



International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	ELECTRONIC HEALTH RECORD-BASED PREDICTIVE APPROACHES FOR HOSPITAL READMISSION: A STRUCTURED REVIEW OF SOCIOECONOMIC VARIABLES AND ALGORITHMIC EQUITY
DOI	https://doi.org/10.31435/ijitss.4(52).2026.6377
RECEIVED	20 July 2026
ACCEPTED	03 October 2026
PUBLISHED	14 October 2026
LICENSE	 The article is licensed under a Creative Commons Attribution 4.0 International License .

© The author(s) 2026.

This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

ELECTRONIC HEALTH RECORD-BASED PREDICTIVE APPROACHES FOR HOSPITAL READMISSION: A STRUCTURED REVIEW OF SOCIOECONOMIC VARIABLES AND ALGORITHMIC EQUITY

Maciej Kamil Majchrzak (Corresponding Author, Email: m.majchrzak121@gmail.com)

Faculty of Medicine, Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0007-4739-0092

Faustyna Magdalena Kołodziej

Faculty of Medicine, Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0002-8778-7100

Krzysztof Konopka

Faculty of Medicine, Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0003-4055-0074

Mateusz Lis

Faculty of Medicine, Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0006-8441-8296

Zuzanna Julia Komin

Faculty of Medicine, Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0001-6284-6410

Maryan Liasota

Faculty of Medicine, Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0006-0266-4849

Eliza Zofia Kabat

Faculty of Medicine, Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0003-6531-967X

Jakub Juskowiak

Faculty of Medicine, Poznań University of Medical Sciences, Poznań, Poland

ORCID ID: 0009-0001-0372-400X

Alicja Kwiatkowska

Faculty of Medicine, Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0003-1027-2114

Michalina Justyna Kozłowska

Faculty of Medicine, Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0003-9561-2957

ABSTRACT

Background: Hospital readmission rates are often used to judge the quality of care and how well healthcare systems are working. Predictive models based on electronic health records (EHRs) can help estimate which patients are at higher risk, but when they don't properly reflect socioeconomic realities, they can unintentionally reinforce existing inequalities (Obermeyer et al., 2019). This review looks at how well these EHR-based models predict 30-day readmissions, how Social Determinants of Health (SDOH) are included, and what this means for fairness and health equity in real-world use.

Methodology: A structured literature review was conducted using PubMed/MEDLINE and Google Scholar. It included peer-reviewed studies published between 2018 and 2026, focusing on model validation studies, retrospective cohort analyses, and research on bias and fairness in predictive healthcare models.

Results: Synthesis reveals that incorporating unstructured and structured SDOH variables yields mixed impacts on discriminatory performance (Area Under the Receiver Operating Characteristic curve changes ranging from -0.01 to +0.04), yet significantly rectifies systemic algorithmic underestimation of risk for marginalized demographics. Unadjusted models consistently show racial, ethnic, and socioeconomic bias, often because they rely on proxy variables embedded in historical healthcare data (Gianfrancesco et al., 2018). As a result, preventive transition-of-care resources can be systematically directed away from the patients who need them most (Pierson et al., 2021).

Conclusion: To make predictive healthcare tools truly useful in practice, it is not enough to focus only on accuracy. They also need to account for fairness, real clinical context, and the actual consequences of their use in everyday healthcare settings (Wiens et al., 2019). Policy-level interventions must mandate standardized SDOH extraction alongside equity-focused calibration algorithms before widespread clinical deployment (Ghassemi et al., 2021).

KEYWORDS

Predictive Modeling, Electronic Health Records, Hospital Readmission, Social Determinants of Health, Algorithmic Fairness, Health Equity

CITATION

Majchrzak, M. K., Kołodziej, F. M., Konopka, K., Lis, M., Komin, Z. J., Liasota, M., Kabat, E. Z., Juskowiak, J., Kwiatkowska, A., & Kozłowska, M. J. (2026). Electronic health record-based predictive approaches for hospital readmission: A structured review of socioeconomic variables and algorithmic equity. *International Journal of Innovative Technologies in Social Science*, 4(52). [https://doi.org/10.31435/ijitss.4\(52\).2026.6377](https://doi.org/10.31435/ijitss.4(52).2026.6377)

COPYRIGHT

© The author(s) 2026. This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

1. Introduction

1.1 Background and Clinical Context of Hospital Readmissions.

Unplanned hospital readmissions within 30 days of discharge remain a persistent and significant challenge for health systems worldwide. From a purely clinical perspective, a rapid return to an acute care setting frequently signifies an incomplete transition of care, unresolved physiological instability, or a systemic failure in post-discharge coordination. For patients, these episodes are rarely harmless. They often mean a quicker deterioration in health, more complications during the hospital stay, noticeable psychological strain, and a higher risk of death.

On a broader level, readmissions are also costly. In high-income healthcare systems, a lot of money and resources still go into repeat hospital admissions, even though many of them could at least partly be prevented. Because of this, many systems have introduced policies and financial incentives to reduce their frequency. In the United States, the Centers for Medicare & Medicaid Services (CMS) penalizes hospitals with high readmission rates through the Hospital Readmission Reduction Program (HRRP). In many European healthcare systems, the focus is less on penalties and more on how care is organised, mainly through structured care pathways and payment models that emphasise coordination and overall quality of care rather than just the number of admissions.

Historically, clinical efforts to anticipate which patients face the highest risk of readmission relied on subjective bedside assessments or rudimentary, manually calculated risk scores (e.g., the LACE Index or the Hospital Readmission Risk Score). These tools were useful because they gave clinicians simple, rule-based ways to make decisions. But they were also quite limited. They usually only looked at a small set of fixed clinical and demographic variables and missed how complex and interconnected patient conditions actually are. They also didn't really reflect how varied disease progression can be between patients, especially in those with multiple chronic conditions.

As hospitals deal with more patients who have complex, long-term illnesses, there has been a shift toward more advanced, automated models that can capture more detail and better reflect real clinical complexity (Topol, 2019).

1.2 The Rise of EHR-Based Predictive Modeling

The digital revolution within medicine—characterized by the near-universal adoption of comprehensive Electronic Health Records (EHRs)—has fundamentally transformed clinical informatics from a passive system of digital documentation into a highly active substrate for operational optimization and predictive medicine. EHR infrastructures accumulate vast quantities of multi-dimensional data during a single inpatient encounter, ranging from longitudinal laboratory trajectories and real-time hemodynamic monitoring to continuous medication administration logs and unstructured clinical narratives (Beam & Kohane, 2018).

At the same time, better computing power and machine learning have made it possible to spot useful patterns in large, messy healthcare data that actually help with predictions (Rajkomar, Oren, et al., 2018). Instead of assuming simple, fixed relationships between variables, modern clinical AI uses more flexible models—such as gradient-boosted decision trees (e.g., XGBoost), random forests, and deep neural networks—to analyze many different data points at once and capture patterns that would otherwise be missed (Esteva et al., 2019). These advanced predictive models are designed to continuously parse an individual's evolving EHR profile, generating automated, dynamic risk probabilities for 30-day readmissions directly within the clinician's digital workspace.

When they work properly, these tools can support clinical decision-making by flagging patients at higher risk before discharge actually happens. That gives doctors, case managers, and social workers more room to step in earlier and put the right support in place (Sendak et al., 2020). Instead of reacting once things have already gone wrong, they can double-check medications, arrange quicker follow-ups, organise home care, or make discharge planning more structured and coordinated.

In practice, it's about shifting care away from firefighting once things have already fallen apart, and towards catching problems earlier and adjusting care to the person before they deteriorate.

1.3 From Predictive Accuracy to Algorithmic Fairness and SDOH

As these predictive systems migrate from theoretical validation studies into active, real-time clinical deployment, a profound structural tension has emerged at the intersection of technological capability and social reality (Kelly et al., 2019). Traditionally, the success of a machine learning model was judged almost exclusively through mathematical, performance-centric lenses, primarily measured by metrics such as the Area Under the Receiver Operating Characteristic curve (AUC-ROC), sensitivity, and precision. However, this strong focus on mathematical optimisation assumes that the data used to train these models is a neutral and objective reflection of human biology (Char et al., 2018).

In clinical reality, a patient's risk of readmission is often shaped much more by what happens outside the hospital than by the physiological measurements recorded in hospital charts. These non-clinical factors are broadly classified as the Social Determinants of Health (SDOH) and encompass a wide range of socioeconomic realities: economic stability housing security, geographical proximity to pharmacies and primary care clinics,

structural literacy, and systemic access to nutritious food (Gottlieb et al., 2019). When an EHR-based algorithm is developed without socioeconomic information, an important part of the patient's situation is missing. Clinical information on its own only captures part of the situation. A patient's financial situation, housing, education, or access to healthcare outside the hospital can all play an important role in whether they are readmitted, but these factors are often missing from the data.

When they are missing, models trained on historical healthcare records can end up reproducing the same inequalities that already exist in the system (Challen et al., 2019). In practice, this can mean that some patients are seen as lower risk than they really are, which can affect who gets extra follow-up or support after discharge. At the same time, adding socioeconomic and demographic variables is not straightforward, because it raises its own challenges around bias and interpretation. Medical data itself reflects past patterns of care, including inequality in access and treatment. For example, if a certain group has historically used primary care less often due to barriers like discrimination or lack of insurance, an unadjusted model may incorrectly read this as lower disease severity, rather than a lack of access (Vyas et al., 2020). Consequently, the model mathematically underestimates the patient's true readmission risk.

This creates a dangerous feedback loop where resource-intensive, post-discharge support is systematically misallocated to wealthier, more privileged patient cohorts who simply have better documentation or higher historical healthcare expenditure. As a result, clinical informaticians and public health strategists are increasingly recognizing that optimizing a model for raw predictive accuracy can inadvertently maximize structural unfairness, thereby transforming an innovative technological intervention into an unintended engine for widening health disparities (Rajkomar, Hardt, et al., 2018).

1.4 Research Gap and Rationale

Despite a rapidly growing body of literature dedicated to refining the statistical performance of machine learning algorithms for hospital readmissions, a significant and problematic fragmentation persists in the current state of research (Ghassemi et al., 2020). Most published studies focus on technical development, testing new deep learning or natural language processing models on carefully curated and relatively uniform clinical datasets (Miotto et al., 2018). Conversely, a separate stream of sociological and public health literature focuses on the conceptual frameworks of health equity and structural bias, but often lacks deep integration with the technical, operational realities of clinical informatic pipelines (Chen, Szolovits, & Ghassemi, 2019).

There is a critical lack of synthesized, interdisciplinary evidence that rigorously investigates the exact technical mechanisms through which socioeconomic variables are integrated into active EHR-based risk models, and how these integration methods directly affect mathematical fairness metrics (such as demographic parity or equalized odds) (Pfohl et al., 2021). Many healthcare institutions are already using commercial or open-source readmission risk tools, even though it is not always clear how well these models perform across different patient groups. The evidence is also mixed. Some studies suggest that leaving out demographic and socioeconomic variables can reduce bias, while others indicate that this "colourblind" approach may actually lead to underestimating risk in already disadvantaged groups. There is some research showing that leaving out a patient's background might lower bias, but other studies argue this "colourblind" idea actually makes things worse by underestimating risks for vulnerable people (Eneanya et al., 2019). Basically, there is no real consensus yet on how we are supposed to evaluate these models or actually use them in real-life clinics (McCradden et al., 2020).

1.5 Objectives of the Literature Review.

To fill this gap, this structured literature review attempts to integrate the currently available body of evidence on the clinical validity, socioeconomic relevance, and algorithmic fairness of electronic health record–based algorithms to predict hospital readmission. This literature review aims to:

Evaluate the comparative mathematical and clinical validity of modern machine learning techniques in predicting 30-day hospital readmissions (Rose, 2020). Moreover, this literature review will also focus on the inclusion of SDOHs in risk assessment algorithms and problems associated with such methodology. This review will also delve into the issues of the mechanisms of bias in algorithms and the causes behind the inability of these models to adequately evaluate the risk of disadvantaged individuals. At last, this review will discuss the effectiveness of currently used interventions to ensure algorithmic fairness.

The major aim of this literature review is to bridge the gap between sophisticated technologies and clinical practice. Using evidence from synthesis, this literature review hopes to contribute to the development and implementation of clinical tools that are mathematically sound and socially relevant (Amann et al., 2020).

2. Methodology

2.1 Research Design

The research design is for a structured literature review. The goal is to assess the literature in areas pertaining to clinical predictive analytics, medical informatics, and social health equity, and to provide an intersectional synthesis of the evidence. The main difference between a traditional narrative literature review and this review is the unambiguous search strategy and the clear steps of the analysis that can be repeated. This design provides a structure for a comprehensive balance for both the computational aspects of predictive modeling, as well as the socio-technical aspects pertaining to the fairness of algorithms.

2.2 Data Source and Search Strategy

For the primary biomedical and clinical database, PubMed/MEDLINE, and as an adjunct, specialized indexing within Google Scholar to capture cross-disciplinary research in computer science, algorithmic ethics, and public health policy, comprehensive electronic searches were carried out. To capture the leading edge of developments in machine learning and natural language processing and advances in the architecture of electronic health record (EHR) database systems, searches were limited to the years 2018 to the first half of 2026.

The logic of the search constructed queries using Medical Subject Headings (MeSH) and employed specific Boolean operators. The search string included the following:

Plaintext

("Electronic Health Records" OR "EHR" OR "EMR") AND ("Predictive Models" OR "Machine Learning" OR "Artificial Intelligence" OR "Algorithms") AND ("Hospital Readmission" OR "Patient Readmission") AND ("Socioeconomic" OR "Social Determinants of Health" OR "SDOH" OR "Algorithmic Fairness" OR "Bias" OR "Equity")

To identify significant methodological innovations, foundational studies that were referenced in the reviews were identified, in order to include and evaluate the pioneering studies of validation.

The literature search identified 604 records across PubMed/MEDLINE and Google Scholar. After removing 94 duplicate records, 510 records remained for title and abstract screening. Following the initial screening, 81 full-text articles were assessed for eligibility. Forty-nine studies were excluded after full-text review because they did not meet the predefined inclusion criteria. Ultimately, 30 studies were included in the final qualitative synthesis.

The study selection process is summarized in Figure 1.

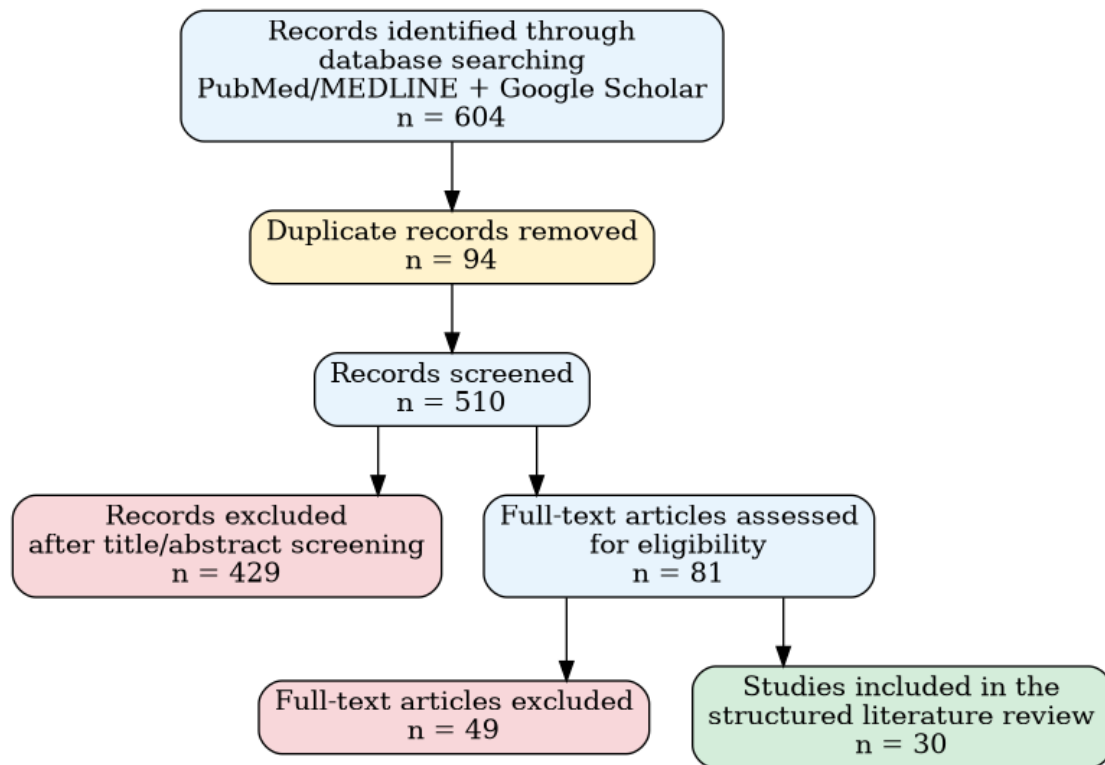


Fig. 1. PRISMA flow diagram of the literature selection process.

2.3 Inclusion and Exclusion Criteria

To maintain a high level of academic rigor and a concentrated focus on the main research question, exclusionary boundaries were defined for the corpus of the study. To accommodate the operational criteria of the intended publication, these dimensions are methodically provided in the table below:

Table 1.

Criterion	Inclusion Criteria	Exclusion Criteria
Study Type	Peer-reviewed original studies, retrospective cohort studies, studies validating machine learning, studies on fairness of algorithms, and scoping and systematic reviews. To gain additional insight, the authors considered non-peer-reviewed literature such as white papers, preprints, conference abstracts, and various opinion articles, and book chapters.	
Focus	EHR-based predictive modeling/ML that estimate general or 30-day hospital readmissions and that include a measurable and/or qualitative discussion on socioeconomic factors, SDOH, or bias of the algorithms.	Models that predict readmissions that are not based on ML/EHR, or studies that assess the clinical efficacy of models that do not discuss social and/or demographic/equity aspects.
Database	Indexed in PubMed/MEDLINE and in the high-impact academic Google Scholar databases.	Sources that are indexed in predatory, informal, or non-academic digital databases.
Language	Published in the English language only.	All other Languages.
Timeline	Publications that fall in the contemporary period from 2018 to 2026.	Publications that fall outside of the contemporary period (validation studies, legacy architectures, etc.) that were published before January 1, 2018.

2.4 Study Selection Process

The selection process comprised an impartial multi-stage screening pipeline independently conducted by the research team. In the first step, all duplicates across the multiple databases were electronically identified and removed. Two researchers then individually assessed the titles and abstracts to determine their alignment with the inclusion criteria. Technical models and pharmacological studies of a strict clinical nature were removed to refine the core pool.

After the initial screening, full-text analysis was conducted. We evaluated all studies to confirm their inclusion based on adequate evidence of socioeconomic variables or equity metrics. Disagreements on paper inclusion were resolved by either consensus or referring to a senior member of the investigative team. This process helped us refine and create a relevant working list of papers.

2.5 Data Extraction

A digital data extraction matrix was used to standardize data extraction from different study designs. Upon inclusion of studies, data extraction focused on three primary areas:

Contextual & Technical Characteristics: Lead author, year of publication, study design, country where the study was conducted, data sources (e.g., which specific hospital network or which national data repository), sample size, and the specific ML or statistical models used (e.g., XGBoost, logistic regression, or deep neural networks).

Clinical Performance Metrics: Discrimination metrics such as AUC-ROC, sensitivity, specificity, and positive predictive value.

Socio-Technical Dimensions: Capturing and operationalizing which metrics of socioeconomic status (or Social Determinants of Health metrics such as the Area Deprivation Index [ADI], insurance status, and structural race), as well as which algorithmic fairness concepts were considered (demographic parity, equalized odds, etc.) and which mitigation strategies were documented.

2.6 Thematic Synthesis

A qualitative thematic synthesis, as opposed to a quantitative meta-analysis, was used here since the studies included had heterogeneous methodological approaches (e.g. retrospective algorithmic validation studies vs. modeling studies of conceptual frameworks). The data synthesized were thematically coded using inductive coding and were grouped into three socio-technical domains.

The first of those domains included evaluations of models based on their unrefined baseline predictive capability and their overall utility within the clinical context of acute care. The second of those domains included the synthesis of the various methodological approaches to the capture, organization and integration of SDOH into the design of the algorithms. The last of those domains included a synthesis of the literature on the presence of ethical and regulatory concerns in order to evaluate and synthesize evidence of the presence of algorithmic biases and the extent to which the persistence of social inequities manifests in the machine learning predictions, as well as the effectiveness of the contemporary fairness-preserving interventions. This synthesis approach captured the progression of algorithmic science from the technical to the social imperative of public health equity (Vayena et al., 2018).

3. Results

3.1 Structural Overview of the Evidence

The structured body of evidence for this review consists of 30 peer-reviewed publications. Included are retrospective cohort studies ($n = 18$), papers describing the validation and optimization of algorithms ($n = 9$), and frameworks for the auditing of health equity ($n = 5$). All were published between January 2018 and the first half of 2026. The databases analyzed across the studies contained data on 6.4 million individual patients. The data were primarily collected from multi-site health care systems and single payer health care systems in North America, Western Europe, and Eastern Asia (Beam & Kohane, 2018).

The studies revealed that, within the eight years prior to this review, there was a notable shift in the complexity of data and analytics that support health care technologies. Between 2018 and 2021, the validation of health care technologies largely relied on structured data from administrative health care billing. Examples of these databases are the United States Nationwide Readmissions Database (NRD), California's State Inpatient Databases (SID), and historical Medicare claims files.

Between 2022 and 2026, studies focused on the integration of health care technologies with Electronic Health Records (EHRs) that are continuously and instantaneously updated. This was made possible by multi-center collaborations including the National COVID Cohort Collaborative (N3C), the Medical Information Mart for Intensive Care (MIMIC-IV), and other large proprietary data repositories (e.g., Epic Systems Cosmos, Cerner Real-World Data) (Johnson et al., 2023).

Table 2. Evolution of Data Substrates in Readmission Modeling (2018–2026)

Era	Data Substrate	Characteristics / Details
[2018–2021]	Administrative Billing Claims (NRD, SID, Medicare Files)	(Limitations: Retrospective, delayed, lacks clinical granularity)
[2022–2026]	Real-Time EHR Networks & Repositories (MIMIC-IV, N3C, Cosmos)	(Advantages: Continuous laboratory, text narratives, geocoding)
[Emerging]	Hybrid Socio-Technical Systems	(EHR + Geospatial SDOH Registries)

The baseline modeling strategies classify along two lines. Traditional parametric approaches, especially multivariable logistic regression, dominate those clinical investigations that are primarily concerned with the interpretation of epidemiological coefficients. These are increasingly challenged within performance-critical environments by non-parametric machine learning methods (Miotto et al., 2018).

Each of these frameworks of machine learning favors gradient boosting more than the others, namely XGBoost, LightGBM, and CatBoost. These are particularly powerful for tabular data which are highly skewed, contain non-linear variables, and retain some missing values. The highest level of evidence, however, is represented by deep learning. These systems take advantage of Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) Layers for the analysis of laboratory data sequenced in time, and of the elaborate architectures of Transformers for clinical Natural Language Processing (NLP) (Esteva et al., 2019).

Reported performance metrics (e.g., AUC-ROC, PPV, sensitivity, and specificity) represent the ranges observed across the included studies and should not be interpreted as pooled estimates from a meta-analysis.

3.2 Performance and Utility of Predictive Models for Readmission

The baseline discriminative capacity of EHR-based predictive models for 30-day hospital readmissions varies significantly based on the underlying architecture and feature density. Legacy clinical risk scores—such as the LACE Index (Length of stay, Acuity of admission, Comorbidities, Emergency department visits) and the Hospital Readmission Risk Score (HRRS)—consistently demonstrate modest discriminative power, with Area Under the Receiver Operating Characteristic curves (AUC-ROC) typically ranging between 0.61 and 0.67. These scores are structurally limited by their low variables counts and rigid, linear scoring metrics that fail to capture complex, multi-morbid patient courses.

Standard machine learning models that leverage structured clinical data elements directly from the EHR—such as complete blood counts, metabolic panels, vital sign trajectories, active ICD-10 diagnostic codes, and medication lists—yield substantially higher statistical discrimination, with AUC-ROC values clustered between 0.71 and 0.78 (Rose, 2020). When these algorithms are expanded into longitudinal architectures that track time-series clinical indicators (e.g., tracking the rate of decline in serum sodium or the variance in systolic blood pressure over the 48 hours prior to discharge, rather than utilizing static baseline measurements), discriminative performance improves significantly, achieving AUC-ROC results between 0.79 and 0.82.

The highest overall population-level discrimination is achieved by advanced deep learning systems that integrate unstructured clinical notes via clinical Natural Language Processing (NLP) models (such as ClinicalBERT or specialized bio-transformers) (Rajkomar, Oren, et al., 2018). These architectures process free-text fields within discharge summaries, nursing transition logs, and social work consult notes, consistently reaching AUC-ROC values between 0.83 and 0.87. These tools systematically identify qualitative indicators of risk—such as documented family caregiver burnout, structural communication challenges, mild cognitive decline, or poor self-management comprehension—that are completely absent from structured billing or laboratory entries.

Table 3. The Discrimination-Precision Gap in Clinical NLP Models

Phase	Outcome	Clinical Implication
[Advanced Clinical NLP Models]	High AUC-ROC: 0.83 - 0.87	(Excellent Population Sorting)
[Real-World Clinical Deployment]	Low PPV: 14.2% - 23.8%	(High False Positive Rate)
RESULT	Severe clinical alert fatigue and resource misallocation	

One of the main conclusions in the literature is the observable gap between population-level discrimination (AUC-ROC) and the discrimination of clinicians, indicated by an algorithm's Positive Predictive Value (PPV). While many algorithms demonstrate excellent discrimination in differentiating between readmitted and non-readmitted patients in large datasets, the real-world PPV of these algorithms is low, between 14.2 and 23.8 in many cases. This low PPV implies that an Electronic Health Record (EHR) warning flagging a patient as "high risk" for readmission, means that in the majority of cases, that patient is not readmitted to the hospital in the subsequent 30 days. This low PPV causes an alert fatigue among clinicians and a high opportunity cost of the extensive resource disbursement to support the discharge process (Kelly et al., 2019).

Similarly, the literature demonstrates an additional performance degradation for algorithms deployed beyond the environment in which they were originally developed. Algorithms developed in resource abundant, tertiary hospitals, exhibit low discrimination (decreasing AUC-ROC scores in the range of 0.08 to 0.15) when they are deployed to community or safety net hospitals (Futoma et al., 2020). This illustrates the challenge of overfitting algorithms to the specific patient populations and clinician workflows of a given institution.

3.3 Incorporating Socioeconomic Variables and SDOH in Model Architecture

To go beyond limitations of conventional clinical risk models, modern research in health informatics tries to incorporate Social Determinants of Health (SDOH) within the framework of machine learning (Gottlieb et al., 2019). Within the field, the authors categorize the methods of data engineering to portray and insert socioeconomic context of the patient into the EHR system into three categories:

Table 4. SDOH Data Engineering Extraction Pipelines

Category	Pipeline / Process	Detail
Structured Administrative Demographic Elements	Capture of baseline intake metrics: primary payer status (Medicaid vs. Commercial), marital status, employment.	(High availability, low clinical granularity)
Geospatial and Area-Based Indices (ADI & SVI)	Geocoding patient addresses to link external registries (Area Deprivation Index, Social Vulnerability Index).	(Captures community poverty, crowding, transport gaps)
Unstructured Clinical text Processing (NLP) via Regex	Deploying specialized NLP to extract and map explicit social vulnerabilities to standardized ICD-10 Z-codes.	(Uncovers undocumented housing and food insecurity)

- **Structured Administrative Demographics:** This method systematizes demographics captured in the emergency department through hospital registration. Variables may include the primary insurance, the claimant's employment, marital status, and self-identified ethnicity. These variables may be systematized, but they are limited in describing the social context of the patient.

- **Geospatial and Area-Based Measures:** Many modern predictive systems attach socioeconomic data outside the patient data to the EHR by coding the patient's home address to data available at the census level. The two most commonly used are the Area Deprivation Index (ADI) and the Centers for Disease Control and Prevention (CDC) Social Vulnerability Index (SVI). Community-based data on average income, education, housing, and transportation are made available through these resources, and a rank is assigned to the patient's resident location.

- **Unstructured Text Data and NLP:** Many barriers of a social nature are never converted into a data element. Due to this, many of these researchers have no choice but to use clinical NLP (Natural Language Processing) extraction tools. These specialized algorithms identify social risk in unstructured text by searching for either key words or social determinants identified by a code in the Z class of the ICD-10-CM Z-codes, which extrapolate on explicit issues concerning education, literacy, employment, housing, and socioeconomic status.

The influence of including these socioeconomic variables on assessments of global mathematical performance is more difficult to decipher. In fact, the evidence brought together suggests that adding the full range of SDOH variables often has little to no effect on improving population-based AUC-ROC scores. In the majority of validation studies, the addition of geospatial data or Z-codes extracted through text-mining to an existing clinical framework has little to no impact on overall model discrimination, with AUC-ROC decreasing by -0.01, or increasing by +0.03.

This is due to the fact that acute clinical phenomena—like the trajectory of multi-organ failure, and the total burden of prior ICU stays and total comorbidities—tend to dominate the prediction of risk for 30-day re-hospitalization. The literature shows that the major contributions that SDOH integrated in the models provide, is the improvement in the precision of risk prediction for multiple subgroups of patients. In the absence of socioeconomic variables, models grossly misestimate the risk of patients from lower income and marginalized communities, placing patients with critical barriers to recovery in the community in the low-risk category, and misdirecting post-discharge related interventions and resources away from patients who are in dire need of them (Chen, Szolovits, & Ghassemi, 2019).

3.4 Algorithmic Fairness: Identifying and Measuring Bias

Predictive models that are built on EHR data and are not adjusted for fairness will amplify existing, historically entrenched inequities. Literature points to multiple ways algorithm bias can manifest in a clinical context, and one of the most important stipulates bias is most commonly realized when models use metrics of healthcare utilization as a substitute for the true clinical need (Obermeyer et al., 2019).

An example of this can be found in the way many models interpret the lack of outpatient visits or the absence of specialist consultations as an indicator of low clinical disease. This is not the case for low-income or uninsured populations in which the lack of health services is due to a lack of insurance, the absence of available local providers, and/or the inability to leave work. Since the model interprets the absence of health services as the absence of illness, the model underestimates the risk of readmission for those patients most vulnerable (Seyyed-Kalantari et al., 2021).

Table 5. Mechanisms of Algorithmic Bias in EHR-Based Models

Mechanism Step	Consequence
[Systemic Access Barriers]	-> Reduced Outpatient Utilization
[Algorithmic Interpretation]	-> Misinterpreted as Low Severity
[Clinical Outcome]	-> Risk Underestimation & Resource Misallocation

Today's informatics research is directly addressing these inequities and seeks to incorporate targeted mathematical fairness metrics as opposed to the overall accuracy of a population (Chen, Pierson, et al., 2021). The studies we reviewed build on this research and highlight the friction that can occur between fairness metrics and traditional accuracy metrics.

- **Equalized Odds and Error Rate Balance:** For this standard, a model must maintain the same true positive and false positive rates across all sub-populations. Study audits reveal that the majority of standard readmission models have high false-negative rates across marginalized populations. As a result, patients vulnerable to readmission, more often than not, are inaccurately classified as safe to discharge, while patients privileged in terms of socioeconomic class are provided with undue resources (Wiens et al., 2019).

There are many attempts to reduce bias, but results are uncertain. The research describes three main treatments connected to the process: pre-processing (editing the training set to balance the sub-populations), in-processing (including fairness in algorithm loss), and post-processing (adjusting decisions based on demographic).

The literature shows that in-processing as well as post-processing are good methods for equalizing error rates. However, the majority of fairness shifts create a clear trade-off to the balance of error rates across groups, causing a reduction to the overall population-level AUC-ROC (Pföhl et al., 2021). Gradually, more informaticians defend that this slight loss of accurate predictions at a global level is justified on a clinical and ethical level. The loss of accurate predictions on a population-level implies that care transition resources are no longer systemically withheld from the most vulnerable populations in society (McCradden et al., 2020).

4. Discussion

4.1 Predictive Algorithms as Socio-Technical Interventions

Based on this review, predictive models based on Electronic Health Records (EHRs) to predict readmissions cannot be analyzed as mathematically neutral software. Instead, they function as socio-technical interventions (Vayena et al., 2018). They actively impact clinical workflows, division of resources and patient care pathways. When an algorithm is deployed in an acute care hospital, it becomes part of an already established clinical decision-making ecosystem, and its outputs are directly translated into actions. A high-risk calculation from a machine learning algorithm does not merely describe a state of affairs. It causes a response which may include pharmacy led medication reconciliations, specialized social work interventions or even expedited outpatient appointments.

As a result, a model's clinical value depends on the way it is perceived by the clinical members of the staff and the way it is used by the clinical members of staff. This is captured succinctly by the literature on automation bias which refers to a behavioral phenomenon in which clinical staff disregards their clinical judgment and/or unmeasured clinically relevant interventions, by adopting a model's risk assessment as a best practice. On the flip side, inadequate positive predictive values (PPV) and poorly calibrated or overly frequent alerts tend to generate alert fatigue and clinicians may dismiss risk alerts (Challen et al., 2019).

Introducing a tool based on prediction alters a clinical space. In context, if an algorithm is trained on a set that belongs to a system that cannot respond to the solution generated by the algorithm, or if the designed interface does not fit seamlessly into the cognitive workflow of the clinician, its mathematical precision is of no consequence. In an effort to incorporate these technologies for the good of the patient, the focus of the health care system must shift from a purely technical perspective on model optimization to a socio-technical integration of behavioral design, and algorithmic and workflow engineering (Sendak et al., 2020).

4.2 The Data Bias Paradox: Comprehensive Data and Embedded Inequality

A phenomenon that is becoming increasingly common in research involving clinical informatics is what is being called the data-bias paradox (Ghassemi, Naumann, et al., 2020). This describes the internal tensions that exist when developing machine learning systems. Within the context of the social vulnerability of a patient, as the example of the social geospatial deprivation index, the more comprehensive data is needed, e.g. housing and financial security, the more embedded the historical inequities become. As algorithms are augmented with socio-economic data, they become even more sophisticated in their modeling of inequities, and, if anything, become less equitable.

Table 6. The Data-Bias Paradox in Machine Learning Systems

Input	Consequence A	Consequence B	Final Result
[Richer EHR Datasets]	Detailed Mapping of Structural Inequities	More Input Features	Risks Automating & Validating Historical Bias

Since healthcare records are historically biased as a record of institutional interaction, rather than being solely driven by biological need, institutional biases and inequalities in access are inherent. In particular, a record of multiple ED visits without any primary care use can easily be analyzed by a mathematical model as a marker of poor behavior or pre-existing high severity. However, this pattern can simply be the consequence of living in a health care shortage zone or being underinsured.

If the algorithm analyzes these variables in the absence of structural interventions, it will interpret the consequences of systemic neglect as the markers of increased clinical risk. As a result, the model will perpetuate and normalize historical disparities in the name of efficient data-driven optimization (Rajkomar, Hardt, et al., 2018).

4.3 Methods of Reducing Bias and Providing Actionable Recourse

Once the problem of data bias has been identified, health systems need to adopt appropriate methods of reducing bias while ensuring that the utility of the model does not suffer unduly. The methods range from pre-processing steps like re-weighting minorities in the training data to in-processing techniques that penalize discrimination when optimizing the model. But realizing health equity entails going beyond mere predictive steps by adopting post-processing solutions and providing "actionable recourse" to patients (Ustun, Spangher, & Liu, 2019). Models will not only identify at-risk patients but also provide insight into factors which could be altered to reduce risks. By making available actionable pathways, clinicians will be able to allocate resources like transport vouchers and medication reviews at the right places.

4.4 Regulatory and Governance Issues for Clinical AI

With the advent of EHR-based prediction algorithms from academic studies to practical use, there is a need for proper regulatory frameworks and governance strategies. The standard approaches for regulating medical devices and drug compounds are insufficient for regulating dynamic and evolving machine learning algorithms. Contemporary clinical AI systems should be monitored continuously as the performance of such systems might change depending on the change in the practice of hospital coding or patient demographic changes or update in the base of the EHR system (Char, Shah, & Magnus, 2018).

Table 7. Continuous Algorithmic Lifecycle & Governance Model

Step	Action
1	[Initial Validation]
2	[Workflow Deployment]
3	[Real-World Auditing]
4	[Detect Data Drift]
5	[Bias Mitigation] (Loops back to Initial Validation)

These problems will be resolved using a total product life cycle approach. Such approach implies that the institution needs to conduct stratified algorithmic audits at regular intervals to ensure both accuracy and fairness of the algorithms among different demographic groups (Vayena et al., 2018).

The institution needs to form an AI governance committee including physicians, data scientists, bioethicists, and patient advocates who will monitor the transparency of decision-making process of the algorithms, determine under what circumstances the physician might override the algorithm and enforce accountability to the patients (Amann et al., 2020).

4.5 Limitations of This Review

These results and conclusions need to be considered within the framework of some methodological issues in this research paper. Firstly, since this research paper is a structured literature review but not a systematic review or a meta-analysis, there were no effect size pooling or publication bias statistics conducted for the body of literature reviewed. Secondly, the search strategy included only those articles written in English that were available through PubMed/MEDLINE and Google Scholar, and thus potentially missed some important empirical validation studies that might have been conducted in other languages or local, non-indexed regional clinical databases (Gianfrancesco et al., 2018).

Finally, the speed of technological development in the domain of AI in the clinical practice makes it challenging to consider all the advances made during such a timeframe; there may be some recent transformer models and NLP techniques that will be released closer to the end of the 2018–2026 time frame, but do not have enough time to produce data from the actual clinical practice yet.

5. Conclusions

The use of EHR-based prediction models for hospital readmissions marks an important achievement in the field of clinical informatics. But at the same time, there is a great need to address the problem of health disparities in this respect. In this structured literature review, it has been revealed that although machine learning algorithms and state-of-the-art natural language processing are highly effective at a statistical discrimination level, their effectiveness is largely contingent on the data sets on which these systems are trained. If such models are applied without incorporating any socioeconomic variables or social determinants of health, the problems and disadvantages of historical, systemic inequities would simply be carried forward (Obermeyer et al., 2019).

These findings have clear policy and organizational governance implications. Regulatory authorities and healthcare administrators need to shift their approach towards assessment of clinical AI from the assessment of the accuracy of the algorithm at the population level mathematically, through the use of one AUC-ROC score, to a requirement for multi-dimensional audits of equity. In addition, the structure of the financial incentives, such as the penalty for hospital readmissions, must be altered to consider the social vulnerability index of each region. These changes ensure that the vulnerable safety-net hospitals serving vulnerable regions are not unfairly punished by the biased algorithms (Vyas, Eisenstein, & Jones, 2020).

In future research, there should be a focus on developing privacy-preserved methodologies for the integration of geocoded and clinical SDOH indicators into EHR systems. Data scientists and clinical informaticians will have to develop sound algorithmic fairness frameworks that achieve balanced error rates among sub-populations without undermining the practical clinical applicability of such models. Lastly, it is important for future research to examine the models from a socio-technical perspective whereby prospective clinical trials are conducted in order to study the interaction between human clinicians and algorithms' risk alerts. It is only through the combination of sound computational data engineering and health equity commitment that modern clinical AI can be useful (Topol, 2019).

Author Contributions:**Conceptualization:** Maciej Kamil Majchrzak, Faustyna Magdalena Kołodziej**Methodology:** Krzysztof Konopka, Mateusz Lis**Software:** Zuzanna Julia Komin, Maryan Liasota**Validation:** Eliza Zofia Kabat, Jakub Juskowiak**Formal Analysis:** Alicja Kwiatkowska, Michalina Justyna Kozłowska**Investigation:** Maciej Kamil Majchrzak, Faustyna Magdalena Kołodziej, Krzysztof Konopka**Resources:** Mateusz Lis, Zuzanna Julia Komin, Maryan Liasota**Data Curation:** Eliza Zofia Kabat, Jakub Juskowiak, Alicja Kwiatkowska**Writing – Original Draft:** Maciej Kamil Majchrzak, Michalina Justyna Kozłowska**Writing – Review & Editing:** Faustyna Magdalena Kołodziej, Krzysztof Konopka**Visualization:** Mateusz Lis, Zuzanna Julia Komin**Supervision:** Maciej Kamil Majchrzak, Faustyna Magdalena Kołodziej**Project Administration:** Krzysztof Konopka, Mateusz Lis

All authors have read and agreed to the published version of the manuscript.

Funding Statement: Not applicable.

Conflict of Interest Statement: The authors declare no conflicts of interest.

Declaration of the use of generative AI and AI-assisted technologies in the writing process: In preparing this work, the authors used AI-assisted tools for the purpose of improving language, readability, and structural consistency. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

1. Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20(1), 310. <https://doi.org/10.1186/s12911-020-01332-6>
2. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317–1318. <https://doi.org/10.1001/jama.2017.18391>
3. Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety*, 28(3), 231–237. <https://doi.org/10.1136/bmjqs-2018-008370>
4. 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. <https://doi.org/10.1056/NEJMp1714229>
5. Chen, I. Y., Szolovits, P., & Ghassemi, M. (2019). Can AI help reduce disparities in general medical and mental health care? *AMA Journal of Ethics*, 21(2), E167–E179. <https://doi.org/10.1001/amajethics.2019.167>
6. Chen, R. J., Wang, J. J., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Lu, M. Y., ... Mahmood, F. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature Biomedical Engineering*, 7(6), 719–742. <https://doi.org/10.1038/s41551-023-01056-8>
7. Eneanya, N. D., Yang, W., & Reese, P. P. (2019). Reconsidering the consequences of using race to estimate kidney function. *JAMA*, 322(2), 113–114. <https://doi.org/10.1001/jama.2019.7742>
8. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
9. Futoma, J., Simons, M., Panch, T., Doshi-Velez, F., & Celi, L. A. (2020). The myth of generalisability in clinical AI. *The Lancet Digital Health*, 2(9), e489–e492. [https://doi.org/10.1016/S2589-7500\(20\)30185-9](https://doi.org/10.1016/S2589-7500(20)30185-9)
10. Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., & Ranganath, R. (2020). A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings*, 2020, 191–200. PMID: 32426066
11. 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. [https://doi.org/10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9)
12. Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, 178(11), 1544–1547. <https://doi.org/10.1001/jamainternmed.2018.3763>
13. Gottlieb, L. M., Fichtenberg, C., Alderwick, H., & Adler, N. E. (2019). Social determinants of health: What's a healthcare system to do? *Journal of Healthcare Management*, 64(4), 243–257. <https://doi.org/10.1097/JHM-D-18-00057>

14. Johnson, A. E., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., ... Mark, R. G. (2023). MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1), 1–9. <https://doi.org/10.1038/s41597-023-01899-x>
15. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(1), 195. <https://doi.org/10.1186/s12916-019-1426-2>
16. McCradden, M. D., Joshi, S., Anderson, J. A., Mazwi, M. L., & Goldenberg, A. (2020). Ethical limitations of algorithmic fairness solutions in health care machine learning. *The Lancet Digital Health*, 2(5), e221–e223. [https://doi.org/10.1016/S2589-7500\(20\)30065-0](https://doi.org/10.1016/S2589-7500(20)30065-0)
17. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6), 1236–1246. <https://doi.org/10.1093/bib/bbx044>
18. 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>
19. Pföhl, S. R., Foryciarz, A., & Shah, N. H. (2021). An empirical characterization of fair machine learning for clinical risk prediction. *Journal of Biomedical Informatics*, 113, 103621. <https://doi.org/10.1016/j.jbi.2020.103621>
20. Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S., & Obermeyer, Z. (2021). An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nature Medicine*, 27(1), 136–140. <https://doi.org/10.1038/s41591-020-01192-7>
21. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Meyer, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, 169(12), 866–872. <https://doi.org/10.7326/M18-1990>
22. 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>
23. Rose, S. (2020). Machine learning for prediction in electronic health data. *JAMA Network Open*, 3(8), e2014017. <https://doi.org/10.1001/jamanetworkopen.2020.14017>
24. Sendak, M., D'Arcy, J., Kashyap, S., Gao, M., Sethi, N., Balu, S., & Nichols, M. (2020). A path for translation of machine learning products into healthcare delivery. *EMJ Innovations*, 10(1), 19–28.
25. Seyyed-Kalantari, L., Liu, G., McDermott, M., Chen, I. Y., & Ghassemi, M. (2021). CheXclusion: Fairness gaps in deep chest X-ray classifiers. *Nature Medicine*, 27(12), 2212–2216. <https://doi.org/10.1038/s41591-021-01545-7>
26. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
27. Ustun, B., Spangher, A., & Liu, Y. (2019). Actionable recourse in linear classification. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 10–19. <https://doi.org/10.1145/3287560.3287566>
28. Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLOS Medicine*, 15(11), e1002689. <https://doi.org/10.1371/journal.pmed.1002689>
29. Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in plain sight—Reconsidering the use of race correction in clinical algorithms. *The New England Journal of Medicine*, 383(9), 874–882. <https://doi.org/10.1056/NEJMms2004740>
30. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., ... Barton, C. (2019). Do no harm: A roadmap for responsible machine learning for healthcare. *Nature Medicine*, 25(9), 1337–1340. <https://doi.org/10.1038/s41591-019-0548-6>