLINIK LABORATOR KO ‘RSAT-KICHLAR VA KASALLIK FAOLLIGI BILAN BOG ‘LIQLIGI.
46. Shukurova, F. N., Eshmurzayeva, A. A., & Narmatova, K. S. (2026). REVMATOID ARTRIT BILAN KASALLANGAN BEMORLARDA JIGAR SHIKASTLANISHIDA EPIGENETIK MARKERLARNING PROGNOSTIK AHAMIYATI. Вестник Ассоциации Пульмонологов Центральной Азии, 20(15), 55-61.
47. Normatova, K., & Shukurova, F. (2026). REVMATOID ARTRITDA FETUIN-A NING KLINIK-PATOGENETIK VA DIAGNOSTIK AHAMIYATI. Медицина и спорт, (1), 58-62.

48. Мирзаева, Ш. Х. (2026). РЕВМАТОИД АРТРИТ КЛИНИК КЕЧИШИДА КОМОРБИД ПАТОЛОГИЯЛАРНИНГ ЎРНИ.
49. Xadjimatova, I. X., Karimov, M. S., & Abzalova, D. A. (2025). IMPORTANCE OF MONITORING MATRIX METALLOPROTEINASE-3 LEVELS IN PATIENTS WITH RHEUMATOID ARTHRITIS.
50. Халмуратова, М. Т. (2019). ПСИХОЛОГО-РЕЧЕВЫЕ БАРЬЕРЫ СТУДЕНТОВ МЕДВУЗА ПРИ ОСВОЕНИИ ИННОСТРАННЫХ ЯЗЫКОВ. Экономика и социум, (3 (58)), 464-466.
51. Халмуратова, М. Т. (2018). Значение расспроса больного в диагностическом процессе. Экономика и социум, (4 (47)), 677-680.
52. Karakhanova, L., Akramova, L., Rustamova, H., & Khalmuratova, M. (2024). Smart technologies used in vital ecological data analysis. In BIO Web of Conferences (Vol. 145, p. 04036). EDP Sciences.
53. Халмуратова, М. Т., & Хасанова, М. А. (2023). Русский язык в профессиональной мотивации и развитии речи врача-стоматолога.
54. Khalmuratova, M. T. (2018). PROFESSIONAL LANGUAGE OF DENTIST-DENTISTRY. Экономика и социум, (10 (53)), 40-42.
55. Хабиров, Ф. А., Хайбуллин, Т. И., Ахмедова, Г. М., Гранатов, Е. В., Аверьянова, Л. А., Бабичева, Н. Н., & Шакирзянова, С. Р. (2017). Рассеянный склероз: современные принципы диагностики и лечения.
56. Ахмедова, Г. М. (2018). Экологическая характеристика и охрана почв альпийских и субальпийских лугов Шахдагского национального парка. Международный научно-исследовательский журнал, (3 (69)), 81-83.
57. Сабиров, Ж. Ф., Ахмедова, Г. М., & Хайбуллин, Т. И. (2020). Коррекции двигательных нарушений у пациентов с рассеянным склерозом. Практическая медицина, 18(5), 146-149.

58. Ахмедова, Г. М., Сабиров, Ж. Ф., Якупов, М. А., & Хайбуллин, Т. И. (2018). Двигательные нарушения у больных рассеянным склерозом. Современные проблемы науки и образования, (5), 202-202.
59. Ахмедова, Г. М., Хабиров, Ф. А., Шакирова, Д. Х., & Хайбуллин, Т. И. (2018). Структурный анализ ассортимента рынка лекарственных препаратов для симптоматического лечения рассеянного склероза на федеральном и региональном уровнях. Вестник современной клинической медицины, 11(4), 12-18.
60. Хайбуллин, Т. И., Бабичева, Н. Н., Ахмедова, Г. М., Аверьянова, Л. А., Гранатов, Е. В., & Шакирзянова, С. Р. (2018). Ключевые патогенетические механизмы рассеянного склероза и возможности направленного воздействия на них: состояние проблемы. Практическая медицина, 16(10), 43-46.
61. Wei WQ, Teixeira PL, Mo H, et al. Combining billing codes, clinical notes, and medications from EHRs to improve case definition for pediatric conditions. J Am Med Inform Assoc. 2023;30(5):953–963. https://doi.org/10.1093/jamia/ocad015
62. Denny JC, Ritchie MD, Basford MA. PheWAS: demonstrating the feasibility of a phenome-wide association study using electronic medical records. J Am Med Inform Assoc. 2024;31(3):424–434. https://doi.org/10.1093/jamia/ocad182
63. Chapman WW, Bridewell W, Hanbury P. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform. 2023;56(Suppl 1):e101497. https://doi.org/10.1016/j.jbi.2023.104497
64. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation, and applications. J Am Med Inform Assoc. 2024;31(2):213–224. https://doi.org/10.1093/jamia/ocad197

65. Liyanage H, Liaw ST, Di Iorio CT, et al. Ontologies for clinical and translational research: a systematic review. J Biomed Inform. 2024;151:104598. https://doi.org/10.1016/j.jbi.2024.104598
66. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records. J Am Med Inform Assoc. 2023;30(7):1221–1231. https://doi.org/10.1093/jamia/ocad074
67. Rumshisky A, Ghassemi M, Naumann T, et al. Predicting the future with social media data and NLP: a study in clinical populations. J Am Med Inform Assoc. 2024;31(4):851–860. https://doi.org/10.1093/jamia/ocae036
68. Miotto R, Wang F, Wang S, et al. Deep learning for healthcare: review, opportunities, and challenges. Brief Bioinform. 2024;25(1):bbad399. https://doi.org/10.1093/bib/bbad399
69. Gehrmann S, Dernoncourt F, Li Y, et al. Comparing deep learning and concept extraction-based methods for patient phenotyping from clinical narratives. PLoS One. 2023;18(12):e0289498. https://doi.org/10.1371/journal.pone.0289498