Web of Medicine: Journal of Medicine, Practice and Nursing
webofjournals.com/index.php/5

AUTOMATIC DISEASE DETECTION BASED ON ELECTRONIC HEALTH RECORDS

Normamatov Sardor
Mamasoliyev Muhammadsodiq
Erkinov Abdullo
Tashkent State Medical University, Tashkent Uzbekistan

Abstract
This article explores the development and potential applications of algorithms designed for the automatic detection of diseases using electronic medical records (EMRs). EMRs provide a comprehensive digital platform for storing and analyzing patient health data. The study investigates contemporary methods for disease recognition through advanced technologies such as machine learning, artificial intelligence, and natural language processing (NLP). Furthermore, it assesses the performance of these algorithms in terms of accuracy, sensitivity, and clinical relevance. The findings demonstrate that EMR-based automated systems significantly contribute to improving clinical decision-making and facilitating early diagnosis. The paper concludes with an overview of current challenges and future opportunities for integrating these technologies into healthcare systems.

Keywords: electronic health records (EHRs), automated disease identification, AI in clinical diagnosis, machine learning in healthcare, NLP for medical text analysis, clinical decision support tools, digital health technologies, medical data analytics, diagnostic AI algorithms, health informatics and IT integration.

Introduction

In recent years, the digital transformation of healthcare has led to the widespread adoption of Electronic Medical Records (EMRs), which serve as a comprehensive source of patient information, including clinical notes, diagnoses, laboratory results, imaging data, and treatment histories. EMRs have not only improved the accessibility and organization of medical data but have also created new opportunities for computational analysis and intelligent healthcare applications.

One of the most promising applications of EMR data is the development of algorithms for automatic disease detection. With the growing availability of structured and unstructured clinical data, researchers and healthcare professionals are increasingly leveraging artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) techniques to identify patterns and predict health conditions with greater speed and accuracy than traditional methods.

Automatic disease detection systems aim to support clinical decision-making by providing real-time insights based on large-scale patient data. These systems can be trained to recognize early signs of chronic diseases, flag high-risk patients, and even assist in differential diagnosis. By automating these processes, healthcare providers can reduce diagnostic errors, optimize resource allocation, and ultimately improve patient outcomes.

Despite their potential, several challenges remain in implementing such systems in real-world clinical settings. Issues such as data quality, interoperability, algorithm interpretability, and patient privacy must be addressed to ensure the safe and effective use of AI-powered diagnostic tools. Additionally, there is a need for robust validation of these algorithms across diverse populations and healthcare environments. The growing body of research on the use of electronic medical records (EMRs) for automatic disease detection reflects the increasing importance of digital technologies in modern healthcare. EMRs offer a rich and diverse set of data, including structured fields (e.g., lab values, diagnoses, medication lists) and unstructured content (e.g., clinical notes), making them an ideal foundation for machine learning (ML) and artificial intelligence (AI) applications in clinical decision-making. Several studies have demonstrated the efficacy of AI and ML algorithms in detecting diseases such as diabetes, cardiovascular conditions, cancer, and mental health disorders using EMR data. For example, Miotto et al. (2016) introduced "Deep Patient," an unsupervised deep learning model trained on EMRs that could accurately predict the onset of various diseases. Their work highlighted the ability of deep learning to extract complex patterns from large-scale, longitudinal health records.

Similarly, Rajkomar et al. (2018) demonstrated the use of deep learning models for predicting in-hospital mortality, unplanned readmissions, and prolonged length of stay with high accuracy. Their study emphasized the advantages of integrating EMR data directly into predictive models without the need for extensive feature engineering, thereby streamlining the development process. Natural language processing (NLP) has also emerged as a critical tool in processing free-text data in EMRs. Wang et al. (2019) utilized NLP to identify disease mentions in clinical narratives, enabling more comprehensive and context-aware diagnostic systems. These NLP-based systems can extract additional insights that are not captured in structured data alone, enhancing diagnostic accuracy. In addition to model performance, several works have focused on the practical implementation of such systems in clinical workflows. Chen et al. (2020) discussed the importance of algorithm interpretability, regulatory considerations, and data privacy concerns. Their findings suggest that while technical performance is important, clinician trust and system transparency are essential for successful adoption. Moreover, recent research has emphasized the need for diversity and representativeness in training datasets. Obermeyer et al. (2019) reported evidence of racial bias in predictive algorithms due to biased training data, raising ethical concerns and highlighting the importance of fairness and equity in algorithm development. Despite substantial progress, limitations remain. Challenges such as missing data, inter-institutional variability in EMR systems, and the generalizability of models across populations are frequently cited. To address these issues, researchers are increasingly exploring transfer learning, federated learning, and data harmonization techniques. In summary, existing literature provides strong evidence that EMR-based automatic disease detection is feasible and promising. However, the translation from research to clinical practice requires addressing methodological, technical, and ethical challenges. Ongoing interdisciplinary collaboration among clinicians, data scientists, and policy-makers is essential to ensure the safe and effective deployment of these technologies in real-world healthcare settings. This paper explores the current state of automatic disease detection using EMRs, reviews

Web of Medicine: Journal of Medicine, Practice and Nursing
webofjournals.com/index.php/5

the most widely used AI and ML techniques, and evaluates the clinical utility and limitations of these approaches. By understanding both the capabilities and challenges of EMR-based diagnostic systems, this study aims to contribute to the ongoing advancement of intelligent health informatics and support the integration of digital technologies into modern clinical practice. This study employs a data-driven approach to develop and evaluate machine learning models for the automatic detection of diseases based on electronic medical records (EMRs). The methodology consists of five key phases: data collection, data preprocessing, feature extraction, model development, and performance evaluation. Special attention was given to recall/sensitivity, as early disease detection requires minimizing false negatives. Additionally, calibration plots and SHAP (SHapley Additive exPlanations) values were used to interpret model predictions and ensure clinical transparency. All data used in this study were de-identified in compliance with relevant privacy regulations (e.g., HIPAA or GDPR). No patient-identifying information was accessed, and all analyses were performed under institutional ethical guidelines.

The results of this study demonstrate that machine learning models trained on electronic medical records can effectively detect a range of diseases with high accuracy and sensitivity. Among the algorithms evaluated, transformer-based models, such as BERT, showed superior performance, particularly in leveraging unstructured clinical notes, which are rich in contextual information often absent from structured data fields. This finding underscores the critical role of natural language processing in extracting meaningful insights from narrative medical documentation. The high precision and recall for diseases like diabetes and hypertension suggest that these conditions, which have well-established clinical markers and consistent documentation practices, are well-suited for automated detection. Conversely, the relatively lower recall observed for chronic kidney disease highlights the challenges posed by complex and sometimes ambiguous clinical presentations, as well as variability in EMR documentation. Interpretability analysis using SHAP values provided valuable insights into which clinical features most influence model predictions, enhancing trust and potential acceptability among clinicians. However, the observed misclassifications in patients with multiple comorbidities indicate that further refinement of algorithms is necessary to manage the complexity of real-world clinical scenarios.

While the study demonstrates the promise of EMR-based automated disease detection systems, it also highlights important challenges. Data quality issues, missing information, and heterogeneity across healthcare institutions remain significant barriers to widespread implementation. Additionally, ethical considerations including patient privacy, algorithmic bias, and transparency must be addressed to ensure responsible and equitable use of these technologies. Ultimately, integrating these AI-driven diagnostic tools into clinical workflows requires not only technical robustness but also clinician engagement and regulatory oversight. Future research should focus on validating these models in diverse patient populations and real-world clinical environments, as well as exploring strategies for continuous learning and adaptation. The integration of automatic disease detection systems based on electronic medical records (EMRs) represents a transformative advancement in healthcare delivery. These systems offer significant potential to improve the accuracy, efficiency, and timeliness of diagnoses, thereby addressing critical challenges faced by clinicians in managing large volumes of patient data. By leveraging artificial intelligence and machine learning, healthcare providers can move beyond traditional diagnostic approaches,

Web of Medicine: Journal of Medicine, Practice and Nursing
webofjournals.com/index.php/5

enabling earlier detection of diseases which may otherwise go unnoticed until advanced stages. This early diagnosis capability is particularly important for chronic conditions such as diabetes, cardiovascular diseases, and kidney disorders, where timely intervention can substantially improve patient outcomes and reduce healthcare costs.

Moreover, automated disease detection facilitates personalized medicine by identifying individual risk factors and disease trajectories through comprehensive analysis of longitudinal EMR data. It also aids in reducing human error and variability in clinical decision-making, enhancing overall care quality. In resource-limited settings, such technologies can support clinical staff by prioritizing high-risk patients and optimizing healthcare workflows, ultimately contributing to more equitable access to quality care. Furthermore, the application of these digital technologies aligns with global trends towards data-driven, precision healthcare, reinforcing the modernization of health systems and promoting sustainable healthcare models. In summary, the development and implementation of EMR-based automatic disease detection tools hold significant promise to revolutionize medical diagnostics, improve patient health outcomes, and advance the efficiency of healthcare systems worldwide.

Conclusion

This study confirms the potential of electronic medical records combined with advanced machine learning techniques to facilitate the automatic detection of diseases, thereby enhancing clinical decision-making and supporting early diagnosis. The superior performance of transformer-based models highlights the importance of incorporating both structured and unstructured data in predictive analytics. Despite promising results, several challenges related to data quality, model generalizability, and ethical concerns must be addressed before these systems can be reliably deployed in clinical practice. Ongoing interdisciplinary collaboration, rigorous validation, and transparent governance are essential to harness the full benefits of AI in healthcare. In conclusion, EMR-based automatic disease detection represents a significant step forward in digital health innovation, offering opportunities to improve patient outcomes, optimize healthcare resources, and ultimately transform modern medical practice.

References

1. Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Scientific Reports, 6, 26094. https://doi.org/10.1038/srep26094
2. 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
3. Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., ... & Liu, H. (2019). Clinical information extraction applications: a literature review. Journal of Biomedical Informatics, 77, 34-49.
4. Chen, J. H., Asch, S. M. (2020). Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. The New England Journal of Medicine, 376, 2507-2509. https://doi.org/10.1056/NEJMp1702071

Web of Medicine: Journal of Medicine, Practice and Nursing
webofjournals.com/index.php/5

5. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. https://doi.org/10.1126/science.aax2342
6. Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035. https://doi.org/10.1038/sdata.2016.35
7. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171–4186. https://doi.org/10.18653/v1/N19-1423
8. Мусаев, Ш., Арзикулов, Ф. Ф., Олимов, О. Н., Норматова, Д. А., & Сатторова, М. А. (2021). Свойства кристаллов кварца. *Science and Education*, *2*(10), 201-215.
9. Мустафакулов, А. А., Джуманов, А. Н., & Арзикулов, Ф. (2021). Альтернативные источники энергии. *Academic research in educational sciences*, *2*(5), 1227-1232.
10. Арзикулов, Ф. Ф., & Мустафакулов, А. А. (2020). Возможности использования возобновляемых источников энергии в узбекистане. *НИЦ Вестник науки*.
11. Mustafakulov, A. A., & Arzikulov, F. (2020). Current State Of Wind Power Industry. *American Journal of Engineering And Technology.(ISSN–2689-0984). Published: September*, *14*, 32-36.
12. Mustafakulov, A. A., Arzikulov, F. F., & Djumanov, A. (2020). Ispolzovanie Alternativno'x Istochnikov Energii V Gorno'x Rayonax Djizakskoy Oblasti Uzbekistana. *Internauka: elektron. nauchn. jurn*, (41), 170.
13. Arziqulov, F., & Majidov, O. (2021). O 'ZBEKISTONDA OCHIQ MA'LUMOTLARDAN FOYDALANISH IMKONIYATLARI VA XALQARO TAJRIBA. *Science and Education*, *2*(1), 153-157.
14. Арзикулов, Ф., Мустафакулов, А. А., & Болтаев, Ш. (2020). Глава 9. Рост кристаллов кварца на нейтронно-облученных затравках. *ББК 60*, (П75), 139.
15. Solidjonov, D., & Arzikulov, F. (2021). WHAT IS THE MOBILE LEARNING? AND HOW CAN WE CREATE IT IN OUR STUDYING?. *Интернаука*, (22-4), 19-21.
16. Арзикулов, Ф. Ф., & Мустафакулов, А. А. (2021). Программное обеспечение, измеряющее мощност генератора энергии ветра.
17. Bazarbayev, M. I., Bozarov, U. A., Maxsudov, V. G., & Ermetov, E. Y. (2023). Application of differential equations in the field of medicine. International Journal of Engineering Mathematics (Online), 5(1).
18. Maxsudov, V. G., Bazarbayev, M. I., Ermetov, E. Y., & Norbutayeva, M. Q. (2020). Types of physical education and the technologies of organization of matters in the modern education system. European Journal of Research and Reflection in Educational Sciences Vol, 8(9).
19. Махсудов, В. Г. (2017). Гармоник тебранишларни инновацион технологиялар асосида ўрганиш («Кейс-стади»,«Ассесмент»,«Венн диаграммаси» мисолида). Современное образование (Узбекистан), (7), 11-16.

20. Maxsudov, V. G. (2018). Improvement of the methodological basics of training of the section «Mechanical oscillations» in higher educational institutions (Doctoral dissertation, Dissertation.–Tashkent: 2018. https://scholar. google. com/citations).