A Systematic Review on the Evolution of AI in Healthcare ecosystems

Rashmi S[1], Updendra Kumar[2]
[1] Lincoln University College, Malaysia ; Department of Computer Science & Engineering, Institute of Engineering & Technology, Lucknow ;
[1] pdf.rashmi@lincoln.edu.my, [2]*upendra.ietlko@gmail.com*

Abstract: Clinical workflows are the organized sequence of processes and decision nodes associated with patient care, ranging from diagnostics and treatment planning through to monitoring and administrative processes. Clinical workflows tend to be multifaceted, time-sensitive, and open to variability, which is why they are an excellent fit for development using Artificial Intelligence (AI). Artificial intelligence technologies hold the promise to rationalize clinical workflows by enhancing diagnostic precision, streamlining repetitive activities, aiding in decision-making, and facilitating personalized treatment regimens. A large number of methods ranging from statistical models and machine learning algorithms to rule-based systems and hybrid architectures are being used more and more to realize these promises. This paper focus on these methods and compares based on their transparency, precision, computational complexity, and training data needs.

Keywords: Artificial Intelligence; Machine Learning; Healthcare; LLMs; Clinical workflow.

Introduction

Artificial Intelligence is reforming the healthcare domain [1] by revamping the mode of medical services that includes the treatment plans, customizing patient care, better diagnostic accuracy and automating administrative workflows. AI has significant applications in healthcare, boosting both clinical and operational efficiency.

The major applications of AI in healthcare include advanced diagnosis and medical imaging, accelerated drug discovery, personalized treatment planning, intelligent virtual health assistants, and the optimization of clinical workflows.

As AI continues to evolve, it has taken on a bigger role in healthcare, leading to big changes and improvements in many areas of medical practice. The advent of innovative gadgets such as fitness trackers and other wearable devices, smart sensors, monitoring devices has contributed to the integration of AI in healthcare.

In the 1970s, Rule-based AI systems were first developed for healthcare. Later its usage peaked during the 1980s and 1990s and in the 2000s onwards evolved into hybrid AI systems.

Thomas D a v e n p o r t a n d R a v i K a l a k o t a in [2] explores how AI is transforming clinical medicine, practical challenges and its future prospects. The key technologies of AI relevant to healthcare include

Rule Based Expert Systems: It was commonly adopted for Clinical decision support systems in the 1970s. . They are typically based on a collection of rules composed of 'if-then' logic. They required humans with domain knowledge expertise to formulate rules. These systems were replaced by data driven Machine learning models. Although Electronic Health Record (EHR) could be maintained well it efficiency reduces in a complex environment.

Machine Learning (ML): ML, a statistical technique that trains models with data, has its most widely used application in healthcare [3] in precision medicine. Precision Medicine is a treatment method that is contrary to one-size-fits-all. It relies on data analysis and predicts outcomes of the treatment considering patient attributes. While the majority of the applications are based on Supervised Learning, neural networks and Deep Learning are also employed. Deep neural networks that are capable of identifying hidden patterns are used to spot suspicious lesions that exhibit cancer-related abnormalities. This feature is most suited for radiology for detecting features not visible to the human eye.

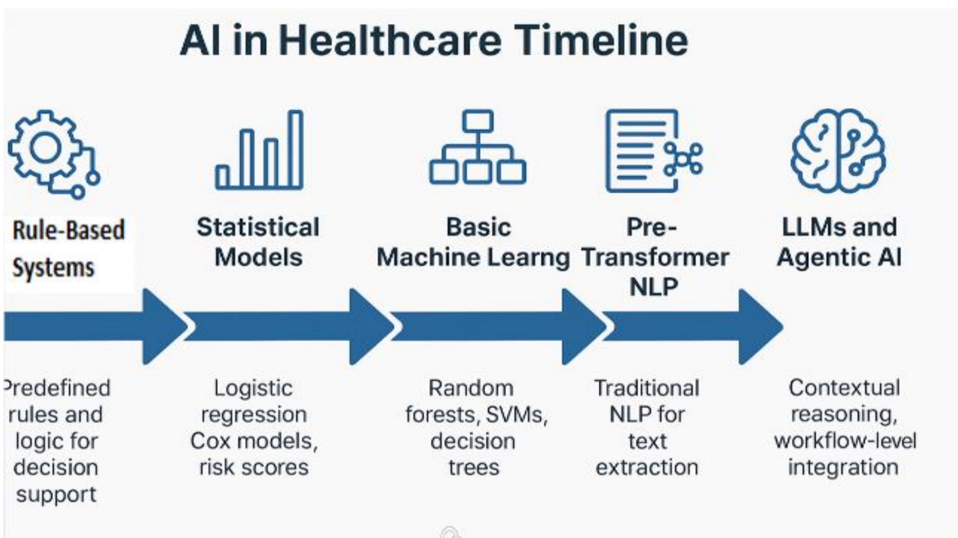


Figure 1: Role of AI in Healthcare Timeline

Natural Language Processing with Deep Learning: This duo can be used in healthcare [4] [5] for converting speech-to-text during doctor -patient interactions, extracting and analyzing useful insights from Electronic Health Records (EHR). It can also be used in clinical documentation reducing overhead of manual entries, saving time.one of the most efficient usage can be directly transcribing verbal conversation of the healthcare professional into records. Decision making can be quickened by deriving insights from EHR

Robotic process automation: RPA, together with Machine Learning can be called Intelligent Automation. RPA uses domain rules, workflows taking the role of a semi-intelligent user of the systems. In healthcare, they can be used to detect fraud insurance claims, retrieving text from scanned images to mention a few. AI driven RPA, visual automation are the future of Smart health solutions [6]

Figure 1 shows the summary of the evolution of different technologies of AI in healthcare. This paper presents a review of the research work in the field of healthcare technology, contemporary improvements and state-of-the-art developments

Related work

Xiaolan Chen et al..[7] reviews how Large Language Models(LLMs) and LLM-based agents can contribute to healthcare settings and that it calls for a transformation from a static evaluation models to dynamic, adaptive real world medical practice. It refers to the medical exams that present insights used to gauge LLMs responsiveness to evolving medical science. They include the National Pharmacist Licensing Examination in China, United States Medical Licensing Examination (USMLE), Chinese Master's Degree Entrance Examination, National Medical Licensing Examination in China,
Otolaryngology-Head and Neck Surgery Certification Examinations , National Nurse Licensing Examination in China , American Board of Neurological Surgery (ABNS), European Board of Radiology exam, the Royal College of General Practitioners Applied Knowledge Test and many more. Some research uses manually collected multimodal data in real world scenarios.
The paper categorizes LLM evaluation into four levels:
Closed Ended Tasks: It mainly uses MCQs to assess factual data and can objectively measure recall and precision
Open Ended Tasks: It involves information extraction such as disease prediction, treatment plans, summarization of EHRs.
Image Processing Tasks: It refers to visual questions answering, image classification, segmentation and report generation.
LLM agents: simulating workflows with planning and tool usage .The paper considers both human and automated evaluation methods using tools and metrics such as DISCERN, JAMA and ROUGE, METEOR and F1-scores respectively.
It also identifies key challenges such as standardization of evaluation metrics, non-reliability of automated evaluation, scarcity of data sets etc., in deploying LLMs in medical practice.
Major tips for going forward include building diverse, good quality datasets and to devise automated evaluation tools.

Patrick Schober et al in [8] shows how Logistic Regression, a statistical method can be used for binary classification predicting the probability of disease or no disease. In clinical research, it supports decision-making and risk prediction. To illustrate, it considers a neonate undergoing a surgical procedure called Pyloromyotomy. Logistic regression was used for a study to assess the risk of hypoxemia where oxygen saturation is less than 90% linear (straight line) relationship with the logit (the natural logarithm of the odds) of the outcome.

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X$$

The probability (P), the probability has a sigmoidal relationship with the independent variable, and the estimated probabilities are now appropriately constrained between 0 and 1. The study compares traditional RSI and modified RSI induction techniques with an outcome of adjusting the confounders. It also concludes that the traditional RSI had three times higher risk.

[9] LLMs-7 analyzes the way in which LLMs are implemented in real world clinical workflows. The study is specific to GPTs, the generative pre-trained transformers discussing the applications, quality assessments and the challenges involved.. Amongst the multiple studies, four of them were deployed in real time. It included outpatient communication, data extraction, Psychological counseling and therapy services, responses to patients and clinicians and similar services. The adoption of LLM-based solutions have shown effectiveness in patient care delivery, reduction in the clinician workload and patient satisfaction as well. However, performance was dependent on the actual task and use cases and it degraded over time.

Patrick Schober et al in [8] in their article offer a brief and applied introduction to logistic regression, a statistical method commonly utilized in medical studies to determine the association of one or more independent variables with a dichotomous outcome (e.g., having or not having a condition). The authors clarify that logistic regression is well adapted to estimating probabilities and odds, and they highlight its utility in observational studies where adjustment for confounding variables is critical. Following a description of how logistic regression controls for many covariates and presents results in the form of odds ratios—rendering them extremely interpretable and clinically meaningful—the author applies an example of forecasting hypoxemia during induction methods in neonates to demonstrate the effect of controlling for several covariates using logistic regression. It adds testing tools such as the Hosmer–Lemeshow goodness-of-fit test and the area under the ROC curve to gauge model performance. By emphasizing interpretability, ease of access, and stability of logistic regression, the authors recommend its use in clinical investigations particularly when lucidity in comprehending variable-outcome relationships is crucial for informing medical decisions.

Alam et al in [10] documents a strong methodology for enhancing medical data classification with the help of a combination of feature ranking methodologies and the Random Forest algorithm. The authors meet the challenge of dealing with high-dimensional medical informatics datasets by initially applying different feature ranking techniques (Information Gain, Gain Ratio, ReliefF) to select the most informative features. Subsequently, these features are utilized to train Random Forest classifiers, greatly eliminating noise and computational expense while improving predictive accuracy. The authors use this framework on ten benchmark datasets involving various diseases and show that models based on top-ranked features outperform baseline models based on all features indiscriminately.

The assessment entails exhaustive metrics like accuracy, precision, recall, F1-score, AUROC, AUPR, and RMSE with the support of 10-fold cross-validation and independent testing. These results show that feature selection not only reduces computation but also enhances classification performance across various diseases. The article concludes by promoting the cross-applicability of the method and the prospect for use across a variety of clinical settings, providing a blueprint for developing high-performing, domain-independent diagnostic devices. The approach also highlights the centrality of explain ability via feature importance, albeit recognizing the unavoidable trade-off with the interpretability of the Random Forest model in itself.

Kierner et al in [11] explores the integration of rule-based reasoning with machine learning (ML) in clinical decision systems (CDS), addressing the growing demand for systems that balance predictive accuracy with transparency. After reviewing 957 publications and analyzing 71 that met inclusion criteria, the authors identify five primary motivations behind hybrid architectures: improving prediction accuracy, handling limited data, conforming to clinical knowledge, enhancing explain ability, and improving health outcomes. Based on these motivations, the paper defines five architectural types for combining ML and rules: Rules Embedded in ML (REML), ML pre-processing for Rules (MLRB), Rules pre-processing for ML (RBML), Rules influencing ML training (RMLT), and Parallel Ensemble of Rules and ML (PERML).

The authors emphasize that while embedded and sequential hybrid approaches dominate the literature, parallel systems offer untapped potential for combining interpretability and performance. The paper critically discusses how integrating expert knowledge via rules can enhance trust and regulatory acceptance, especially in high-stakes medical domains. However, it also acknowledges practical limitations, such as the complexity of implementation and limited real-world deployment. The study offers a foundational taxonomy that researchers and developers can use to guide the design of transparent, reliable, and effective hybrid AI systems in healthcare.

Müller & Rahm in [12] intend to solve a major weakness of conventional workflow systems in medicine: their inability to dynamically change in response to unforeseen clinical events or exceptions. The authors introduce a rule-based approach that supports automated, real-time adaptation of medical workflows at runtime. Looking within the space of -distributed cancer therapy-, the system identifies -semantic exceptions —like new allergy or out-of-range lab results—and how different elements of the current workflow must be changed. The method relies on formal logic (Frame Logic and Transaction Logic) to describe events and rules so that the system possesses a clearly defined semantic foundation for invoking changes.

Two programs, drcd and p-algorithm , are designed to identify impacted workflow areas and implement required adjustments, like inserting, deleting, or substituting certain activities. This targeted and smart adjustment lessens human intervention, enabling doctors to modify patient treatment plans and reducing man-in-the-loop error. While the system emphasizes control flow over data flow and does not yet allow for major structural changes, it does prove that rule-based agents are able to enable dynamic, patient-adaptive treatment adjustments. The framework is used in the HEMATOWORK system, proving the practical applicability of the technique and the suitability for use in other adaptive clinical processes.

Key Contribution

A comparative study of these papers have been provided in the following table.

Table 1.1: Comparative study based on different parameters

Aspect	Logistic Regression (Paper 1)	Random Forest (Paper 2)	Rule-Based (Scoping Review) (Paper 3)	Rule-Based (Workflow Mod) (Paper 4)
Title	Logistic Regression in Medical Research	A Random Forest based predictor for medical data classification	Taxonomy of hybrid architectures involving rule-based reasoning and ML	Rule-Based Dynamic Modification of Workflows in a Medical Domain
Core Method	Binary Logistic Regression	Feature Ranking + Random Forest Classifier	Hybrid Systems (Rules + ML: REML, RBML, PERML, etc.)	Rule-based detection and handling of workflow exceptions
Primary Focus	Explaining logistic regression for binary outcomes	Improving prediction accuracy with feature ranking before classification	Reviewing hybrid system architectures and motivations	Dynamic adjustment of medical workflows based on clinical events
Use Case	Predicting hypoxemia risk in neonates (RSI technique)	Classifying 10 diseases using UCI datasets	Various clinical decision support applications (diagnosis, prognosis, etc.)	Cancer therapy treatment adjustments
Advantage Highlighted	Interpretability via odds ratios	Generalizability and high accuracy across datasets	Transparency, accuracy, and trust in hybrid models	Flexibility and automation in managing clinical workflow changes
Limitations Discussed	Assumptions (e.g., linearity, independence)	Assumes high-quality ranked features	Complexity of integration, lack of real-world deployment data	Domain-specific tuning; limited to local changes in workflows
Evaluation	Conceptual and example-driven	Empirical (10 datasets, metrics like accuracy, F1, ROC)	Literature review of 71 papers, taxonomy creation	Prototype development within HEMATOWORK system

Transparency vs. Accuracy Tradeoff	High transparency, limited complexity modeling	Higher accuracy, less interpretability	Strives for both via structured hybrid designs	High transparency; focus on explainability through rules
Model Type	Statistical	Ensemble ML	Mixed (ML + Rule-Based)	Rule-Based System with semantic exception detection

Fig 2 shows the Accuracy vs Transparency graph that depicts how various healthcare AI approaches compromise between interpretability and predictive accuracy. Accuracy, which indicates how well a model classifies or predicts outcomes, is on the x-axis, and transparency, or how clear the model is to clinicians, is on the y-axis. Logistic regression (Paper 1) is high in transparency but low in accuracy; it produces clear, understandable outputs by odds ratios, so clinical explanations are good, although it does not handle complex data well. Random Forest (Paper 2) is low in transparency but high in accuracy, as it has an ensemble structure that makes the predictions very good but interpretation very bad. Hybrid rule-based and machine learning models (Paper 3) take a middle ground—moderately transparent and fairly accurate—by blending rule logic into ML models to enhance trust at the cost of not compromising performance greatly. Rule-based workflow systems (Paper 4) rank high on transparency but low on accuracy because they do not entail learning from data but are high on adaptability and traceability in clinical settings. This trade-off is particularly critical in medicine, where interpretability usually has higher priority than pure performance, especially in high-stakes clinical decision-making. The graph serves to inform model choice according to the requirements of a particular clinical application.

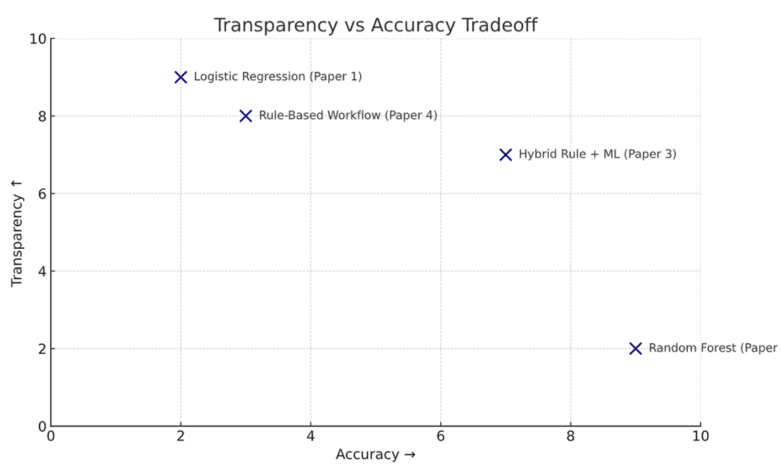
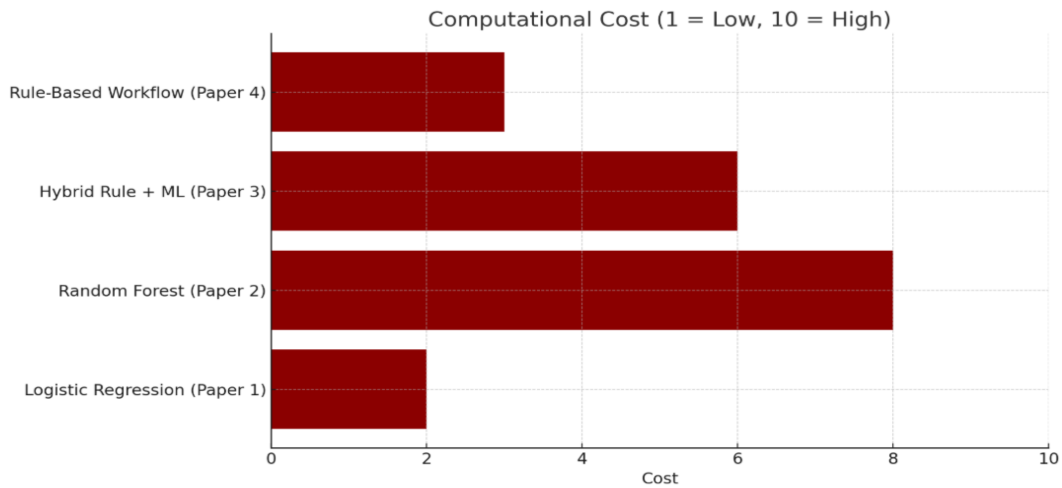


Fig 2: Transparency vs Accuracy Tradeoff



Results

Fig 3: Computational Cost comparison graph

Fig 3 shows the resource demand of each healthcare AI technique in terms of processing capacity, memory consumption, and execution time. Models with taller bars on this graph require more computational resources , so they are better suited for adequately equipped clinical setups or offline batch processing. Those with shorter bars are lighter and suitable for real-time or resource-limited setups.

Random Forest (Paper 2) is the most computationally expensive because of its ensemble approach, where many decision trees are learned and combined. This algorithm is particularly expensive when used with feature ranking methods and cross-validation as in the paper. Hybrid Rule + Machine Learning systems (Paper 3) comes in a comparably high cost, especially when intricate combination of rule engines and learning models needs to be achieved—usually including both symbolic reasoning and statistical computation.

At the opposite end of the spectrum, Logistic Regression (Paper 1) and Rule-Based Workflow Systems (Paper 4) have zero computational cost . Logistic regression is a linear model with closed-form solution and efficient convergence, so it's very quick to train and deploy. Rule-based workflow systems depend on pre-specified logic and don't do data-driven learning, so their execution is quick and computationally light. This chart facilitates stakeholders' estimation of the practical feasibility of implementing these models, particularly in resource-constrained or time-critical clinical settings.

Fig 4 shows the amount of data each model or approach generally requires in order to function well in healthcare environments. Models further to the right along the x-axis need larger datasets to achieve peak performance, whereas those to the left can function adequately with limited or smaller datasets.

Logistic Regression (Paper 1) and Rule-Based Workflow Systems (Paper 4) fall on the lower spectrum , since they can function well even with limited quantities of data. Logistic regression is based on clearly defined statistical relationships and formatted variables, whereas rule-based systems actually implement expert knowledge in a direct decision rule form and reduce reliance on training data as a whole.

Conversely, Random Forest (Paper 2) and Hybrid Rule + Machine Learning systems (Paper 3) are placed higher on the continuum. These techniques gain much from larger, high-dimensional datasets to properly learn patterns and interactions between variables. Random Forest specifically relies on ensemble learning from multiple decision trees, which needs varied and abundant input data to prevent overfitting or bias. Hybrid solutions also need significant amounts of data, particularly when the machine learning component dominates prediction or decision-making support. This graph points to the significance of data availability when selecting an AI method for a given clinical setting.

Fig 4: Training Data Requirement graph

Conclusions

As health systems increasingly embrace digital technologies, the incorporation of AI into clinical processes both offers great promise and poses real-world issues. This comparative analysis suggests that there is no one AI method that dominates others across all contexts; instead, each has some strengths best suited for particular clinical situations. Logistic regression is highly interpretable and simple, which makes it highly suitable for transparent decision-making in resource-constrained environments. Random Forest models are highly accurate predictively but require a lot of computation and large amounts of data, potentially restricting their application in real-time or resource-constrained contexts. Hybrid rule-based and machine learning systems provide a compelling middle ground, allowing for enhanced accuracy with some degree of transparency maintained through inherent clinical logic. Last but not least, rule-based workflow systems emphasize flexibility and human understandability, which is especially important in dynamic, high-exception situations such as long-term therapy management.

Finally, the choice of AI techniques in clinical workflows must be informed by the particular objectives of the healthcare application—whether that is optimizing outcomes, preserving clinical trust, or guaranteeing operational efficiency. Being aware of the trade-offs between transparency, accuracy, requirements on data, and computational cost is critical to implementing safe, effective, and ethically justifiable AI in actual medical settings.

References

1. Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: Transforming the practice of medicine. Future Healthcare Journal, 8(2), e188–e194. https://doi.org/10.7861/fhj.2021-0095
2. Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94
3. Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. Innovative Infrastructure Solutions, 7(3), Article 66. https://doi.org/10.1016/j.ijin.2022.05.002
4. Shastry, K. A., & Shastry, A. (2023). An integrated deep learning and natural language processing approach for continuous remote monitoring in digital health. Digital Applications in Organic and Hybrid Electronics, 100301. https://doi.org/10.1016/j.dajour.2023.100301
5. Chakraborty, C., Bhattacharya, M., Pal, S., & Lee, S.-S. (2023). From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare. Current Research in Biotechnology, 5, 100164. https://doi.org/10.1016/j.crbiot.2023.100164
6. Nimkar, P., Kanyal, D., & Sabale, S. R. (2022). Increasing trends of artificial intelligence with robotic process automation in health care: A narrative review. Cureus, 14(9), e29556. https://doi.org/10.7759/cureus.29556
7. Chen X, Xiang J, Lu S, Liu Y, He M, Shi D. Evaluating large language models and agents in healthcare: key challenges in clinical applications. *Intell Med*. 2025; **7**:100123. https://doi.org/10.1016/j.imed.2025.03.002
8. Schober, P., & Vetter, T. R. (2021). Logistic Regression in Medical Research. Anesthesia & Analgesia, 132(2), 365–366. https://doi.org/10.1213/ANE.0000000000004915
9. Artsi Y, Sorin V, Glicksberg BS, Korfiatis P, Nadkarni GN, Klang E. Large Language Models in Real-World Clinical Workflows: A Systematic Review of Applications and Implementation. *medRxiv* [Preprint]. 2025 Jun 11. https://doi.org/10.1101/2025.06.10.25329323
10. Alam, M. Z., Rahman, M. S., & Rahman, M. S. (2019). A random forest based predictor for medical data classification using feature ranking. Informatics in Medicine Unlocked, 15, 100180. https://doi.org/10.1016/j.imu.2019.100180
11. Kierner, S., Kucharski, J., & Kierner, Z. (2023). Taxonomy of hybrid architectures involving rule-based reasoning and machine learning in clinical decision systems: A scoping review. Journal of Biomedical Informatics, 144, 104428. https://doi.org/10.1016/j.jbi.2023.104428
12. Müller, R., & Rahm, E. (1999). Rule-based dynamic modification of workflows in a medical domain. In A. P. Buchmann (Ed.), Datenbanksysteme in Büro, Technik und Wissenschaft (pp. 429–448). Springer-Verlag.