A Survey on Optimization and Machine Learning-Based Fair Decision Making in Healthcare

Zequn Chen[a]; Wesley J. Marrero[a*]

[a] Thayer School of Engineering, Dartmouth College, Hanover, NH, USA

***Corresponding author:** Wesley J. Marrero

wesley.marrero@dartmouth.edu

Phone: (603) 646-3457

Address: 15 Thayer Dr, Hanover, NH 03755https://orcid.org/0000-0002-7092-2292

Zequn Chen, M.S.; zequn.chen.th@dartmouth.edu; (603) 646-3457 ; 15 Thayer Dr, Hanover, NH 03755; ORCID: 0009-0000-0693-1250

Wesley J. Marrero, Ph.D.; wesley.marrero@dartmouth.edu; (603) 646-3457; 15 Thayer Dr, Hanover, NH 03755; ORCID: 0000-0002-7092-2292

Abstract

The unintended biases introduced by optimization and machine learning (ML) models are a topic of great interest to medical professionals. Bias in healthcare decisions can cause patients from vulnerable populations (e.g., racially minoritized, low-income, or living in rural areas) to have lower access to resources and inferior outcomes, thus exacerbating societal unfairness. In this systematic literature review, we present a structured overview of the literature regarding fair decision making in healthcare until April 2024. After screening 782 unique references, we identified 103 articles within the scope of our review. We categorize the identified articles into the following three sections: algorithmic bias, fairness metrics, and bias mitigation techniques. Specifically, we identify examples of algorithmic, data, and publication bias as they are

typically encountered in research and practice. Subsequently, we define and discuss the fairness metrics previously considered in the literature, including notions of fairness through unawareness, demographic parity, equal opportunity, and equal odds. Lastly, we summarize the bias mitigation techniques available in the optimization and ML literature by classifying them into pre-processing, in-processing, and post-processing approaches. Fairness in decision making is an emerging field, poised to substantially reduce social inequities and improve the overall well-being of underrepresented groups. Our review aims to increase awareness of fairness in healthcare decision making and facilitate the selection of appropriate approaches under varying scenarios.

Keywords: Fairness, Decision making, Optimization, Machine learning, Systematic literature review

1. Introduction

An increasing number of healthcare researchers and practitioners are leveraging quantitative methodologies to improve decision making. These methodologies aim to achieve desirable resource allocation strategies, testing procedures, or treatment protocols. Optimization and machine learning (ML) models have been proven to be extremely helpful in medical decision making and health policy [1–7]. For example, researchers have been applying optimization models to schedule patients based on their predicted no-show rates; and reinforcement learning has been used for treatment recommendation to ensure effective prescription and low mortality rate [8, 9]. However, such advanced decision-making methods can lead to inequitable outcomes as they do not give sufficient attention to underrepresented groups [10], such as those who are racially minoritized or low-income. Our survey presents a review of fair decision-making techniques and concepts in healthcare, revealing how we can leverage optimization and ML approaches to ensure equitable access to healthcare.

The unintended biases introduced by optimization and ML algorithms are of special interest to practitioners and researchers [11–14]. For example, compared with policies generated for White veterans, medical policies generated by reinforcement learning often offer Black veterans fewer opportunities to receive cardiovascular screenings. Hispanic patients are disproportionately underdiagnosed by convolutional neural networks because they tend to have limited access to healthcare resources and the convolutional neural networks cannot perform satisfyingly with inadequate data [15]. Such bias may cause decision-makers to distribute fewer medical resources to racially minoritized subgroups. Clinicians have utilized machine learning models in Warfarin dosing, which shows superior performance in European patients but unsatisfying outcome in Asian patients [16, 17].

Current literature reviews focus on the biases brought or perpetuated by ML in prediction settings. For example, Ahmad and collaborators show that ML-based predictions often yield fewer satisfying outcomes in underrepresented groups due to their insufficient data [18]. Mishler and coauthors reveal that predictors are sensitive to proportions of different populations, and incorporating fairness definitions can help avoid such issues [19]. In contrast to previous reviews centering on prediction, we focus on fairness within the context of decision making. It is worth noting that our review has some overlaps with Smith et al.'s survey [14]. The main differences between our surveys are: 1) their work focuses on fairness in reinforcement learning exclusively, while our paper explores other in-processing techniques such as mixed-integer programming and stochastic programming; 2) we review different pre-processing techniques, such as fair data transformers and natural language processing; and 3) we consider post-processing methods such as Laplacian smoothing and multi-accuracy approaches. The fair reinforcement learning methods cited in Smith et al. are included in our review for completion purposes [20–23]. However, we refer the interested reader to their review for an in-depth description of these works.

Our review begins by describing our literature search strategy. Then, we portray bias categories commonly encountered in decision making, followed by a summary of fairness metrics. Next, we present bias mitigation techniques to extend the traditional decision-making framework while achieving fair choices. Finally, we outline our survey's contributions and limitations, as well as promising future directions.

2. Search Strategies

For our systematic review, we searched the Google Scholar database for records related to fair decision making in healthcare. The electronic search strategy used the terms "decision making"

and "healthcare", combined with one of the terms in "bias", "fairness", or "equity", and one of the terms in "optimization", "machine learning", "deep learning", "reinforcement learning", "game", or "network". The keyword combinations we used are demonstrated in Figure 1. We included all records mentioning these keywords in the publication title, abstract, or full text. The publication language was restricted to English. The search's last update was in April of 2024.

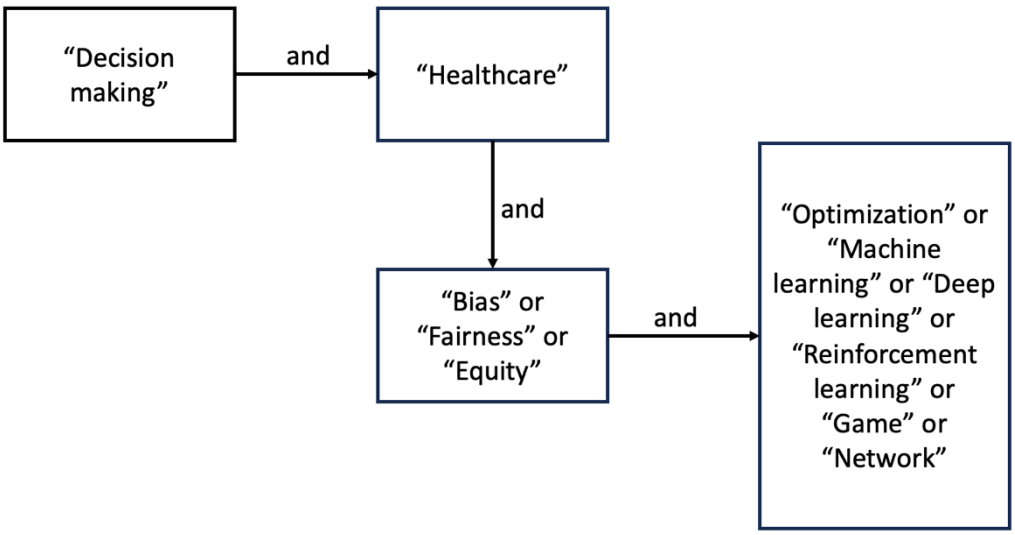

Figure 1: Key words combination for literature review

The titles and abstracts of the articles were screened by one investigator, then the selected manuscripts were double-checked by another researcher. Records were excluded if the publications: 1) did not address healthcare topics and could not be extended to healthcare easily; 2) did not focus on decision making (e.g., papers focusing on predictions); or 3) only included introductory text or conference abstract. Of the remaining publications on methodology or review of bias, fairness metrics, and bias mitigation methods, the full text was screened by investigator before the final discussion with researcher. The articles and surveys are categorized into three sections: bias in healthcare (section 3.1), fairness metrics (section 3.2),

and fair decision-making models and algorithms (section 3.3). Discordance between authors was settled through discussion until the consensus had been achieved.

Fair decision-making concepts (i.e., types of biases, fairness metrics, or bias mitigation approaches) were extracted from the full text of the selected works by investigator. To ensure the accuracy and completeness of the extracted aspects, the selected concepts were double-checked by researchers. All chosen concepts were grouped into section 3.1, section 3.2 or section 3.3 to reflect their relations.

3. Literature Review Results

The systematic review led to 782 records; 208 of them were unrelated and could not be easily transferred to the healthcare domain, and 212 of them did not focus on decision making. Furthermore, 70 records were excluded because they were introductory text or conference abstracts. Of the remaining articles, 153 papers discussed fair decision making in healthcare without a focus on methodology or specific examples of application. In 139 papers left, we excluded 36 additional articles since they did not meet the inclusion criteria described in section 2. Of the remaining 103 papers, 21 were review papers and 82 were original papers. The flow diagram for literature review is shown in Figure 2.

Of the 103 included papers, all concepts related to fair decision-making in healthcare were extracted. We created a structured coverage of the key findings in the following sections. Section 3.1 categorizes biases into three groups: 1) algorithmic bias, 2) data bias, and 3) publication bias. Section 3.2 classifies fairness metrics as one of the following three types: 1) fairness through unawareness, 2) demographic parity, and 3) equal opportunity. Section 3.3

categorizes bias mitigation techniques into three distinct classes: pre-processing, in-processing, and post-processing methodologies.

For the in-processing section, we describe the cutting-edge optimization and ML methods used to alleviate bias. At a high level, there are two mainstream approaches to achieving fairness in these methods: incorporating fairness in objectives or adding fairness-enhancing constraints.

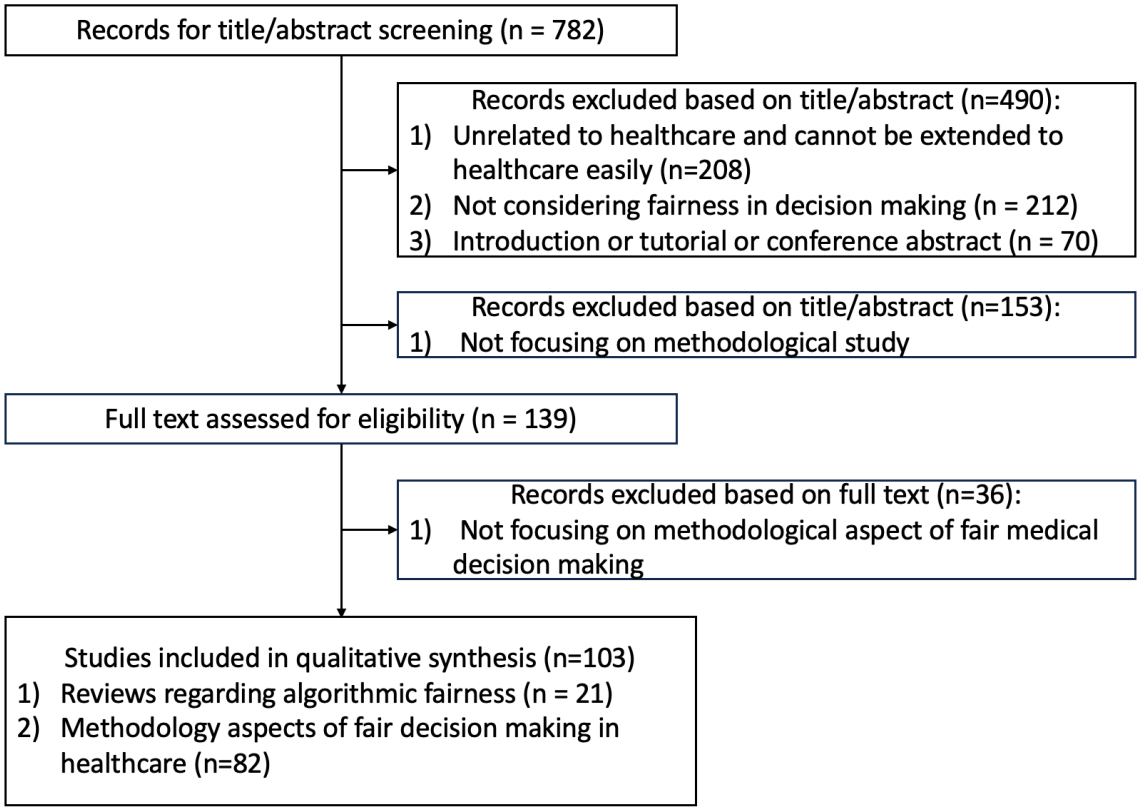

Figure 2: Flow diagram for the systematic literature review of fair decision making in healthcare

3.1 Bias in Healthcare Decision Making

Biases can exist in data and algorithms, which may impede decision-making systems from generating equitable outcomes among subgroups. In this section, we summarize some sources

of bias impacting decision making across healthcare domains. These biases can be categorized into one of the following classes: algorithmic bias, data bias, and publication bias. Section 3.3 of our survey will summarize methods addressing these biases.

3.1.1 Algorithmic Bias

Algorithmic bias stems from computational procedure failing to consider fairness in their execution. This type of bias is a result of improper algorithmic design, which consequently may influence user behavior [24]. For example, optimization-based vaccine allocation algorithms aiming to maximize overall social welfare may exacerbate demographic disparities since underrepresented populations may have less access to vaccines under this objective [25]. Similarly, ambulance allocation models may fail to consider fairness of unit availability across different populations by solely maximizing the overall survival rate. While the survival rate may be high among the entire population, it can be low among patients from vulnerable populations, such as people with lower socioeconomic status. Algorithmic bias also exists in ML models. For instance, Samorani and coauthors have found that ML-based scheduling models have a higher likelihood of assigning Black patients to overbooked slots [26], resulting in worse service experience and longer waiting time.

3.1.2 Data Bias

Data bias refers to the unfairness generated by prejudiced data sources. Unbiased datasets are often necessary for high-quality decision-making model [27]. However, socioeconomic and racial disparities in resource availability may lead to skewed datasets [28]. Two common data biases in healthcare are aggregation biases and representation biases. Aggregation bias refers to the effect of aggregating data without considering disparities among subgroups [29]. For example, hemoglobin A1c level, a widely accepted indicator of diabetes, vary across sex and

ethnicity. If we ignore the subgroup differences in the data and draw conclusions for subpopulations based on the entire population, we may introduce aggregation biases [10, 11, 30, 31]. Representation bias occurs when the data cannot represent the characteristics of all subgroups [32–36]. For instance, providers may have fewer electronic health records (EHR) for people from lower socioeconomic status as they may have limited access to electronic healthcare systems. Hence, we can expect more missing data among people from lower socioeconomic status, which indicates their data cannot demonstrate their overall characteristics. If decision-making models are built with data underrepresenting this population, the developed models may be biased against those with lower socioeconomic status [37, 38].

Another source of data bias is response bias. Response bias occurs when data are labeled inconsistently or collected by unreliable methods. Response bias frequently happens in self-reported data or surveys due to participants' inaccurate answers. For instance, when medical students rate their mental health conditions in surveys, they tend to provide socially acceptable answers. However, such answers often do not align with their true mental conditions as students [39]. Since policymakers may harness data to make public health decisions, response bias can skew decision making [40]. Therefore, models built from data with response bias may underestimate the seriousness of medical students' mental health problems [28].

3.1.3 Publication Bias

Publication bias happens when the researchers' decision to publish a paper depends on the study results. Compared with studies without positive results, publishing medical studies with positive results is generally easier [41]. This phenomenon may lead to the overestimation of certain clinical treatments as evidence against a treatment is not made available. Medical

practitioners may then make treatment decisions based on biased outcomes, giving rise to degraded treatment effects.

An example of publication bias occurred during the COVID-19 pandemic. The academic papers concerning COVID-19 treatment were published rapidly within this period. Most publications showed promising outcomes of COVID-19 treatments, while less satisfying research results only had a slim chance of being published [42]. When related treatments were applied to the general population, many of these treatments produced inferior outcomes compared to publication results [42]. Another example comes from the Cochrane Review on topical benzoyl peroxide, a widely used acne treatment. In 2019, Yang and coauthors found that most benzoyl peroxide studies before 2015 were still unpublished because of their negative results [43]. This finding suggests published literature could not reliably reflect the overall effect of benzoyl peroxide, and medical practitioners applying benzoyl peroxide may not have achieved the desired treatment effects.

3.2 Fairness Metrics

In this section, we present the evaluation of three types of metrics: fairness through unawareness, demographic parity and equal opportunity. To portray the ideas of different fairness metrics, we use a vaccine distribution example. For illustration purposes, we restrict our attention to only one binary sensitive attribute, denoted by A. An example of this type of attribute may be a dichotomized version of race, which includes White and people of color as its categories. The metrics covered by the section are summarized in Table 1.

Table 1: Definition of fairness metrics

	Fairness through unawareness	**Demographic parity**	**Equal opportunity**
Sensitive attribute	No	Yes	Yes
Individual qualification	No	No	Yes
Definition	$X \cap g = \emptyset$	After feeding X to decision-making models, $E(r_{g_1}) = E(r_{g_2})$	After feeding X to decision-making models, within truly qualified patients, $E(r_{g_1}) = E(r_{g_2})$

X denotes entire characteristics of the population, g denotes the sensitive attribute (in our case sensitive attribute is race), X^+ denotes the characteristics of qualified patients, r_{g_1} denotes the cumulated rewards of White patients while r_{g_2} denotes the cumulated rewards of people of color. A qualified subgroup represents a subset of the general population that may be of special interest to decision makers. For example, certain patients, such as senior citizens and those residing in areas with inadequate healthcare resources, are considered qualified patients as they may be more susceptible to the disease.

3.2.1 Fairness Through Unawareness

Fairness through unawareness is the base fairness metric [15]. It does not consider any sensitive attribute during the decision-making process. Within the context of our vaccination distribution example, fairness through unawareness means the model is considered fair if it does not consider race while deciding the vaccination distribution. However, simply ignoring sensitive attributes may not remove inequity, because other variables can be highly correlated with the sensitive traits. This method has been proven to be invalid in many cases [44].

3.2.2 Demographic Parity

The demographic parity fairness metric aims to ensure the expected reward of a decision-making model is independent of sensitive attributes [13]. Independence of sensitive attributes indicates the outcomes (i.e., expected cumulative rewards) must be equivalent in privileged and unprivileged groups [45]. In our vaccine distribution example, demographic parity ensures that, when other information is the same (e.g., age, socioeconomic status), White patients and patients of color have the same expected cumulated rewards. The problem with demographic parity is that it does not consider the population's ground-truth qualifications. For example, suppose patients of color are more vulnerable to a disease; this indicates that these patients have a higher ground-truth qualification and thus should receive a greater expected cumulative reward from the vaccine. If we apply demographic parity in this case, we fail to consider the varied ground-truth qualification across races in decision making.

3.2.3 Equal Opportunity

Similar to demographic parity, equal opportunity verifies whether the expected cumulated rewards for privileged and unprivileged groups are the same [24]. However, demographic parity applies to the entire population, while equal opportunity applies solely to a qualified

population [45]. Within our example, this metric requires that within a qualified population, the cumulated expected reward of receiving vaccine among the unprivileged group ($A = people\ of\ color$) is the same as the cumulated expected reward of receiving vaccine among the privileged group ($A = White$). However, equal opportunity fails to investigate fairness among truly unqualified people.

In summary, fairness through unawareness is the most straightforward metric but is invalid in many settings. Demographic parity evaluates if decision making is independent of sensitive attributes within the entire population. Equal opportunity measures whether decision making is independent of sensitive attributes among a qualified subgroup, which applies to smaller populations compared to demographic parity.

3.3 Bias Mitigation

In this section, we summarize different bias mitigation approaches used across healthcare domains. The methods can be categorized into pre-processing, in-processing, and post-processing. Pre-processing mechanisms clean and manipulate the input data before it is used in decision-making models [46]. In-processing methodologies refer to building unbiased algorithms directly [47]. Post-processing methods calibrate algorithmic outcomes to achieve fairness [48].

3.3.1 Pre-Processing

Datasets may be biased, which can cause skewed decisions. For instance, when a dataset is imbalanced, decisions may be biased toward subpopulations with smaller sizes [49]. Pre-processing methods can help circumvent possible biases in this setting. There are five

commonly used approaches for pre-processing: reweighting the underrepresented populations, resampling, natural language processing, post-survey analysis, and fair data transformers. The identified pre-processing methods, along with their respective reference, areas of application, fairness metrics, and targeted bias categories, are summarized in Table 2.

Table 2: Pre-processing bias mitigation methods

Method(s)	**Reference(s)**	**Area of application**	**Fairness metric**	**Targeted bias**
Reweighting	Nilsson et al. [50]	Medical diagnosis	Demographic parity	Representation bias
	Kumar et al. [51]	Medical diagnosis	Demographic parity	Representation bias
	Peacock et al. [52]	Resource allocation	Demographic parity	Representation bias
Resampling	Chawla et al. [53]	Clinical treatment	Demographic parity	Representation bias
	Mohamed et al. [54]	Medical diagnosis	Demographic parity	Representation bias
	Chawla et al. [53]	Medical diagnosis	Demographic parity	Representation bias
Natural language processing	Minot et al. [55]	Medical diagnosis	Equal opportunity	Representation bias
Post-survey analysis	Serra et al. [56]	Medical diagnosis	Demographic parity	Response bias

Fair data transformer	Biswas et al. [46]	N/A	Demographic parity	Aggregation bias

Reweighting. Reweighting assigns greater weights to underrepresented instances [50]. Biases may be introduced to decision-making tasks if we fail to process underrepresented population's data. Skewed data can also lead to threatening consequences for underrepresented populations. For example, African Americans and Asians have fewer instances in genome studies, which gives rise to higher misclassification rates for the two subgroups in clinical research [11]. Classification methods can investigate the features of each piece of data and use them to decide which category or label the data belongs to. Since physicians rely on classification methods for diagnosis and treatment design, varying misclassification rates for different subgroups can trigger bias in decision making. A remedy for similar issues is to assign greater weights to underrepresented instances, hence data from all populations play an equal role in the modeling process [50]. Kumar and coauthors have shown distributing more weight to underrepresented groups can improve fairness in medical image classification by 8% [51]. However, some reweighting methods, such as inverse propensity score weighting, can potentially increase biases since they calibrate the distributions of all variables simultaneously [57].

Resampling. Resampling is a technique to ensure the data is balanced (i.e., has an near-equal number of instances from each subgroup) by repeatedly drawing samples from the same data [49]. In practice, the majority groups may disproportionately outnumber the remaining groups. Using imbalanced data directly may favor the majority groups while disregarding the minority ones. To avoid this concern, we can resample from the minority groups, so the majority and minority groups have approximately the same size. For example, Chawla and collaborators deploy a synthetic minority resampling technique to decide whether a patient needs diabetes

treatment. Their method successfully shrinks the gap of true positive rates between majority and minority groups [53]. However, medical data (such as EHR) are typically complex, and resampling may lead to overfitting [57]. Researchers and practitioners may avoid overfitting by using cross-validation, which provides a more accurate estimate of a model's performance on unseen data.

Natural language processing. Natural language processing removes biased information from text data before feeding data to decision-making algorithms [58]. Due to the complexity of healthcare data, natural language processing is becoming increasingly popular in fair pre-processing settings. This step ensures the algorithms do not consider sensitive attributes during decision making. For example, Minot and coauthors identify and remove gender-related languages from EHRs by using bidirectional encoder representations. Then, they deploy classification algorithms to evaluate health conditions and give clinical suggestions. Their results show that fairness across genders improves with only a mild degradation in performance [59].

Post-survey analysis. Post-survey analysis for data bias mitigation refers to the process of analyzing survey data after its collection to identify, assess, and correct various types of biases that may have been introduced during the data collection phase [27]. For example, some respondents might misremember their health history; hence the treatment effect for them is likely to be inferior compared to respondents remembering correctly. Researchers can mitigate this by cross-referencing survey responses with medical records or by shortening the recall period to mitigate such bias [56].

Fair data transformers. Fair data transformers extract features from the input data in a fair way. Such transformers modify the input data to achieve fairness [46]. Data transformers, such as principal component analysis, are popular techniques for pre-processing data. In practice, multiple transformers are typically evaluated for a specific fairness metric (e.g., demographic parity). The transformers achieving the fairest output are used to produce inputs for ML and optimization algorithms. Biswas and collaborators have shown that a proper data transformer can significantly improve the fairness of outcomes [46]. Researchers have observed that among data transformers, selecting a subset of features can introduce unfairness [60]. Feature standardization and non-linear transformers are relatively fair transformers, although they can be biased under special conditions such as having too many outliers. These observations indicate that the appropriate transformer must be selected on a case-by-case basis. Though fair data transformers have not been deployed in healthcare to the best of our knowledge, it is easy to extend a fair data transformer to healthcare data preprocessing. For example, in the context of vaccination distribution, we can collect data such as age, sex, population density, and incidence rate in the areas where individuals live. Then, we apply several fair data transformers and feed the transformed data into the same decision-making algorithm. After the algorithm outputs vaccination distribution decisions, we can evaluate the fairness of policies and choose the data transformer that produces the fairest output.

3.3.2 In-Processing

In-processing methodologies incorporate one or more fairness metrics directly in the design of algorithms to lessen biases. These bias mitigation techniques are attracting increased attention within domains such as resource allocation, scheduling, and clinical treatment [26, 61]. Overall, we find six typically used in-processing methods in the literature: mixed-integer programming, stochastic programming, deep reinforcement learning, survival analysis, multi-objective

Markov Decision Process, and constrained Markov Decision Process. Table 3 demonstrates the methods and the corresponding references and applications.

Table 3: In-processing bias mitigation methods

Method(s)	**Reference(s)**	**Area of application**	**Fairness metric**
Mixed-integer programming	Acuna et al. [62]	Resource allocation	Demographic parity
	Lodi et al. [63]	Resource allocation	Equal opportunity
	Radovanović et al. [64]	Resource allocation	Demographic parity
	Ala et al. [65]	Scheduling	Demographic parity
	Zhong et al. [66]	Scheduling	Demographic
	Argyris et al. [67]	Resource allocation	Equal opportunity
	Neophytou et al. [68]	Resource allocation	Demographic parity
	Rastegar et al. [69]	Resource allocation	Equal opportunity
	Wolbeck et al. [70]	Scheduling	Demographic parity
	Gunnarsson et al. [71]	Resource allocation	Demographic parity
	Sepulveda et al. [72]	Resource allocation	Demographic parity
	Klyve et al. [73]	Scheduling	Demographic parity

Method(s)	Reference(s)	Area of application	Fairness metric
	Gross et al. [74]	Scheduling	Demographic parity
	Akshat et al. [75]	Resource allocation	Demographic parity
	Azizi et al. [76]	Scheduling	Demographic parity
	Proano et al. [77]	Scheduling	Demographic parity
	López et al. [25]	Resource allocation	Equal opportunity
Stochastic programming	Ala et al. [78]	Scheduling	Demographic parity
	Yin et al. [79]	Resource allocation	Demographic parity
Deep reinforcement learning	Budhiraja et al. [80]	Scheduling	Equal opportunity
	Yang et al. [21]	Clinical treatment	Demographic parity
	Yu et al. [81]	Clinical treatment	Demographic parity
	Li et al. [82]	Resource allocation	Demographic parity
	Atwood et al. [83]	Resource allocation	Equal opportunity
Fair survival analysis	Keya et al. [84]	Resource allocation	Demographic parity

Method(s)	Reference(s)	Area of application	Fairness metric
Multi-objective Markov Decision Process	Ge et al. [85]	N/A[a]	Equal opportunity
Constrained Markov Decision Process	Ge et al. [20]	N/A[a]	Demographic parity

[a] This method may be easily extended to healthcare settings.

Mixed-integer programming. Mixed-integer programming has been widely used to ensure fairness in decision making [62–77]. Emergency department overcrowding has become a nationwide crisis over the last decade [62]. To resolve the overcrowding issue, researchers have applied mixed-integer programming to build fair medical resource distribution models. Mixed-integer programming is a type of constrained optimization problem that allows for both integer and continuous variables in its objective and constraints [86]. For example, Acuna and coauthors added equity constraints to ensure that the minimal quality of care for every emergency is greater than or equal to a threshold β in an ambulance allocation situation [62]. These constraints guarantee patients suffering from uncommon diseases still receive necessary clinical support. Their equity constraints are demonstrated below:

$$\sum_{j \in J} q_{\{i,j\}} X_{\{i,j\}} \geq \beta, \ \forall i \in I,$$

where I denotes the set of all possible diseases and $i \in I$ denotes disease i. Moreover, J denotes the set of all emergency departments, $j \in J$ refers to the emergency department j, $q_{\{i,j\}}$ is the quality of care for disease i offered by emergency department j, and $X_{\{i,j\}}$ is a binary decision variable. If department j provides the care for disease i, then $X_{\{i,j\}}$ =1,

otherwise $X_{\{i,j\}} = 0$. Lastly, $\beta \in [0,1]$ is selected based on domain experts' suggestions, where 0 denotes the worst quality and 1 the best quality.

Another ubiquitous way to fulfill fairness requirements in healthcare decision making is to modify the objective function of an optimization approach. When medical resources are scarce, people from vulnerable groups may have lower access to them. To ensure fairness towards vulnerable populations, the objective function of an algorithm can be set to maximize the smallest number of allocated resources across all population subgroups [25]. This objective ensures that each subgroup receives their required medical support.

Stochastic programming. Researchers have also leveraged stochastic programming techniques to generate in-processing bias mitigation techniques [78, 79]. These techniques optimize an objective function while representing uncertainty through probability distributions [87]. To optimize patients' waiting time, we can add fair constraints in stochastic optimization models to limit the expected difference between the maximum waiting time and minimum waiting time. The constraint can be formularized as [78]:

$$\max_{n,k} \mathrm{E}\left(W_{\{k\}}^{\{n\}}\right) - \min_{n,k} \mathrm{E}\left(W_{\{k\}}^{\{n\}}\right) \leq \alpha.$$

Here, $k = 1,2, \ldots T$ denotes the time slots when decisions are made, n denotes the n-th patient, and $\alpha \geq 0$ is the threshold suggested by domain experts. The expected waiting time of the n-th patient scheduled to interval k is represented by $E\left(W_{\{k\}}^{\{n\}}\right)$. These constraints guarantee the expected waiting time among all patients does not vary drastically.

Deep reinforcement learning. Reinforcement learning and deep learning play a pivotal role in in-processing methods. Reinforcement learning is a type of ML where the algorithms learn to make decisions by performing actions and observing the results in an environment of interest [88]. Deep learning uses multiple neural network layers and activation functions to extract new features from the input data [89], being capable of recapitulating and modeling complex patterns in data. Deep reinforcement learning is a combination of deep learning and reinforcement learning. Deep reinforcement learning can be used to solve Markov Decision Process (MDP) models. An MDP is a mathematical framework used for modeling decision making in situations where outcomes are partly random and partly under the control of a decision-maker [90]. In practice, an MDP may encompass a massive number of system configurations (i.e., states), becoming computationally intractable by traditional reinforcement learning methods. However, deep reinforcement learning can take advantage of deep learning to represent a policy (i.e., sequence of procedures for decision making at each state) as a neural network and learn to find a policy that optimizes model outcomes (i.e., rewards) [80–83]. Yang et. al redefine the rewards of deep reinforcement learning to achieve fairness [21]. In their approach, the absolute value of rewards of a certain subgroup are smaller if the size of the group is large. The reward function is demonstrated below:

$$R(s_t, a_p, l_p) = \begin{cases} \lambda_p, & if\ a_p = l_p \\ -\lambda_p, & if\ a_p \neq l_p \end{cases}$$

Here, s_t denotes the state at time t, a_p denotes the diagnosis of the model for a person in group p, and l_p denotes the ground-truth disease of the patient from group p. The parameter λ_p is the reward of group p adjusted by its size. Specifically, a positive reward is given if the agent gives the correct diagnosis, and a negative reward is given otherwise. The authors require the

absolute reward for minorities becomes greater than the absolute reward of majorities. This definition of rewards helps the solution approach give more attention to minority groups.

Fair survival analysis. Fair survival models provide an additional tool for decision making in healthcare settings [84]. Traditional survival analysis estimates the time until an event of interest [91]. Fair survival models incorporate event probabilities and fairness violations. The objective of the fair model is below:

$$g(\beta) = -\left(L_{x(\beta)} + \lambda F_{x(\beta)}\right),$$

where $L_{x(\beta)}$ is the log-likelihood of a Cox proportional-hazards model that measures the probability of getting a disease during a certain period and $F_{x(\beta)}$ is the fairness penalty. Moreover, λ is the weight of the fairness penalty in the objective. The difference between the highest and lowest probabilities of disease incidence within a cohort is utilized as the metric for evaluating fairness. Then, they feed the input data to train the model (i.e., learn the parameters β to optimize the objective). The outcome is used to generate a waitlist of patients, which decides the sequence of resource allocation. Their numerical experiment shows the fair survival model can substantially boost the group disease risk range.

Multi-objective Markov Decision Process. While it has not been applied to healthcare settings to the best of our knowledge, the multi-objective MDP is a promising approach to alleviate the potential effect of bias. This model is an extension of the traditional MDP with the difference that the reward function depends on a utility objective and a fairness objective [92]. Ge and coauthors have applied the Pareto frontier to identify the policy that optimizes both utility and fairness elements [85]. The modified reward function is:

$$f_w\big(R(s,a)\big) = w^T R(s,a)$$

where $R(s,a)$ is a reward vector containing rewards r for all objectives after taking the action a at state s, and w is the weight for each objective. They apply reinforcement learning to learn the weight w. Their result shows that there exists a trade-off between utility and fairness performance, and we can choose the final policy based on user preferences [85]. Though multi-objective MDP has not been applied to fair decision making in healthcare to the best of our knowledge, it is possible to deploy these methods to generate fair clinical decisions. For example, if we need to guarantee similar vaccination rates between males and females, we can add this fairness objective into our model. The Pareto frontier can return optimal policies that consider both vaccine utility and distribution fairness.

Constrained Markov Decision Process. The Constrained Markov Decision Process (CMDP) is another prospective direction. Compared to traditional MDP, CMDP can accommodate fair constraints in decision making [93]. In CMDP, we can formulate the cost function regarding fairness, and then choose policies leading to fairness cost less or equal to the threshold [20]. The constraint can be formulated as:

$$E\left[\sum_{t=0}^{\infty} \gamma\, C_t\right] \leq d,$$

where C_t denotes the fairness cost at time t, $\gamma \in (0,1)$ is a discounted factor representing the fairness violations at the current time over the future, and d denotes the threshold for accumulated discounted fairness cost. This model has not been utilized in fair healthcare decision making to the best of our knowledge. However, we can model fairness metrics as

constraints, and choose a set of policies that satisfy these constraints. Afterward, we can investigate which policy in this set gives the optimal discounted accumulated reward.

3.3.3 Post-processing

Post-processing methods calibrate algorithmic outcomes to achieve fairness. We identify the following post-processing methods relevant to healthcare applications: Laplacian smoothing, multi-accuracy approaches, and expert systems. The post-processing bias mitigation methods included in this review are shown in Table 4.

Table 4: Post-processing bias mitigation methods

Method(s)	Reference(s)	Area of application	Fairness metrics
Laplacian smoothing	Petersen F et al. [48]	N/A	Demographic parity
Multi-accuracy	Kim et al. [94]	N/A	Demographic parity
Expert systems	Tang et at. [22]	Clinical treatment	N/A
	Lu et at. [95]	Clinical treatment	N/A
	Tang et at. [22]	Clinical treatment	N/A
	Yu et al. [96]	Clinical treatment	N/A

Laplacian smoothing method. The Laplacian smoothing method is a technique to reduce the noise of the data while preserving the important characteristics of the solution technique [97]. For instance, we can take advantage of this method to guarantee comparable results among similar individuals while preserving the performance (i.e., satisfying loss) of the algorithms [48]. Researchers have shown this technique may improve outcome consistency by approximately significantly. The Laplacian smoothing method can be extended to healthcare. For example, after a reinforcement learning algorithm produces treatment plans, a Laplacian smoothing method can guarantee comparable treatment plans are assigned to similar patients.

Multi-accuracy approaches. We can also apply multi-accuracy approaches to combine several weak learners to achieve high accuracy rates among all subpopulation groups [94]. These methods play a vital role in classification techniques by assigning larger weights to samples identified incorrectly in weak learners. Subsequently, the following weak learners pay extra attention to misidentified samples and adjust their results accordingly [98]. This post-processing technique can improve the accuracy rate of subgroups with the worst classification error, which shows a promising future for complex problems such as population health assessment. With accurate and fair classification for target populations, physicians can deploy algorithm-based treatment design to achieve desirable medical outcomes.

Expert systems. Lastly, the clinical expertise of medical practitioners may help increase the fairness of algorithms [96, 99, 100]. For example, reinforcement learning techniques can suggest several near-equivalent actions, then we can rely on clinicians to decide what actions can lead to the fairest outcome [22]. This approach may enable improved decisions to overcome potential biases while leveraging practitioners' experience.

The distribution of bias mitigation techniques across the identified papers is demonstrated in Figure 3. Moreover, we include the distribution of papers across areas of application in Figure 4.

- Bias mitigation appraoch
  - Pre-processing (9)
    - Reweight (3)
    - Resampling (3)
    - Natural language processing (1)
    - Post-survey analysis (1)
    - Fair data transformer (1)
  - In-processing (27)
    - Mixed integer programming (17)
    - Stochastic programming (2)
    - Deep reinforcement learning (5)
    - Fair survival analysis (1)
    - Multi-objective Markov Decision Process (1)
    - Constrained Markov Decision Process (1)
  - Post-processing (6)
    - Laplacian smoothing (1)
    - Multi-accuracy (1)
    - Expert systems 4)

Figure 3: Distribution of papers in bias mitigation approach

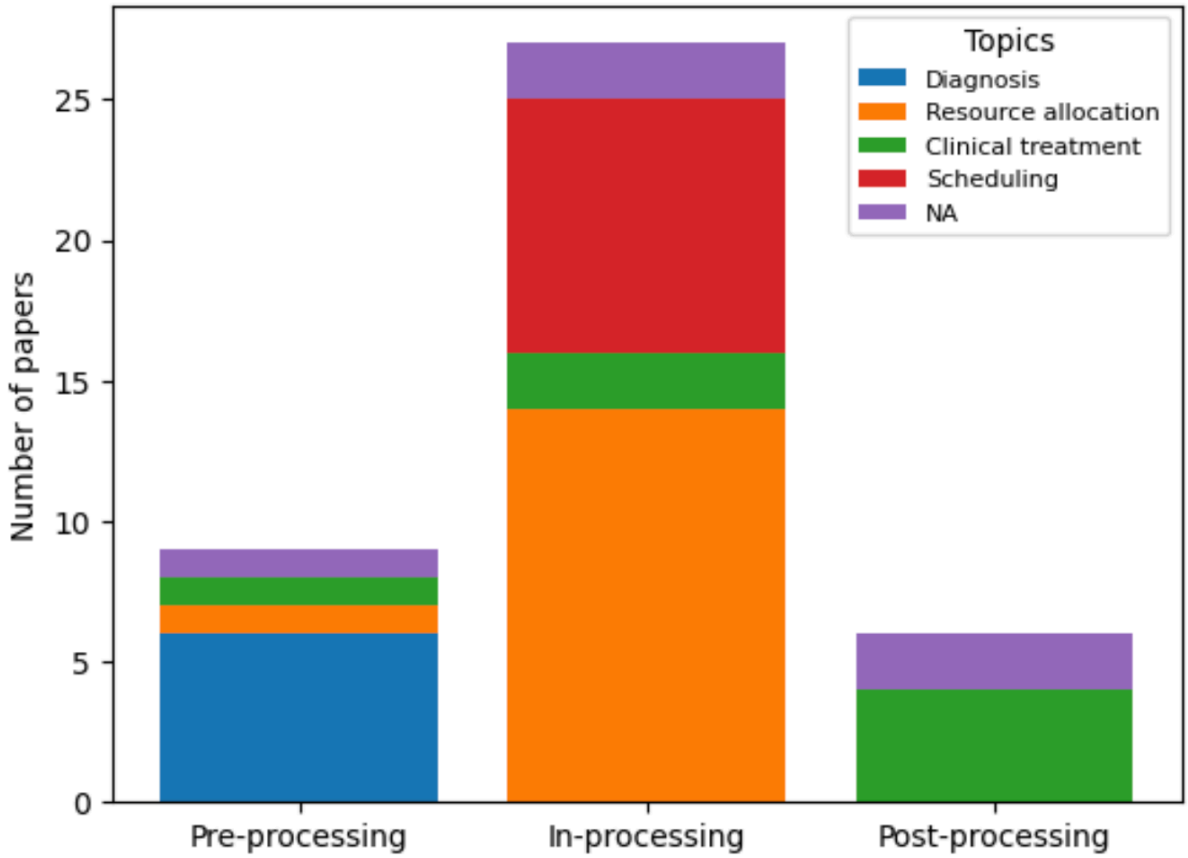

Figure 4: Distribution of papers across topics

4. Conclusion and Future Research Directions

Compared to traditional decision-making techniques, fair decision-making approaches attempt to yield near-optimal and equitable outcomes. This review summarized the state-of-the-art fair decision-making approaches in healthcare settings. We found that even though a plethora of fairness methods has been proposed, most of them focus on prediction rather than decision making, and our survey bridges this gap. First, we presented different categories of biases for data and models. Then, we described multiple fairness metrics that have been used in existing literature. One of the main contributions of our review is that we categorized the literature on decision making in healthcare into pre-processing, in-processing, and post-processing bias mitigation methods. We elaborated on the high-level ideas and examples for methodologies mentioned in our survey. Another important contribution of this systematic review is that we summarized multiple fairness metrics and pointed out their use across applications. Lastly, we explored multiple bias mitigation technologies that have not been applied in healthcare and illustrated how they may be employed in healthcare settings.

Since the most relevant research projects to this review were conducted in the United States, most examples in our paper are cases in this country. Thus, this review may not sufficiently reflect the reality in other parts of the world. Additionally, we might have missed keywords during our literature review, leading to the omission of works that used excluded terms. Finally, we limited our search to articles in English, so we were unable to capture potentially insightful publications in other languages.

Several areas are worth exploring for future research directions. The first promising field is algorithm explainability. Many decision-making algorithms in healthcare are considered black boxes that are hard to understand. The lack of explainability is an obstacle for practitioners to

identify if the model is relying on biased features [89]. Explainable models can resolve this concern since they can reveal underlying structures in a clear way, contributing to removing potential decision biases. Another related emerging field is the combination of interpretability and fairness. Fair interpretable models guarantee the algorithmic outputs align with professionals' instincts. While increased interpretability can win more trust among practitioners, it may hurt the model's fairness. Hence, we need to consider how to strike a balance between fairness and interpretability [101].

Another promising field is the study of context-aware fairness metrics. Researchers have found that different fairness metrics can be incompatible. Thus, we cannot expect a model to satisfy all fairness metrics [98]. In such contexts, it is critical to understand which type of metric we should consider in a specific circumstance. Identifying the best fairness metric for a specific problem will likely require cooperation between modelers and domain experts [102]. Exploring the combination of multiple fairness metrics in decision making is also a potential direction, allowing algorithms to satisfy multiple fairness requirements simultaneously.

It is also worthwhile to bridge the gap between prediction and fair decision making. Current research usually follows a "prediction then optimization" pipeline, but innovative approaches can be explored to incorporate the decision error induced by prediction into the objective function of optimization [103]. These approaches have the potential to achieve fair prediction and optimization simultaneously.

Given the growing importance of decision-making approaches in healthcare, fairness considerations and bias-mitigation approaches are increasingly vital. Our survey may aid practitioners in 1) understanding potential sources of biases in decision making; 2) choosing

the appropriate fairness metric to evaluate decision-making models; and 3) selecting the appropriate pre-processing, in-processing, and post-processing techniques to reduce bias. In conclusion, this survey sheds light on the current state and challenges of fair decision-making in healthcare, highlighting the crucial need for continuous improvement in policies and practices to ensure equitable healthcare outcomes for all individuals.

References

1. Awaysheh A, Wilcke J, Elvinger F, et al (2019) Review of Medical Decision Support and Machine-Learning Methods. Vet Pathol 56:512–525. https://doi.org/10.1177/0300985819829524

2. Brnabic A, Hess LM (2021) Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. BMC Med Inform Decis Mak 21:54. https://doi.org/10.1186/s12911-021-01403-2

3. Zerouaoui H, Idri A (2021) Reviewing Machine Learning and Image Processing Based Decision-Making Systems for Breast Cancer Imaging. J Med Syst 45:8. https://doi.org/10.1007/s10916-020-01689-1

4. Peiffer-Smadja N, Rawson TM, Ahmad R, et al (2020) Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect 26:584–595. https://doi.org/10.1016/j.cmi.2019.09.009

5. Abdalkareem ZA, Amir A, Al-Betar MA, et al (2021) Healthcare scheduling in optimization context: a review. Health Technol 11:445–469. https://doi.org/10.1007/s12553-021-00547-5

6. Wang L, Demeulemeester E (2023) Simulation optimization in healthcare resource planning: A literature review. IISE Trans 55:985–1007. https://doi.org/10.1080/24725854.2022.2147606

7. Ahmadi-Javid A, Jalali Z, Klassen KJ (2017) Outpatient appointment systems in healthcare: A review of optimization studies. Eur J Oper Res 258:3–34. https://doi.org/10.1016/j.ejor.2016.06.064

8. Wang L, Zhang W, He X, Zha H (2018) Supervised Reinforcement Learning with Recurrent Neural Network for Dynamic Treatment Recommendation. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, New York, NY, USA, pp 2447–2456

9. Mizan T, Taghipour S (2022) Medical resource allocation planning by integrating machine learning and optimization models. Artif Intell Med 134:102430. https://doi.org/10.1016/j.artmed.2022.102430

10. Balayn A, Lofi C, Houben G-J (2021) Managing bias and unfairness in data for decision support: a survey of machine learning and data engineering approaches to identify and mitigate bias and unfairness within data management and analytics systems. VLDB J 30:739–768. https://doi.org/10.1007/s00778-021-00671-8

11. Mehrabi N, Morstatter F, Saxena N, et al (2022) A Survey on Bias and Fairness in Machine Learning. ACM Comput Surv 54:1–35. https://doi.org/10.1145/3457607

12. Swain S, Bhushan B, Dhiman G, Viriyasitavat W (2022) Appositeness of Optimized and Reliable Machine Learning for Healthcare: A Survey. Arch Comput Methods Eng 29:3981–4003. https://doi.org/10.1007/s11831-022-09733-8

13. Caton S, Haas C (2024) Fairness in Machine Learning: A Survey. ACM Comput Surv 56:166:1-166:38. https://doi.org/10.1145/3616865

14. Smith B, Khojandi A, Vasudevan R (2024) Bias in Reinforcement Learning: A Review in Healthcare Applications. ACM Comput Surv 56:1–17. https://doi.org/10.1145/3609502

15. Chen RJ, Wang JJ, Williamson DFK, et al (2023) Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng 7:719–742. https://doi.org/10.1038/s41551-023-01056-8

16. Sharabiani A, Bress A, Douzali E, Darabi H (2015) Revisiting Warfarin Dosing Using Machine Learning Techniques. Comput Math Methods Med 2015:e560108. https://doi.org/10.1155/2015/560108

17. Syn NL, Wong AL-A, Lee S-C, et al (2018) Genotype-guided versus traditional clinical dosing of warfarin in patients of Asian ancestry: a randomized controlled trial. BMC Med 16:104. https://doi.org/10.1186/s12916-018-1093-8

18. Ahmad MA, Patel A, Eckert C, et al (2020) Fairness in Machine Learning for Healthcare. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, New York, NY, USA, pp 3529–3530

19. Mishler A, Dalmasso N (2022) Fair when trained, unfair when deployed: Observable fairness measures are unstable in performative prediction settings. ArXiv Prepr ArXiv220205049

20. Ge Y, Liu S, Gao R, et al (2021) Towards Long-term Fairness in Recommendation. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining. Association for Computing Machinery, New York, NY, USA, pp 445–453

21. Yang J, Soltan AAS, Eyre DW, Clifton DA (2023) Algorithmic fairness and bias mitigation for clinical machine learning with deep reinforcement learning. Nat Mach Intell 5:884–894. https://doi.org/10.1038/s42256-023-00697-3

22. Tang S, Modi A, Sjoding M, Wiens J (2020) Clinician-in-the-Loop Decision Making: Reinforcement Learning with Near-Optimal Set-Valued Policies. In: Proceedings of the 37th International Conference on Machine Learning. PMLR, pp 9387–9396

23. Mingyu Lu (2020) Is Deep Reinforcement Learning Ready for Practical Applications in Healthcare? A Sensitivity Analysis of Duel-DDQN for Hemodynamic Management in Sepsis Patients - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8075511/. Accessed 26 Apr 2024

24. Grote T, Keeling G (2022) On Algorithmic Fairness in Medical Practice. Camb Q Healthc Ethics 31:83–94. https://doi.org/10.1017/S0963180121000839

25. Munguía-López ADC, Ponce-Ortega JM (2021) Fair Allocation of Potential COVID-19 Vaccines Using an Optimization-Based Strategy. Process Integr Optim Sustain 5:3–12. https://doi.org/10.1007/s41660-020-00141-8

26. Samorani M, Harris SL, Blount LG, et al (2022) Overbooked and Overlooked: Machine Learning and Racial Bias in Medical Appointment Scheduling. Manuf Serv Oper Manag 24:2825–2842. https://doi.org/10.1287/msom.2021.0999

27. Seker E, Talburt JR, Greer ML (2022) Preprocessing to Address Bias in Healthcare Data. In: Challenges of Trustable AI and Added-Value on Health. IOS Press, pp 327–331

28. Aleem S, Huda N ul, Amin R, et al (2022) Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions. Electronics 11:1111. https://doi.org/10.3390/electronics11071111

29. James LR (1982) Aggregation bias in estimates of perceptual agreement. J Appl Psychol 67:219–229. https://doi.org/10.1037/0021-9010.67.2.219

30. Dhabliya D, Dari SS, Dhablia A, et al (2024) Addressing Bias in Machine Learning Algorithms: Promoting Fairness and Ethical Design. E3S Web Conf 491:02040. https://doi.org/10.1051/e3sconf/202449102040

31. Nazer LH, Zatarah R, Waldrip S, et al (2023) Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health 2:e0000278. https://doi.org/10.1371/journal.pdig.0000278

32. Chang C, Deng Y, Jiang X, Long Q (2020) Multiple imputation for analysis of incomplete data in distributed health data networks. Nat Commun 11:5467. https://doi.org/10.1038/s41467-020-19270-2

33. Furukawa MF, Raghu TS, Shao BBM (2010) Electronic Medical Records, Nurse Staffing, and Nurse-Sensitive Patient Outcomes: Evidence from California Hospitals, 1998–2007. Health Serv Res 45:941–962. https://doi.org/10.1111/j.1475-6773.2010.01110.x

34. Hu Z, Melton GB, Arsoniadis EG, et al (2017) Strategies for handling missing clinical data for automated surgical site infection detection from the electronic health record. J Biomed Inform 68:112–120. https://doi.org/10.1016/j.jbi.2017.03.009

35. Jakobsen JC, Gluud C, Wetterslev J, Winkel P (2017) When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Med Res Methodol 17:162. https://doi.org/10.1186/s12874-017-0442-1

36. Non-clinical influences on clinical decision-making: a major challenge to evidence-based practice - FM Hajjaj, MS Salek, MKA Basra, AY Finlay, 2010. https://journals.sagepub.com/doi/full/10.1258/jrsm.2010.100104. Accessed 22 May 2024

37. Ng JH, Ye F, Ward LM, et al (2017) Data On Race, Ethnicity, And Language Largely Incomplete For Managed Care Plan Members. Health Aff (Millwood) 36:548–552. https://doi.org/10.1377/hlthaff.2016.1044

38. Waite S, Scott J, Colombo D (2021) Narrowing the Gap: Imaging Disparities in Radiology. Radiology 299:27–35. https://doi.org/10.1148/radiol.2021203742

39. IsHak W, Nikravesh R, Lederer S, et al (2013) Burnout in medical students: a systematic review. Clin Teach 10:242–245. https://doi.org/10.1111/tct.12014

40. Gopal DP, Chetty U, O'Donnell P, et al (2021) Implicit bias in healthcare: clinical practice, research and decision making. Future Healthc J 8:40–48. https://doi.org/10.7861/fhj.2020-0233

41. Scherer RW, Dickersin K, Langenberg P (1994) Full Publication of Results Initially Presented in Abstracts: A Meta-analysis. JAMA 272:158–162. https://doi.org/10.1001/jama.1994.03520020084025

42. Raynaud M, Zhang H, Louis K, et al (2021) COVID-19-related medical research: a meta-research and critical appraisal. BMC Med Res Methodol 21:1. https://doi.org/10.1186/s12874-020-01190-w

43. Yang Z, Zhang Y, Mosler EL, et al (2020) Topical benzoyl peroxide for acne. Cochrane Database Syst Rev. https://doi.org/10.1002/14651858.CD011154.pub2

44. Datta A, Tschantz MC, Datta A (2014) Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. ArXiv Prepr ArXiv14086491

45. Wen M, Bastani O, Topcu U (2021) Algorithms for Fairness in Sequential Decision Making. In: Proceedings of The 24th International Conference on Artificial Intelligence and Statistics. PMLR, pp 1144–1152

46. Biswas S, Rajan H (2021) Fair preprocessing: towards understanding compositional fairness of data transformers in machine learning pipeline. In: Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ACM, Athens Greece, pp 981–993

47. Wan M, Zha D, Liu N, Zou N (2023) In-Processing Modeling Techniques for Machine Learning Fairness: A Survey. ACM Trans Knowl Discov Data 17:1–27. https://doi.org/10.1145/3551390

48. Petersen F, Mukherjee D, Sun Y, Yurochkin M (2021) Post-processing for Individual Fairness. In: Advances in Neural Information Processing Systems. Curran Associates, Inc., pp 25944–25955

49. Kamiran F, Calders T (2010) Classification with no discrimination by preferential sampling. Citeseer

50. Nilsson A, Bonander C, Strömberg U, et al (2021) Reweighting a Swedish health questionnaire survey using extensive population register and self-reported data for assessing and improving the validity of longitudinal associations. PLOS ONE 16:e0253969. https://doi.org/10.1371/journal.pone.0253969

51. Kumar N, Shrestha R, Li Z, Wang L (2023) Distributionally Robust Optimization and Invariant Representation Learning for Addressing Subgroup Underrepresentation: Mechanisms and Limitations. In: Wesarg S, Puyol Antón E, Baxter JSH, et al (eds) Clinical Image-Based Procedures, Fairness of AI in Medical Imaging, and Ethical and Philosophical Issues in Medical Imaging. Springer Nature Switzerland, Cham, pp 183–193

52. Peacock WF, Slatkin N, Gagnon-Sanschagrin P, et al (2022) Opioid-Induced Constipation: Cost Impact of Approved Medications in the Emergency Department. Adv Ther 39:2178–2191. https://doi.org/10.1007/s12325-022-02090-9

53. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 16:321–357. https://doi.org/10.1613/jair.953

54. Mohamed R, Azizan NH, Perumal T, et al (2023) Discovering and Recognizing of Imbalance Human Activity in Healthcare Monitoring using Data Resampling Technique and Decision Tree Model. J Adv Res Appl Sci Eng Technol 33:340–350. https://doi.org/10.37934/araset.33.2.340350

55. Minot JR, Cheney N, Maier M, et al (2022) Interpretable bias mitigation for textual data: Reducing genderization in patient notes while maintaining classification performance. ACM Trans Comput Healthc 3:1–41

56. Moreno-Serra R, Anaya-Montes M, León-Giraldo S, Bernal O (2022) Addressing recall bias in (post-)conflict data collection and analysis: lessons from a large-scale health survey in Colombia. Confl Health 16:14. https://doi.org/10.1186/s13031-022-00446-0

57. Borland D, Zhang J, Kaul S, Gotz D (2021) Selection-Bias-Corrected Visualization via Dynamic Reweighting. IEEE Trans Vis Comput Graph 27:1481–1491. https://doi.org/10.1109/TVCG.2020.3030455

58. Zhou B, Yang G, Shi Z, Ma S (2022) Natural Language Processing for Smart Healthcare. IEEE Rev Biomed Eng 1–17. https://doi.org/10.1109/RBME.2022.3210270

59. Joshua R. Minot (2022) Interpretable Bias Mitigation for Textual Data: Reducing Genderization in Patient Notes While Maintaining Classification Performance | ACM Transactions on Computing for Healthcare. https://dl.acm.org/doi/full/10.1145/3524887. Accessed 26 Apr 2024

60. Grgić-Hlača N, Zafar MB, Gummadi KP, Weller A (2018) Beyond Distributive Fairness in Algorithmic Decision Making: Feature Selection for Procedurally Fair Learning. Proc AAAI Conf Artif Intell 32:. https://doi.org/10.1609/aaai.v32i1.11296

61. Uhde A, Schlicker N, Wallach DP, Hassenzahl M (2020) Fairness and Decision-making in Collaborative Shift Scheduling Systems. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. ACM, Honolulu HI USA, pp 1–13

62. Acuna JA, Zayas-Castro JL, Charkhgard H (2020) Ambulance allocation optimization model for the overcrowding problem in US emergency departments: A case study in Florida. Socioecon Plann Sci 71:100747. https://doi.org/10.1016/j.seps.2019.100747

63. Lodi A, Olivier P, Pesant G, Sankaranarayanan S (2024) Fairness over time in dynamic resource allocation with an application in healthcare. Math Program 203:285–318. https://doi.org/10.1007/s10107-022-01904-6

64. Radovanović S, Delibašić B, Marković A, Suknović M (2022) Achieving MAX-MIN Fair Cross-efficiency scores in Data Envelopment Analysis. Proc 55th Hawaii Int Conf Syst Sci

65. Ala A, Alsaadi FE, Ahmadi M, Mirjalili S (2021) Optimization of an appointment scheduling problem for healthcare systems based on the quality of fairness service using whale optimization algorithm and NSGA-II. Sci Rep 11:19816. https://doi.org/10.1038/s41598-021-98851-7

66. Zhong X, Zhang J, Zhang X (2017) A two-stage heuristic algorithm for the nurse scheduling problem with fairness objective on weekend workload under different shift designs. IISE Trans Healthc Syst Eng 7:224–235. https://doi.org/10.1080/24725579.2017.1356891

67. Argyris N, Karsu Ö, Yavuz M (2022) Fair resource allocation: Using welfare-based dominance constraints. Eur J Oper Res 297:560–578. https://doi.org/10.1016/j.ejor.2021.05.003

68. Neophytou N, Taik A, Farnadi G (2024) Promoting Fair Vaccination Strategies through Influence Maximization: A Case Study on COVID-19 Spread. Proc AAAI Conf Artif Intell 38:22285–22293. https://doi.org/10.1609/aaai.v38i20.30234

69. Rastegar M, Tavana M, Meraj A, Mina H (2021) An inventory-location optimization model for equitable influenza vaccine distribution in developing countries during the COVID-19 pandemic. Vaccine 39:495–504. https://doi.org/10.1016/j.vaccine.2020.12.022

70. Wolbeck L, Kliewer N, Marques I (2020) Fair shift change penalization scheme for nurse rescheduling problems. Eur J Oper Res 284:1121–1135. https://doi.org/10.1016/j.ejor.2020.01.042

71. Gunnarsson B, Björnsdóttir KM, Dúason S, Ingólfsson Á (2023) Locating helicopter ambulance bases in Iceland: efficient and fair solutions. Scand J Trauma Resusc Emerg Med 31:70. https://doi.org/10.1186/s13049-023-01114-9

72. Sepúlveda IA, Aguayo MM, De la Fuente R, et al (2024) Scheduling mobile dental clinics: A heuristic approach considering fairness among school districts. Health Care Manag Sci 27:46–71. https://doi.org/10.1007/s10729-022-09612-5

73. Klyve KK, Senthooran I, Wallace M (2023) Nurse rostering with fatigue modelling. Health Care Manag Sci 26:21–45. https://doi.org/10.1007/s10729-022-09613-4

74. Gross CN, Brunner JO, Blobner M (2019) Hospital physicians can't get no long-term satisfaction – an indicator for fairness in preference fulfillment on duty schedules. Health Care Manag Sci 22:691–708. https://doi.org/10.1007/s10729-018-9452-8

75. Akshat S, Gentry SE, Raghavan S (2024) Heterogeneous donor circles for fair liver transplant allocation. Health Care Manag Sci 27:20–45. https://doi.org/10.1007/s10729-022-09602-7

76. Azizi S, Aygül Ö, Faber B, et al (2023) Select, route and schedule: optimizing community paramedicine service delivery with mandatory visits and patient prioritization. Health Care Manag Sci 26:719–746. https://doi.org/10.1007/s10729-023-09646-3

77. Proano RA, Agarwal A (2018) Scheduling internal medicine resident rotations to ensure fairness and facilitate continuity of care. Health Care Manag Sci 21:461–474. https://doi.org/10.1007/s10729-017-9403-9

78. Ala A, Simic V, Pamucar D, Tirkolaee EB (2022) Appointment Scheduling Problem under Fairness Policy in Healthcare Services: Fuzzy Ant Lion Optimizer. Expert Syst Appl 207:117949. https://doi.org/10.1016/j.eswa.2022.117949

79. Yin X, Büyüktahtakın İE (2021) A multi-stage stochastic programming approach to epidemic resource allocation with equity considerations. Health Care Manag Sci 24:597–622. https://doi.org/10.1007/s10729-021-09559-z

80. Budhiraja I, Kumar N, Tyagi S (2021) Deep-Reinforcement-Learning-Based Proportional Fair Scheduling Control Scheme for Underlay D2D Communication. IEEE Internet Things J 8:3143–3156. https://doi.org/10.1109/JIOT.2020.3014926

81. Yu G, Siddique U, Weng P (2023) Fair Deep Reinforcement Learning with Preferential Treatment. In: Gal K, Nowé A, Nalepa GJ, et al (eds) Frontiers in Artificial Intelligence and Applications. IOS Press

82. Li Y, Mao C, Huang K, et al (2023) Deep Reinforcement Learning for Efficient and Fair Allocation of Health Care Resources. ArXiv Prepr ArXiv230908560

83. Atwood J, Srinivasan H, Halpern Y, Sculley D (2019) Fair treatment allocations in social networks. ArXiv Prepr ArXiv191105489

84. Keya KN, Islam R, Pan S, et al (2020) Equitable allocation of healthcare resources with fair cox models. ArXiv Prepr ArXiv201006820

85. Ge Y, Zhao X, Yu L, et al (2022) Toward Pareto Efficient Fairness-Utility Trade-off in Recommendation through Reinforcement Learning. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. Association for Computing Machinery, New York, NY, USA, pp 316–324

86. Achterberg T, Wunderling R (2013) Mixed Integer Programming: Analyzing 12 Years of Progress. In: Jünger M, Reinelt G (eds) Facets of Combinatorial Optimization: Festschrift for Martin Grötschel. Springer, Berlin, Heidelberg, pp 449–481

87. Fouskakis D, Draper D (2002) Stochastic Optimization: a Review. Int Stat Rev 70:315–349. https://doi.org/10.1111/j.1751-5823.2002.tb00174.x

88. Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. MIT Press, Cambridge, Mass

89. Chakraborty S, Tomsett R, Raghavendra R, et al (2017) Interpretability of deep learning models: A survey of results. In: 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). pp 1–6

90. Puterman ML (2014) Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons

91. Kleinbaum DG, Klein M (1996) Survival analysis a self-learning text. Springer

92. Roijers DM, Vamplew P, Whiteson S, Dazeley R (2013) A survey of multi-objective sequential decision-making. J Artif Intell Res 48:67–113

93. Altman E (2021) Constrained Markov Decision Processes. Routledge, New York

94. Kim MP, Ghorbani A, Zou J (2019) Multiaccuracy: Black-Box Post-Processing for Fairness in Classification. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. Association for Computing Machinery, New York, NY, USA, pp 247–254

95. Lu M, Shahn Z, Sow D, et al (2020) Is deep reinforcement learning ready for practical applications in healthcare? A sensitivity analysis of duel-DDQN for hemodynamic management in sepsis patients. American Medical Informatics Association, p 773

96. Yu C, Liu J, Nemati S, Yin G (2023) Reinforcement Learning in Healthcare: A Survey. ACM Comput Surv 55:1–36. https://doi.org/10.1145/3477600

97. Field DA (1988) Laplacian smoothing and Delaunay triangulations. Commun Appl Numer Methods 4:709–712. https://doi.org/10.1002/cnm.1630040603

98. Jo N, Tang B, Dullerud K, et al (2023) Fairness in Contextual Resource Allocation Systems: Metrics and Incompatibility Results. Proc AAAI Conf Artif Intell 37:11837–11846. https://doi.org/10.1609/aaai.v37i10.26397

99. Tang S, Modi A, Sjoding MW, Wiens J Clinician-in-the-Loop Decision Making: Reinforcement Learning with Near-Optimal Set-Valued Policies

100. Lu M, Shahn Z, Sow D, et al Is Deep Reinforcement Learning Ready for Practical Applications in Healthcare? A Sensitivity Analysis of Duel-DDQN for Hemodynamic Management in Sepsis Patients

101. Garcia G-GP, Steimle LN, Marrero WJ, Sussman JB (2024) Interpretable Policies and the Price of Interpretability in Hypertension Treatment Planning. Manuf Serv Oper Manag 26:80–94. https://doi.org/10.1287/msom.2021.0373

102. Xu J, Xiao Y, Wang WH, et al (2022) Algorithmic fairness in computational medicine. eBioMedicine 84:. https://doi.org/10.1016/j.ebiom.2022.104250

103. Elmachtoub AN, Grigas P (2022) Smart "Predict, then Optimize." Manag Sci 68:9–26. https://doi.org/10.1287/mnsc.2020.3922