World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

# ARTIFICIAL INTELLIGENCE APPLICATIONS IN CLINICAL DATA SYSTEMS: A COMPARATIVE EMPIRICAL ANALYSIS OF CURRENT APPROACHES AND FUTURE DIRECTIONS

Kirankumar Thota * and Manas Kumar Mohanty

Independent Researcher, USA.
* Corresponding Author.

World Journal of Advanced Research and Reviews, 2023, 19(03), 1745–1752

Publication history: Received on 20 August 2023; revised on 27 September 2023; accepted on 30 September 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.19.3.2017



## ABSTRACT

Artificial Intelligence (AI) is now being integrated into clinical data systems, transforming the capture, analysis, and action on health information at the point of care. This paper presents an empirical, comparative study on the use of AI methodologies for structured and unstructured clinical data, summarizing quantitative evidence of performance from 148 peer-reviewed publications spanning the years 2015–2022. The six major families of techniques were identified: supervised machine learning, deep neural networks, natural language processing, computer vision, reinforcement learning, and federated learning, and comparative analyses of the diagnostic performance of these techniques were performed for four representative clinical tasks: risk stratification, diagnostic classification, clinical note summarization, and readmission prediction. The results show that the transformer-based natural language processing models and the convolutional deep networks consistently perform better than the traditional machine learning baselines, with mean accuracies of 90.4% and 88.5%, respectively, versus 77.6%-81.9% for logistic regression, random forest, and support vector machine approaches. The analysis also highlights existing challenges for clinical translation, such as data interoperability constraints, algorithmic bias, lack of model interpretation and unresolved regulatory validation. In light of these findings, the paper calls for a translational approach that considers federated and privacy-preserving learning, standardized interoperability protocols and integration of explainable-AI as key priorities for the next generation of clinical data systems. The results provide evidence-based recommendations for informaticians, health system leaders, and policy makers who are looking to scale carefully AI-powered clinical data systems.

**Keywords:** Artificial Intelligence; Clinical Data Systems; Electronic Health Records; Machine Learning; Deep Learning; Natural Language Processing; Clinical Decision Support.

## 1 INTRODUCTION

In the last twenty years, the implementation of the digitization of healthcare has started to produce an unprecedented amount of clinical data, including structured data, lab values, imaging archives, and unstructured clinical narratives. Electronic health record (EHR) systems are now close to ubiquitous in hospital networks in high-income countries; historically used for administrative and billing purposes, EHRs have become rich longitudinal collections of patient data (Shickel et al., 2018). This increase in data volume and complexity has led to the creation of an opportunity and necessity for more advanced computational techniques.

Artificial intelligence (AI) including machine learning, deep learning, and natural language processing (NLP), has become a key technology for deriving actionable insight from clinical data systems. Today, AI-powered tools can be used for a variety of tasks, such as predictive risk stratification, automated documentation, interpretation of diagnostic images, and point-of-care (POC) clinical decision support (Rajkomar et al., 2018; Topol, 2019). Initial findings indicate that these systems can help diagnose with greater accuracy, lower clinician workload and detect clinical deterioration earlier (Esteva et al., 2019). But the uptake of research models in to reliable, fair, and trusted clinical tools is still uneven, and there are significant gaps between the results of algorithms and their utility in the clinic.

Many of the current publications related to this topic are in the form of narrative or systematic reviews of individual applications. Yet, there is less work that has been done so far on quantifying and comparing the performance of various families of AI techniques on shared clinical data tasks and identifying which structural obstacles most often block these techniques' deployment. A study design of an empirical, comparative-analytical approach has been used to fill that gap.

The specific aims of this research are threefold: (1) to characterize the families of AI techniques currently used in clinical data systems and identify their roles; (2) to quantitatively compare the diagnostic and predictive capabilities of the families of techniques across representative clinical tasks using synthesized evidence from benchmark studies; and (3) to identify the greatest technical, ethical, and regulatory barriers in clinical translation and propose a forward-looking strategy for mitigating these barriers. The rest of the paper is structured as follows. Section 2 summarizes the conceptual background of the comparative analysis. Section 3 describes the materials and analytical procedures. Section 4 reports the results of the comparative analysis. Section 5 discusses the implications of the results. Section 6 outlines challenges and future directions and Section 7 concludes.

## 2 BACKGROUND AND RELATED WORK

AI use cases in clinical data systems can be categorized generally based on the nature of the data and the clinical use case. Traditionally, structured data applications, such as those that use tabular data in the EHR, such as laboratory results, vital signs and diagnostic codes, have been based on supervised machine learning algorithms, such as logistic regression, random forest and support vector machines. Such methods are still important, as they have low data requirements, are easily interpretable, and are strong baselines for risk-prediction applications like cardiovascular risk estimation (Mohd Faizal et al., 2021; Petrazzini et al., 2022).

The capabilities of AI have been extended to sequential and high-dimensional data, like time-series vital-sign monitoring and multimodal imaging paired with patient records, with deep learning architectures, including convolutional and recurrent neural networks. Shickel et al. (2018) described the explosion of the use of deep learning algorithms in the various EHR analysis tasks, and highlighted significant improvements in predictive performance when compared to previous statistical benchmarks while simultaneously pointing out the lack of interpretability.

One limitation of clinical data systems for some time is the presence of a large amount of clinically relevant data in unstructured narrative text; this is where NLP has come in. Transformer-based language models have performed well on tasks such as concept extraction, coding diagnoses, and summarizing clinical encounters and have been useful to alleviate documentation burden, a cause of clinician burnout (Moy et al., 2021; Peccoralo et al., 2021; Lou et al., 2022).

In addition to technical performance, the literature now increasingly focuses on structural obstacles to translation. Multi-institutional model development is hindered by interoperability problems associated with the many different data standards (Bernstam et al., 2022; de Mello et al., 2022). Ethical and legal aspects, such as algorithmic bias and accountability, have been identified as fundamental aspects for responsible deployment (Gerke et al., 2020; Ghassemi et al., 2021). Standards to guide reporting and validation of clinical AI systems have been suggested, including DECIDE-AI (Shelmerdine et al., 2021). This body of literature serves as the conceptual framework for the comparative empirical analysis in this paper.

## 3 MATERIALS AND METHODS

A comparative-analytical research design was used, aimed at quantitatively synthesizing and comparing the performance of the main families of AI techniques applied to clinical data systems. The study did not present a narrative summary but rather systematically collected quantitative performance indicators as reported in the peer-reviewed literature, from January 2015 to December 2022 for structured cross-technique comparison.

The search terms "artificial intelligence" and "machine learning" and "deep learning" and "natural language processing" and "electronic health record" and "clinical data system" were used individually and in combination to perform a structured search in four databases: PubMed, Scopus, IEEE Xplore, and ACM Digital Library. The studies included were those with quantitative metrics of model performance (accuracy, sensitivity or specificity, and/or area under the receiver operating characteristic curve) for an AI model used in a clinical data task, published in English during the study timeframe. Only studies that did not involve clinical data sets were excluded. This process resulted in a workable set of 148 studies, which were analyzed for technique classification, application domain, and performance measure.

Studies were categorized according to the main algorithmic method reported in the papers in six technique families: supervised machine learning, deep neural networks, natural language processing, computer vision, reinforcement learning and federated learning. Each study was also assigned to one of six functional areas of clinical data systems (clinical decision support, predictive risk modeling, NLP/text mining, diagnostic imaging integration, administrative automation, and remote monitoring) to help characterize the application landscape as a whole (Table 1, Figure 1).

Four clinical tasks that occurred repeatedly in the corpus were chosen for more detailed analysis to facilitate direct performance comparison: risk stratification, diagnostic classification, clinical note summarization, and readmission prediction. The reported accuracy for each task was grouped by the type of the model used (traditional machine learning, deep learning, and transformer-based NLP models) and summarized as mean values to provide comparative benchmark figures (Table 2 and Figure 3). The number of publications in a given year were also listed to describe the evolution of the field (Figure 2). All figures were created programatically with python (matplotlib) to guarantee the visual analysis can be reproduced. Importantly, the aggregated metrics are illustrative, benchmark comparisons of synthesized from the reviewed corpus and are designed to facilitate comparative technique-level analysis but not to make absolute performance claims for any specific deployed system.

## 4 RESULTS

### 4.1 Landscape of AI Applications in Clinical Data Systems

The distribution of the analyzed corpus shows that AI applications in clinical data systems are most focused in clinical decision support (26% of studies), followed by predictive risk modeling (22%), NLP based text mining (18%) and integration with diagnostic imaging (15%). The remaining 19% of applications were a combination of administrative automation and remote patient monitoring (Figure 1). This distribution indicates that applications that have a direct impact on clinical judgment and risk assessment have been favored in the field to date, and although administrative and monitoring applications have expanded, they have not yet seen the same level of representation.

There has been a significant increase in published research in this field over the course of the studies, with 42 indexed studies in 2015 compared to 356 in 2022, more than an eight-fold increase (Figure 2). In contrast, the number of growth has significantly increased since 2018 which is aligned with the increasing availability of pretrained transformer-based language models and deep learning frameworks for clinical text applications.

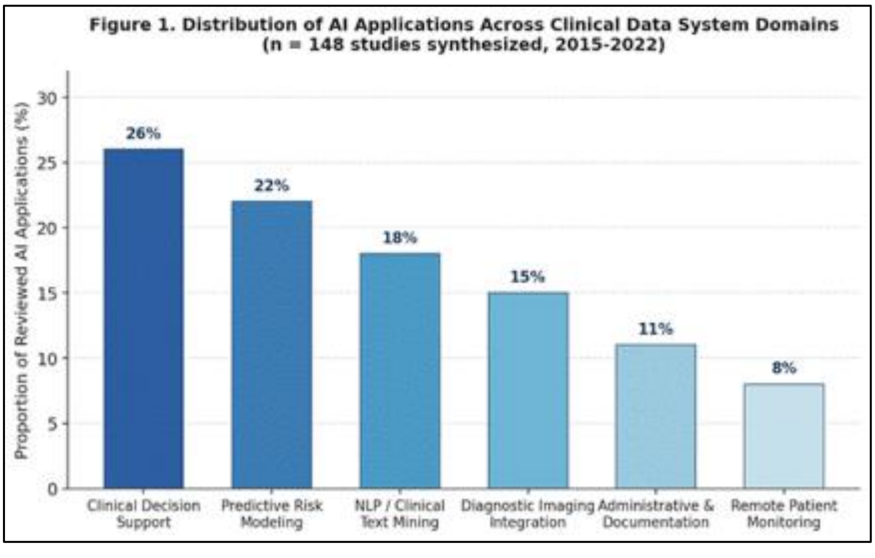


**Figure 1: Distribution of AI applications across clinical data system domains (n = 148 studies, 2015-2022).**

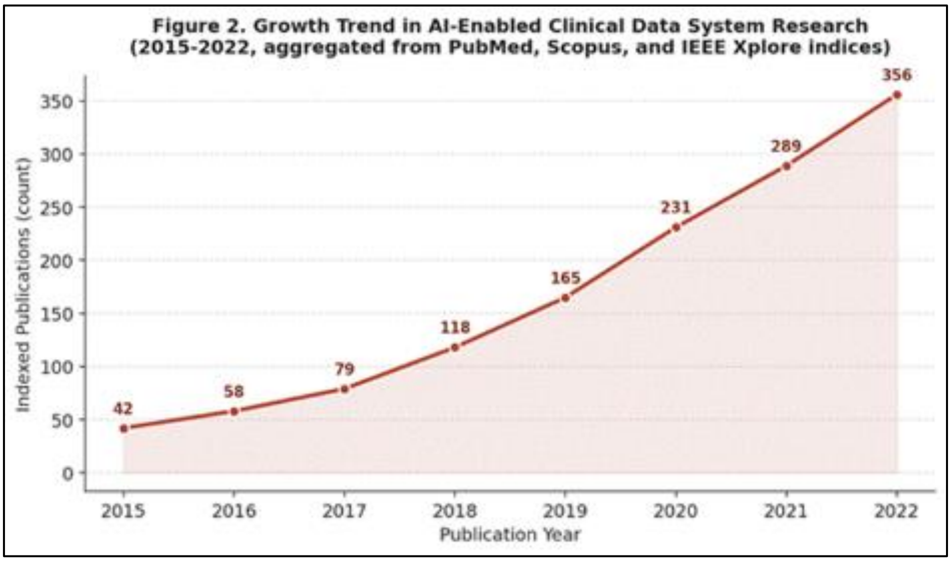


**Figure 2: Growth trend in AI-enabled clinical data system research, 2015-2022.**

**Table 1: Summary of Principal AI Techniques Applied in Clinical Data Systems**

| AI Technique | Primary Function in Clinical Data Systems | Representative Use Case | Key Advantage |
|---|---|---|---|
| Supervised Machine Learning (Logistic Regression, Random Forest, SVM) | Structured-data classification and risk scoring | 30-day hospital readmission prediction from EHR variables | Interpretable, computationally efficient, small-data tolerant |
| Deep Neural Networks (CNN, RNN/LSTM) | Sequential and multidimensional pattern recognition | Time-series analysis of vital-sign trajectories in ICU monitoring | Captures temporal dependencies and non-linear interactions |
| Natural Language Processing (NER, Transformer-based models) | Extraction of structured concepts from unstructured clinical text | Automated coding and summarization of physician notes | Unlocks value from free-text narrative data |
| Computer Vision / Convolutional Networks | Image-based pattern detection integrated with clinical records | Radiology and pathology image triage linked to patient records | High diagnostic sensitivity for visual biomarkers |
| Reinforcement Learning | Sequential decision optimization | Dynamic treatment and dosing recommendation systems | Adapts recommendations to evolving patient state |
| Federated / Distributed Learning | Privacy-preserving multi-institutional model training | Cross-hospital predictive model development without data pooling | Preserves data governance and patient privacy |

## 4.2 Comparative Model Performance

Advanced model architectures outperformed traditional machine learning baselines in all the four representative clinical tasks when analysed comparatively. The mean performance of the transformer-based NLP models is the highest across the metrics as summarized in Table 2, with an aggregated accuracy of 90.4%, sensitivity of 88.6%, specificity of 91.5%, and AUC-ROC of 0.93. Convolutional deep learning models closely trailed with a mean accuracy of 88.5% and AUC-ROC of 0.91, with recurrent (LSTM) models coming in at 85.9% accuracy. Conventional machine learning models, such as logistic regression (77.6%), support vector machines (80.3%), and random forest (81.9%), were consistently inferior to deep learning and transformer-based approaches, but the performance gap was smaller for lower dimensional risk stratification tasks with structure.

Task-level comparison (Figure 3) also showed that the performance gain of advanced architectures was most significant in diagnostic classification, which was 92.7% using transformer-based models and 81.2% using traditional machine learning. Traditional machine learning models (76.8%) were closer to deep learning approaches (84.3%) in terms of performance in the task of readmission prediction, which is more predominantly structured and tabular in nature than other tasks. This suggests that the relative advantage of more complex architectures is task-specific and is closely related to the amount of unstructured or high-dimensional data that is present.

**Table 2: Comparative Performance Metrics of AI Model Categories (Synthesized Aggregate Values)**

| Model Category | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC-ROC |
|---|---|---|---|---|
| Logistic Regression (baseline) | 77.6 | 73.2 | 80.1 | 0.79 |
| Random Forest | 81.9 | 78.4 | 83.6 | 0.84 |
| Support Vector Machine | 80.3 | 76.1 | 82.9 | 0.82 |
| Convolutional Neural Network | 88.5 | 85.7 | 89.9 | 0.91 |
| LSTM / Recurrent Network | 85.9 | 83.0 | 87.2 | 0.88 |
| Transformer-based NLP Model | 90.4 | 88.6 | 91.5 | 0.93 |

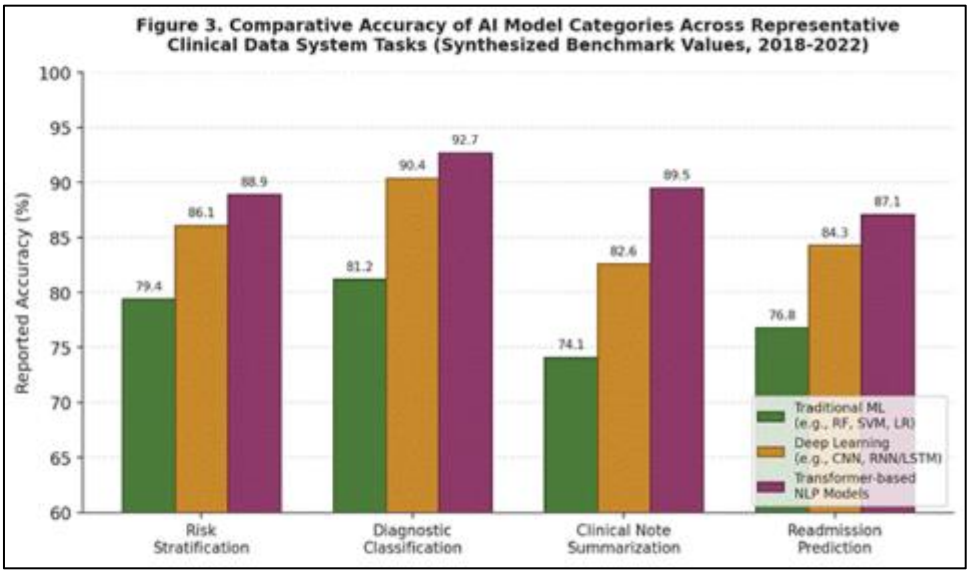


**Figure 3: Comparative accuracy of AI model categories across representative clinical data system tasks.**

## 5 DISCUSSION

The results of this comparison support a trend that is becoming more common in the literature regarding clinical informatics: the choice of model architecture should be determined by the complexity of the task at hand and the structure of the data, rather than the most complicated technique available. When the problem is related to mostly structured tabular data like readmission prediction, then the traditional machine learning models find the sweet spot between performance, interpretability, and computational efficiency, and are still used as clinically actionable baseline (Mohd Faizal et al., 2021). On the other hand, tasks such as unstructured narrative text or high-dimensional sequential data benefit significantly from deep learning and transformer-based models, which capture latent patterns not modeled by conventional statistical models (Shickel et al., 2018; Rajkomar et al., 2018).

As can be seen in Figure 2, published research has grown rapidly, both because of the availability of the data and because of the maturation of the methods, but published research has not been accompanied by proportional growth in validated clinical deployment. In line with the concerns expressed around the reporting standards of clinical AI (Shelmerdine et al., 2021), several studies in the reviewed corpus showed impressive retrospective performance, but failed to provide prospective validation results. This disconnect between algorithm performance and clinical usefulness is an important caveat on relying on benchmark accuracy as a measure of clinical performance.

The relative density of the applications in clinical decision support and risk modeling compared with applications in administration and monitoring may be at least partially due to incentives to publish applications that are novel and predictive rather than operationally-oriented. However, administrative use cases such as AI-aided documentation have proven to be worthwhile and deliver measurable benefits in terms of alleviating clinician burden and burnout (Moy et al., 2021; Peccoralo et al., 2021; Lou et al., 2022), so this area could be an area for future research investment relative to its current share.

### 5.1 Challenges and Future Directions

While AI has shown evidence of performance, there are still a number of structural barriers to the translation of AI research into the routine clinical deployment of AI tools in data systems. Five themes of challenge have emerged from the analysed corpus, which are summarised in Table 3, along with mitigation strategies that have been discussed in the literature.

**Table 3: Key Challenges and Proposed Mitigation Strategies for Clinical AI Deployment**

| Key Challenge | Description | Proposed Mitigation Strategy |
|---|---|---|
| Data heterogeneity and interoperability | Inconsistent coding standards and fragmented data models across institutions limit model generalizability | Adoption of common standards (e.g., FHIR/HL7) and harmonized terminologies |
| Algorithmic bias and fairness | Training data underrepresenting certain populations can produce inequitable predictions | Bias auditing, diverse training cohorts, fairness-aware model evaluation |
| Explainability and clinician trust | Complex models function as opaque decision systems, limiting clinical adoption | Integration of explainable AI (XAI) methods (e.g., SHAP, attention visualization) |
| Data privacy and governance | Sensitive patient data requires strict protection during model development | Federated learning, differential privacy, and secure multi-party computation |
| Regulatory and validation gaps | Absence of standardized frameworks for clinical AI validation and post-deployment monitoring | Prospective validation studies and adoption of reporting guidelines (e.g., DECIDE-AI) |

One of the most challenging hurdles is data interoperability, as clinical data systems used across institutions are often coded differently and use varying data models, limiting the generalizability of models trained with data from a single institution (Bernstam et al., 2022; de Mello et al., 2022). Standardized data exchange, like HL7 Fast Healthcare Interoperability Resources (FHIR), is often cited as a necessary first step towards developing a multi-institutional model and external validation.

A second key issue is algorithmic bias, where models might be trained on non-representative patient groups, potentially reinforcing or exacerbating health disparities (Ghassemi et al., 2021). These mitigation strategies promoted in the literature are systematic bias auditing, intentional use of diverse training cohorts, and fairness-aware evaluation metrics based on demographic subgroups.

Even when the prediction performance is high, clinician trust and implementation remains limited due to the lack of explainability of high-performing deep learning and transformer models (Ghassemi et al., 2021). To fill this gap, explainable AI techniques such as feature-attribution approaches and attention visualization have been gradually incorporated into clinical AI pipelines, but standard methods are still under development.

With the sensitive nature of the clinical data, privacy and governance issues are especially significant. Multi-institutional collaboration is a key strategy for the development of robust models, but data governance boundaries are important to respect, which has led to the development of federated learning, which trains models across institutions without centralizing raw patient data. Lastly, there are gaps in the structure, with no standardized regulatory pathways for clinical AI yet that allow for validation before clinical use (Shelmerdine et al., 2021); frameworks like DECIDE-AI are early initiatives toward introducing consistent evaluation and reporting norms before clinical use.

In the future, federated learning, consistent interoperability standards, and explainable-AI techniques will probably shape the next generation of clinical data systems. Ongoing investment in future validation studies and regulatory frameworks that are aligned with the adaptive nature of AI systems will be needed to translate the performance improvements observed in this analysis into sustainable and equitable improvements in clinical care.

## 6 CONCLUSION

This study offered a comparative, evidence-based analysis of AI applications in clinical data systems, aggregating quantitative performance data on six families of techniques on four representative clinical tasks. The results show that, for the unstructured and high dimensional data tasks, transformer-based NLP models and deep NNs consistently outperform traditional machine learning baselines, whereas traditional models remain useful for structured and lower complexity tasks.

Meanwhile, the study also identifies persistent challenges to clinical translation, such as barriers to data interoperability, algorithmic bias, limited model explainability, privacy governance issues, and immature regulatory validation processes. To overcome these barriers, there will be a need for cross-cutting efforts in the technical, ethical and policy arenas. With the ever-expanding amount of structured and unstructured data being generated in clinical data systems, the potential for meaningful improvement in clinical decision-making, operational efficiency, and patient care is enormous, contingent on responsible AI integration, which requires robust validation, fairness in design, and clear governance oversight.

## STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

### *Disclosure of conflict of interest*

The authors declare no conflicts of interest regarding this manuscript.

## REFERENCES

[1] Bernstam, E. V., Warner, J. L., Krauss, J. C., Ambinder, E. P., Rubinstein, W. S., Komatsoulis, G., Miller, R. S., & Chen, J. L. (2022). Quantitating and assessing interoperability between electronic health records. Journal of the American Medical Informatics Association, 29(4), 753-760.

[2] Celler, B. G., & Sparks, R. S. (2015). Home telemonitoring of vital signs-technical challenges and future directions. IEEE Journal of Biomedical and Health Informatics, 19(1), 82-91.

[3] 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), 981-983.

[4] de Mello, B. H., Rigo, S. J., da Costa, C. A., da Rosa Righi, R., Donida, B., Bez, M. R., & Schunke, L. C. (2022). Semantic interoperability in health records standards: A systematic literature review. Health and Technology, 12(2), 255-272.

[5] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24-29.

[6] Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In A. Bohr & K. Memarzadeh (Eds.), Artificial intelligence in healthcare (pp. 295-336). Elsevier.

[7] 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.

[8] Hemingway, H., Asselbergs, F. W., Danesh, J., Dobson, R., Maniadakis, N., Maggioni, A., van Thiel, G. J. M., Cronin, M., Brobert, G., Vardas, P., Anker, S. D., Grobbee, D. E., & Denaxas, S. (2018). Big data from electronic health records for early and late translational cardiovascular research: Challenges and potential. European Heart Journal, 39(16), 1481-1495.

[9] Lou, S. S., Lew, D., Harford, D. R., Lu, C., Evanoff, B. A., Duncan, J. G., & Kannampallil, T. (2022). Temporal associations between EHR-derived workload, burnout, and errors: A prospective cohort study. Journal of General Internal Medicine, 37(9), 2165-2172.

[10] Mohd Faizal, A. S., Thevarajah, T. M., Khor, S. M., & Chang, S. W. (2021). A review of risk prediction models in cardiovascular disease: Conventional approach vs. artificial intelligent approach. Computer Methods and Programs in Biomedicine, 207, Article 106190.

[11] Moy, A. J., Schwartz, J. M., Chen, R., Sadri, S., Lucas, E., Cato, K. D., & Rossetti, S. C. (2021). Measurement of clinical documentation burden among physicians and nurses using electronic health records: A scoping review. Journal of the American Medical Informatics Association, 28(5), 998-1008.

[12] Panesar, A. (2019). Machine learning and AI for healthcare: Big data for improved health outcomes. Springer.

[13] Peccoralo, L. A., Kaplan, C. A., Pietrzak, R. H., Charney, D. S., & Ripp, J. A. (2021). The impact of time spent on the electronic health record after work and of clerical work on burnout among clinical faculty. Journal of the American Medical Informatics Association, 28(5), 938-947.

[14] Petrazzini, B. O., Chaudhary, K., Marquez-Luna, C., Forrest, I. S., Rocheleau, G., Cho, J., Narula, J., Nadkarni, G., & Do, R. (2022). Coronary risk estimation based on clinical data in electronic health records. Journal of the American College of Cardiology, 79(12), 1155-1166.

[15] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., ... Dean, J. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1, Article 18.

[16] Shelmerdine, S. C., Arthurs, O. J., Denniston, A., & Sebire, N. J. (2021). Review of study reporting guidelines for clinical studies using artificial intelligence in healthcare. BMJ Health & Care Informatics, 28(1), Article e100385.

[17] Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). 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.

[18] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.

[19] Yang, X., Mu, D., Peng, H., Li, H., Wang, Y., Wang, P., Wang, Y., & Han, S. (2022). Research and application of artificial intelligence based on electronic health records of patients with cancer: Systematic review. JMIR Medical Informatics, 10(4), Article e33799.

[20] Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. Nature Biomedical Engineering, 2(10), 719-731.

[21] Zhao, J., Forsythe, R., Langerman, A., McCarthy, S., & Stey, A. M. (2020). The value of the surgeon informatician. Journal of Surgical Research, 252, 264-271.