HHS Public Access

Author manuscript

Circ Cardiovasc Imaging. Author manuscript; available in PMC 2025 February 20.

Published in final edited form as:

Circ Cardiovasc Imaging. 2024 February ; 17(2): e015495. doi:10.1161/CIRCIMAGING.123.015495.

Machine Learning and Bias in Medical Imaging: Opportunities and Challenges

Amey Vrudhula, BSE[1,2], **Alan C Kwan, MD**[2], **David Ouyang, MD**[*,2,3], **Susan Cheng, MD, MPH**[*,2]

[1.]Icahn School of Medicine at Mount Sinai, New York

[2.]Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center

[3.]Division of Artificial Intelligence in Medicine, Department of Medicine, Cedars-Sinai Medical Center

Abstract

Bias in healthcare has been well-documented and results in disparate and worsened outcomes for at-risk groups. Medical imaging plays a critical role in facilitating patient diagnoses but involves multiple sources of bias including factors related to access to imaging modalities, acquisition of images, and assessment (i.e. interpretation) of imaging data. Machine learning (ML) applied to diagnostic imaging has demonstrated the potential to improve the quality of imaging-based diagnosis and the precision of measuring imaging-based traits. Algorithms can leverage subtle information not visible to the human eye to detect underdiagnosed conditions while mitigating cognitive bias in interpretation or derive "new phenotypes" of disease that link imaging features with clinical outcomes. Importantly, the application of ML to diagnostic imaging has the potential to either reduce or propagate bias. Understanding the potential gain as well as the potential risks requires an understanding of how and what ML models learn. Common risks of propagating bias can arise from unbalanced training, sub-optimal architecture design or selection, and uneven application of models. Notwithstanding these risks, ML may yet be applied to improve gain from imaging across all three A's (access, acquisition, and assessment) for all patients. In this review, we present a framework for understanding the balance of opportunities and challenges for minimizing bias in medical imaging, how ML may improve current approaches to imaging, and what specific design considerations should be for introducing ML in an equitable manner to maximize the quality of healthcare for all.

Keywords

Machine Learning; Artificial Intelligence; Imaging; Bias; Health Equity; Disparities

Correspondence: David Ouyang, MD, david.ouyang@cshs.org, Phone: (310)-423-2726, Fax: (310)-423-3522; Susan Cheng, MD, MPH, susan.cheng@cshs.org, Phone: (310) - 423-2726, Fax (310) - 423-3522.
[*]Equal contribution

Disclosures
A.V. reports no disclosures. A.C.K. reports consulting fees from InVision and grant support from AHA 23CDA1053659 and the KL2 sub-award from NIH UL1TR001881. D.O. reports consulting fees from AstraZeneca, Ultromics, EchoIQ, InVision, and research grants from NIH K00-HL157421 and Alexion Pharmaceuticals. S.C. reports consulting fees from UCB and Viz.ai and research grants from NIH R01-HL131532 and NIH R01-HL142983.

Bias in healthcare is a pervasive problem and contributes to variations in process, delivery, and outcomes across the care spectrum. In the context of patient care, data-rich diagnostics provided by medical imaging can play a pivotal role in offering relatively balanced and objective information on the health status of tissues, organs, and systems. However, medical imaging workflows can also be subject to multiple sources of bias – many of which could be mitigated by artificial intelligence (AI) and machine learning (ML). Conversely, some disparities may be perpetuated by the AI/ML driven workflows if applied without thoughtful planning (Figure 1 and Figure 2). In this review, we outline how AI/ML approaches to conventional clinical care workflows offer both opportunities and challenges regarding their potential effects on mitigating bias. Because medical imaging represents a forefront area of AI innovation in healthcare, ongoing lessons can inform our understanding of AI applications in other areas of medical care and their potential influence on bias.

II. Bias in Medicine and Mitigation by Machine Learning

Biases propagate disparities based on socio-demographic traits such as age, sex, gender, race, ethnicity, sexual orientation, socioeconomic status, education, profession, and type of health insurance. Well-documented examples of racial bias in emergency care include the trend that Black adults presenting with acute pain syndromes are less likely to receive adequate analgesia than non-Black adults.[1] Gender bias between cis-gender patients is similarly well-documented: women experiencing chest pain are less likely to be considered for aggressive cardiovascular care and management compared to men.[2] Medical bias can also manifest from more subtle differences in individual characteristics: patients suffering heart attacks in the two weeks after their 80th birthday were less likely to be offered bypass surgery compared to their 79 year-old counterparts.[3] Human medical decision making relies on general impressions and fallible heuristics that can introduce bias. Even risk scores and decision aids, meant to balance ease of use with risk evaluation (e.g. CHA_2DS_2-VASc), rely on a small set of input variables weighted routinely by clinicians. Less restricted computer algorithms could both improve performance and minimize bias when calculating the relationship between clinical variables and outcomes risk for informing treatment decisions.

In the context of medical imaging, the usual types of bias seen in medicine can appear in forms that are specific to clinical imaging workflows (Table 1 and Table 2).[3–5] As in the broader practice of medicine, bias in medical imaging can lead to diagnostic errors as well as missed or inappropriate treatment opportunities.[6] For example, Black and Hispanic children are less likely to receive emergency imaging compared to their White counterparts.[7] Hospitalized compared to non-hospitalized patients are more likely to have their urgent imaging findings interpreted as non-diagnostic.[8] The drivers of bias in medical imaging can be categorized as factors leading to variations in three areas: (1) access to imaging tests, (2) acquisition of imaging data, and (3) assessment or interpretation of imaging findings (Table 2). In each category, AI and ML algorithms could potentially be used to mitigate the drivers. Some methods are currently refined for deployment while others have yet to be developed, tested for feasibility, or evaluated for effectiveness. On the other hand, we also recognize the risk of ML models propagating biases if engineers and clinicians are not careful in their development and application.

I.a. Machine Learning Can Reduce Bias in Access to Imaging

As in other areas of healthcare, disparities related to imaging first begin with access (Table 2, Access – Section I). Relevant clinician practice patterns, which influence whether or not an imaging test is ordered, are often based on an assessment of need, specialty training, years in practice, experience encountering a given type of patient presentation, and local practice patterns.[9,10] More subtly, various "-isms" (racism, sexism, ableism, ageism, etc.) can impact clinician practice patterns, resulting in disparities in access to imaging. Clinician practice patterns also depend on whether a given imaging diagnostic is available locally or regionally, within an payor network, covered in any part by any existing insurance, or technically feasible based on the availability of appropriate equipment, material availability, and personnel.[10] Given the technological, financial, and human capital overhead required, the availability of clinical imaging resources varies substantially across regions (Table 2, Access – Section II & III).[10] For instance, multiple studies have shown that evidence-based imaging (e.g. mammography) is often less accessible to economically disadvantaged patients – not only due to referral bias but also due to travel distance and clinic wait times at accessible centers. In particular, Black women tend to have to travel farther for MRIs, and 2018 data showed that Black and Hispanic patients were more likely to face lack of transportation to healthcare services when compared to their White counterparts.[11,12] Geographic disparities in the ability to physically access imaging facilities are further exacerbated by the uneven distribution of trained imaging personnel and specialists, with some parts of the United States having 4 times fewer radiologists per 100,000 Medicare beneficiaries than others (Figure 3, Section 1.1.1).[13] Not surprisingly, urban compared to rural communities are much more likely to have ready access to advanced imaging and skilled interpretation.

Recognizing the complexity as well as breadth of factors driving variable access to imaging diagnostics across the country, there are several areas where AI and ML algorithms may assist. To initially identify geographic inequities, AI analyses of map data and satellite images can determine disparate travel distances between neighborhoods and imaging centers.[14] For either urban or rural communities, existing intelligent transportation systems[15] can be adapted to integrate with health resource maps to identify minimal travel time and distance required for reach an imaging center. Given constrained options, data driven algorithms have previously guided resource allocation before and during the COVID pandemic as well as in other settings.[16] To address geographic distribution issues, similar algorithms can be developed to guide health systems and policy makers in selecting and incentivizing allocation of scarce material resources in ways that could maximize utility and benefit across a patient population. When considering shortages of personnel in certain geographic areas, AI and ML algorithms can be used to facilitate workflows that allow for some technical tasks to be automated or semi-automated; as further discussed the sections below, these same processes can also be used to mitigate issues related to image acquisition and image interpretation.

Beyond regional access barriers are the yet omnipresent barriers posed by cost (Table 2, Access – Section IV). As advanced cardiac imaging modalities such as CT and MRI have become more integrated into routine care, their availability through more broadly covering insurance plans has increased. However, certain patient subgroups (e.g. patients with cardiovascular disease) may yet opt out of obtaining medically indicated imaging while facing the costs of receiving an imaging test with or without insurance as well as the costs incurred by taking time off and arranging transportation to an imaging center.[17] Although still in development, there are now AI algorithms that may assist with reducing imaging related costs for patients. Multiple algorithms have been developed to guide clinical decisions around which symptomatic patients should receive stress testing or be screening for possible low EF,[18,19] and these can be used to screen out patients in whom testing is unnecessary. At some centers, AI algorithms are being developed to improve patient throughput, increase transportation efficiency, and facilitate visit planning.[20–23] Similarly, advanced algorithms can be developed to enable not only evidence-based but also locale-specific decision support tools that can guide clinicians on when and which diagnostic imaging test to consider for a given patient. For example, the type of cardiac stress test with imaging to order in a given situation may depend not only on patient factors but also on the availability as well as accessibility of a more cost-effective imaging modality.[10]

I.b. Machine Learning Can Reduce Bias in Acquisition of Imaging Data

In patients who receive a diagnostic study, the quality of imaging acquired varies based on multiple factors (Table 2). Quality of acquisition determines the quality of interpretation. Multiple technical factors can impact image acquisition including the type of equipment available and whether associated software is up-to-date and optimized for clinical use (Table 2, Acquisition – Section I). AI algorithms often are brittle and can behave unexpectedly on different equipment and formats of input data, making it critically important to train AI algorithms on diverse imaging datasets.[24] Although some of these factors are not readily improved without financial resources, democratization through open-source AI and ML algorithms may assist in cost-effective and efficient ways. For instance, calculating a coronary artery calcification (CAC) score conventionally requires acquiring images using a cardiac-gated image acquisition protocol from a dedicated scan. Now, several AI and ML advances have made it possible to derive a validated CAC score from archival chest CT images acquired without a specific cardiac protocol.[25,26]

Beyond pure technical variables that affect quality and output of image acquisition are patient factors (Table 2, Acquisition – Section II). Patient body habitus, positioning, or other physical variables can significantly affect image acquisition.[27,28] Almost all imaging modalities have technical parameters that are set to optimally capture images of anatomical tissue that tend to fall within a range of size, type, and distance from the image detector. Ultrasound equipment settings are configured to capture images of tissue located within a certain distance from the ultrasound probe and out-of-range tissue targets will be sub-optimally imaged. Similarly, CT cannot ensure good quality image acquisition of all anatomical structures for body sizes at the extremes. For all these reasons, imaging technicians receive dedicated training on how to adjust parameters before, during, or after image acquisition for a given patient. Many semi-automated tools already have been in use or in development to facilitate this tailoring process. These include gain and contrast adjustment tools for ultrasound and attenuation correction tools for CT and nuclear imaging. Given the shortage of trained MRI technicians, AI guided protocoling and acquisition of cardiac MRI imaging has been in rapid development.[29] Newer tools that rely more heavily on ML and AI algorithms are already in development for application across all the major imaging modalities and particularly those that traditionally have required substantial human technical inputs for operational function, thus reducing the risk of operator bias in difficult-to-image patients.[30–34] Medical acuity of the patient may also affect the extent to which image acquisition needs to be abbreviated and may cause sub-optimal patient positioning. Fortunately, acceleration of acquisition and accommodation for more variable clinical states using AI/ML may reduce barriers to diagnostic imaging for patients with habitus at the extremes, acute medical illness, or other characteristics that can impact image acquisition.

Population level variables such as disease prevalence and social determinants of health can lead to variable quality of image acquisition even when patients can access imaging (Table 2, Acquisition – Section III).[10] For instance, default imaging protocols may be set up to characterize disease entities that are more common for a given population and, thus, tend to miss the opportunity to characterize less common or rare diseases that are optimally imaged using specialized protocols. Even for imaging more common entities, incorrect study protocols will require patients to be recalled for repeat examinations in up to 20% of patients assessed in the outpatient setting, leading to wasted time, resources, and radiation exposure.[35] Improper protocoling also occurs in the inpatient setting, where inadequate imaging can delay time-sensitive clinical decisions. Notably, suboptimal protocoling and image acquisition appears more common among populations cared for in underserved areas compared to other regions (Figure 3 – Section 1.1.2).[10] To this end, AI algorithms can integrate clinical data with provider inputs to provide more patient-centered protocoling while also optimizing image acquisition procedures that may need to be implemented by less trained personnel. Incorporation of AI into existing clinical workflows, especially at understaffed or less specialized centers, can thus allow for more advanced imaging protocols to be performed at these sites while supported by remote readings as needed.

I.c. Machine Learning Can Reduce Bias in Interpretation of Imaging

Interpretation has been the primary focus on AI and ML algorithm development in medical imaging and is the area of greatest opportunity to reduce interpreter bias. In the absence of any clinical decision support systems, the process of interpretation risks bias arising from variables related to the interpreter, the patient, the imaging data, and the population prevalence of any existing disease (Table 2, Interpretation – Section I). The opportunities for AI and ML algorithms to mitigate biases fall into two overlapping areas: (1) automating and standardizing traditional imaging feature measurement; (2) identifying novel disease markers through computer vision approaches to feature analysis and pattern recognition.

I.c.i. Automating and Standardizing Measurement of Conventional Imaging Features—Human measurement of standard structural and functional features suffers from intra- and inter-observer variability, which is a well-recognized challenge to imaging core laboratories (Table 2, Interpretation – Section II).[36,37] Small measurement variability can translate into large deviations of derived or calculated variables, and reclassify patients from healthy to abnormal or vice versa. Human readers may adjust measurements in a manner influenced by patient's sociodemographic or clinical characteristics, unconsciously driving a greater or lesser likelihood of a particular cardiac diagnosis (Table 2, Interpretation – Section III). Thus, unbiased measurement tools that automate standard measures may be valuable not only for saving time, but mitigating this bias in a variety of contexts. For instance, rapidly advancing algorithms can now auto-calculate a number of structural features including ventricular wall thickness, chamber size, and ejection fraction (EF) on echocardiography and cardiac MRI, atherosclerotic plaque size and characteristics on CT, and myocardial perfusion as well as coronary flow reserve indices on SPECT and PET.[38]

Early generations of ML and AI algorithms for measuring structural and functional features had poor performance in less than optimal quality images; however, performance has dramatically improved over just the last few years, as algorithms such as the EchoNet Dynamic now function well on low quality images (Table 2, Interpretation – Section IV).[39,40] These facilitated assessments are especially relevant for closing diagnostic care gaps between men and women – who tend to have smaller caliber coronaries and relatively less cardiac mass – and whose heart disease is more likely to involve inflammation, microvascular disease and dysfunction, and heart failure with preserved EF (Figure 3 – Section 1.2.1.1 and Section 1.2.1.2).[41] Importantly, recent studies have shown not only that ML models reduce variability while also being concordant with human-made measurements, but that the algorithms can leverage information even from poor quality images and perform consistently in men and women.[39,40,42,43] In fact, recent randomized trial data by our group have demonstrated non-inferiority and even superiority of deep learning algorithms compared to conventional human-driven workflows for reproducibly measuring EF from clinical echocardiograms.[39] Beyond the left ventricle, ongoing work is focused on developing algorithms that might facilitate measures of other cardiac chambers and structures. These structures have greater variability in clinical practice, and thus while more technically challenging, have higher potential to mitigate human variability and bias. Notably, efforts are already under way to create a comprehensive AI driven workflow that can facilitate processing and reporting of all standardized measures of cardiac structure and function within a given imaging modality.[44] Similar work is ongoing for advanced measures such as myocardial strain on echocardiography and cardiac MRI, vascular caliber and flow on CT, and microvascular perfusion on PET.[45–48] Because conventional imaging measures and biomarkers may not fully capture disease heterogeneity across all patients, particularly those who have been under-represented, the automated processing of novel imaging phenotypes would allow for broadened investigations of disease phenotypes in understudied populations (Figure 3 – Sections 1.2.2.1 & 1.2.2.2).

I.c.ii. Facilitating Interpretation by Computer Vision Image Assessment— Computer vision assessment of medical images has effectively expanded and enhanced the offerings of radiomics. Prior to the advent of AI, radiomics was able to at least semi-automate approaches to quantifying image-based features such as pixel intensities, their spatial distributions, and relationships with one another – allowing for relatively reproducible measures of detailed and granular characteristics that could not be performed by a human. Radiomics techniques enabled a number of breakthrough discoveries in cardiovascular imaging including myocardial texture analysis and coronary plaque composition analysis.[49–52] However, practically employing radiomics techniques typically still required human effort to facilitate the measurements (e.g. specify the region of interest) and then manage the statistical analysis of the massive amounts of data generated. Now, AI approaches offers not only complete handling of the data but also complete automation the radiomic measures and, furthermore, directly analysis images to identify novel radiomic features – all without need for any human intervention.[53]

Extending from radiomics, AI computer vision techniques can now detect imaging features that are invisible to the human eye and yet specific to clinical disease phenotypes. This capability arises from the high density of imaging data (e.g. multiple complex images) within a single imaging study that computer vision can more efficiently and comprehensively analyze than humans. In cardiology, discoveries have included the ability to discern between amyloid and hypertrophic cardiomyopathy on echocardiography, predict coronary artery calcium score from echocardiography, and predict circulating biomarker levels of cardiac stress (e.g. BNP, troponin).[42,54–56] As in other fields, these models have demonstrated strong performance even after having learned from limited datasets and without involving segmentation or feature extraction. The potential benefit of deriving equal or greater quality prediction from less or limited information can be seen from studies demonstrating, for instance, adequate clinical prediction from a cardiac MRI study that could not be performed with contrast due to presence of chronic kidney disease.[57] Thus, AI algorithms are promising for addressing health disparities related to comorbidity profile or demographic characteristics or both. However, such algorithms are still considered preclinical in their development, and their ability to perform across the range of imaging quality and, moreover, to impact patient outcomes have yet to be evaluated. Nonetheless, there is now growing potential for novel imaging biomarkers to mitigate bias that is intrinsic or related to human interpretation inputs or provide new forms of decision support for clinicians to self-monitor for bias.

While extending the benefits of radiomics, which generates imaging biomarkers of disease, AI methods also offer benefits to multi-source data including genetic, epigenetic, transcriptomic, proteomic, metabolomic, and other types of molecular biomarker data. As recommended by experts in cardiovascular imaging as well as other fields,[58,59] the contributions of AI methods applied to multi-omics data are at least two-fold. First, AI methods add new computational capabilities for harnessing vast amounts of high-dimensional data and analyzing their complex inter-relations. Second, AI methods can effectively expand the overall dataset to be more inclusive such that missingness becomes more meaningful as well as manageable. For instance, some patients in whom advanced imaging data is missing due to limited healthcare access or other sources of bias can have this missingness counted and correlated in analyses of comprehensive multi-source data that allow for such complex relationships to be detected. In addition, more broadly available biomarker data across a diversity of patients can allow for potentially understudied groups to be analyzed through correlative analyses even in the setting of imaging data that are limited or absent at the individual level. In turn, the fusion of AI with multi-omics data analyses offers opportunities to leverage higher dimensionality of measures that can reflect broader variations in risk and treatment responses that not otherwise discernible by studies based on lower dimensional data, and particularly those without diverse population representation.

II.a. Challenges Amidst Opportunities: Risks of Propagating Existing Biases

Recognizing the many ways by which AI models can mitigate bias, we turn our attention to the manners by which models may inadvertently propagate bias (Figure 3 – Section 2). These include: (i) limited or unbalanced training sets (Figure 3 – Section 2.2), (ii) inadequate or incorrect labeling, (iii) incomplete accounting for correlative data, and (iv) inconsistent or selective use. Given that the many potential sources of bias that permeate the AI development and deployment life cycle have been detailed previously,[60] we will focus on key examples relevant to cardiovascular imaging.

First and foremost, training and testing sets that are inherently limited or unbalanced will introduce biases reflective of their limitations (Figure 2, Panel A). Given intrinsic differences in normal ranges for cardiac structural features (e.g. LV mass, aortic root diameter) between men and women, ML and AI models that are imbalanced by sex will be imbalanced in their ability to generate accurate diagnostic measurements and predictions for the under-trained group (Figure 2, Panel B). Similar effects will be seen with models trained on healthy or screening populations applied in disease populations or vice versa. Even imbalances between similar-appearing disease with different population rates will result in bias, such as a model trained to identify amyloid heart disease using data enriched with cardiac images from older patients may be biased to learning features of the TTR amyloid type whereas a model trained using mainly younger patient images may be biased to features of AL amyloid type. Sociodemographic and geographic disparities in data availability for AI model training remains a major concern. Many U.S. cohorts used to train deep learning models published from 2015 to 2019 do not uniformly represent patients from across the U.S.[24] It is now well recognized that many of the largest accessible open source datasets in the world are not representative of the global population and even may be missing demographic information altogether. For instance, the UK Biobank enables correlation between structure, function, and genetics but is comprised predominantly of healthy Caucasians aged 40–69 years living in the UK. Similarly, despite the fact that minority patients have been under-represented in cardiology trials, minority patients can be disproportionately affected by cardiac diseases. For example, non-Hispanic Black persons have greater absolute risks for developing heart failure when compared to other racial groups.[61] Thus, either over-representation and under-representation of any given patient group will pose challenges for any model attempting to derive generalizable results from a skewed dataset. Not surprisingly, ML models trained on data with limited diversity have demonstrated lower performance in minority racial, ethnic, and other underrepresented demographic groups.[62,63] Similarly, algorithms relying on binary categories of sex or algorithms trained on datasets where transgender and gender diverse (TGD) patients are under-represented can result in gender bias, even as sex-related bias is addressed. It is noteworthy that although AI applications in cardiology have led to high-impact discoveries, the majority of studies have failed to report race and ethnicity related demographic characteristics or presented limited data on model performance in these demographic subgroups (Table 3).[19,25,39,40,42,56,64–69]

Bias can also be incorporated into AI models based on inadequate or incorrect labeling. Under- or over-recognized clinical conditions within a given population may also propagate within models that depend on highly sensitive or specific labeling. As mentioned previously, small measurement variations can result in large inaccuracies in calculated variables (e.g., LVEF) with potentially significant clinical consequences wherein healthy patients are reclassified as unhealthy or vice versa. Accordingly, it is important during model development to have deliberate intentions around what should or should not be labeled to maintain overall balance in model training towards a specific diagnostic goal. Full awareness of algorithm decision factors should be explored using model explainability approaches to ensure that prediction is based on consistent, biologically plausible findings.

Bias can be inadvertently produced by ML and AI models if there is incomplete accounting for correlative data during model development. For instance, presence of an implanted medical device on cardiac images may result in specification gaming, where a model associates the presence in ways that are not biologically related to the given disease (e.g. prevalence of certain devices may be higher in more versus less resourced patient care environments). This concern is raised by reports of models being able to predict race or ethnicity from biomedical images.[70] Although these reports have considered potential confounders, it is difficult account for every factor that a ML approach may recognize, such as varying image quality or presence of medical devices associated with greater illness acuity or comorbidity that may, in turn, correlate with race or ethnicity. Accordingly, we and others have found that variable prevalence of comorbidities can serve as key confounders for models that appear to predict race or ethnicity from biomedical imaging.[71] These issues emphasize the importance of context in imaging based models predicting sociodemographic characteristics. Considerable caution in deploying these models should be taken to avoid discrimination, and this issue is likely to grow for models leveraging genetic data in any training set.[72]

Beyond development of an AI model, selective application and individual feelings towards AI can also introduce bias. This has been referred to as ‘automation bias’, the over-reliance for at least some users in assuming model result is almost always correct, and this effect may be exacerbated by user variability in either excessive or selective application of the model.[73] We anticipate that an engaged and expert human reader is still needed to overread images because even if the automated result may align with the correct diagnosis for well-defined clinical scenarios, any scenarios in which the algorithm is insufficiently trained is not likely to process optimally. This phenomenon is likely to remain true even for zero-shot learning systems, which offer the possibility of enabling diagnosis of previously unseen new disease conditions and yet encountered methodological challenges when faced with the COVID-19 chest x-ray upon the start of the pandemic.[74] Thus, even for models designed to solve relatively complex high-dimensional image analysis problems, over-reliance may augment or propagate subtle biases intrinsic to a given model.

II.b. Challenges Amidst Opportunities: Mitigating Present and Future Risks of Bias

Substantially biased AI models are easy to identify because they demonstrate obviously poor performance in practice; less overtly biased models are harder to identify as they may appear to perform and yet produce erroneous results that can propagate harm. Thus, an growing number of approaches have been proposed for mitigating potential bias of AI algorithms developed for healthcare applications including medical imaging (Figure 3). Similar to guidance that has been proposed by others,[75] we can recommend a framework for considering previously detailed approaches to minimizing bias when developing, implementing, and managing an AI algorithm for medical imaging (Table 4).[76–81]

In the earliest stages of model development, careful anticipation and planning is always required prior to any model training. An initial priority is to seek representative and balanced datasets for training and testing, including external datasets for later validation.[76] For prospectively enrolling studies, such as clinical trials, multiple AI algorithms applied to source cohort data can be used to facilitate patient selection and recruitment of underrepresented populations and this, in turn, leads to more optimal training datasets.[82] In addition to prioritizing population balancing, upfront inclusion of non-imaging factors can eventually improve model generalizability. Studies show that models trained using data on imaging features and incorporating data from demographic, clinical, genetic, and cognitive scores demonstrate less bias than those trained on imaging features alone.[83] Such models appear to generalize well across various demographic and clinical subgroups. It may well be that including covariates related to image acquisition, geographic location, type of clinical care setting, and other situational factors could also improve generalizability in parallel with efforts to ensure overall balance across datasets.[83] For this reason, diversity both in the representation of individuals and in the range of variable types in datasets is recommended.

The next priority is to ensure comprehensive and rigorous data handling as well as analyses.[77] Exploratory data analyses, conducted with domain expertise, can identify any data integrity issues while also ensuring no data leakage or improper feature removal, rescaling, imputation during initial data management.[77] If available datasets are found to be imbalanced in the distribution of key variables, rebalancing may be needed. This issue is relevant to cardiovascular imaging, for which there are still relatively few publicly accessible datasets.[76] Newer algorithms can leverage generative AI to produce more balanced synthetic training datasets that involve oversampling of minority classes.[84] Although computationally intensive, these approaches have been developed to accommodate large volume imaging data.[85]

During model training, a number of bias mitigation strategies can be employed in addition to data augmentation approaches (e.g. adversarial training).[60,78,86,87] Importantly, such approaches may yet be subject to algorithmic bias and should be used with care. After initial model training, a growing number of metrics are now available to assist with identifying and minimizing bias that may appear for certain subgroups (e.g. equal opportunity difference, disparate impact, FNR and FOR difference, general entropy index, statistical parity).[88–91] It has been noted that employing multiple such methods may or may not improve overall model performance.[60,86,92] Thus, additional strategies include integrating data from alternate sources such as laboratory, biomarker, multi-omics, and social determinants of health variables in fusion models. Indeed, multiple studies have shown that unsupervised learning approaches applied to a combination of clinical, genetic, and imaging data can identify novel distinct – and likely more inclusive – phenotypes of disease that can serve to improve predictions of both disease outcomes and their responses to interventions.[59]

As with any robust model development process, a comprehensive workflow will involve iterative training and re-training to optimize performance as well as mitigate bias.[93] Upon completion of a final model, initial and then routine reporting of model performance across datasets can be facilitated using tools such as ‘model cards’ for algorithms and ‘datasheets’ for datasets – intended to provide optimal transparency and updated guidance on how to employ, validate, and interpret results from a given model applied to a given dataset.[79–81,94,95] In addition to routinely applied and updated evaluations of overall performance and fairness metrics, these same tools are likely to add value for ongoing assessments of how an algorithm operates in real-world settings. Transparency of differential performance will also for interdisciplinary teams to iterate on and further optimize models across multiple use scenerios.[94,95]

Given that transparency is essential for bias mitigation, concerns have been raised regarding the black box nature of commercial datasets and algorithms. Ideally, even proprietary models will have accessible documentation regarding training and testing dataset features, model assumptions, and metrics of performance and fairness assessed across relevant subgroups. In the absence of transparent model and dataset reporting, practitioners are advised to consider any new algorithms with caution given that suboptimally developed models – even when developed by prominent companies – have demonstrated poor generalization. In some cases, because certain proprietary models may be too expensive for deployment at some centers, generalization may be limited to only resourced environments (Figure 3 – Section 2.1). In other cases, models have converged to use correlative relationships and resulted in biased results that marginalize minority groups.[96,97]

While undergoing rapid evolution, both commercial and non-commercial AI applications in medical imaging and medicine at large are widely recognized as warranting more attention from the FDA and similar oversight agencies worldwide. To date, AI technologies have gained approval without necessarily having robust external validation,[98] and the only high-level guidance is provided to promote fairness without any specific requirements to ensure mitigation of potential bias.[60] Notwithstanding plans to establish processes for monitoring of AI model performance in real-world settings,[95] there are growing calls for the FDA to establish a more structured process for initial evaluation and approval of AI technologies that are ideally aligned with how drugs and devices are regulated.[99,100] For this reason, we offer a framework for considering approaches to bias mitigation that are based on existing workflows for building new AI technologies while emphasizing transparency and also approximating the conventional four phases of drug development and review (Table 4). While undoubtedly more resource intensive, such a process is likely to assist with increasing the rigor and, in turn, the practical utility and generalizability of any AI algorithm intended to eventually serve as an application that improves human health.

III. Conclusion

Rapidly advancing ML and AI methods offer tremendous promise regarding the potential to mitigate biases that are intrinsic to workflows across all the 'three A' operational areas of medical imaging: access, acquisition, and assessment. Using AI and ML approaches to adapt high-resource imaging to be applicable in lower-resource settings may reduce barriers to imaging care in underserved populations. Automation of image measurements may reduce interpretation bias, and novel imaging biomarkers may help inform interpreting physicians of under-recognized conditions. However, both developers and end-users must be cognizant of how AI models can propagate biases just as easily as they can mitigate them. This requires attention and care in the design, training, deployment, and use of the models. Creating fair and equitable models is a goal that must be collaboratively addressed by clinicians and engineers. While AI and ML approaches may seem unbiased, model development will never be perfectly inclusive of all data types, data sources, and inputs. Thus, developers and clinicians should understand the limitations of AI methods, as they do for traditional diagnostic tools, so that decisions remain rooted in the clinical context and serve all of our patients optimally.

Funding

This work was supported in part by the Sarnoff Cardiovascular Research Fellowship (A.V.), the AHA 23CDA1053659 (A.C.K.), the KL2 sub-award from NIH UL1TR001881 (A.C.K), the Erika J Glazer Family Foundation (S.C.), and NIH grants R00-HL157421 (DO) and R01-HL131532 (DO, SC).

REFERENCES

1. Lee P, Le Saux M, Siegel R, Goyal M, Chen C, Ma Y and Meltzer AC. Racial and ethnic disparities in the management of acute pain in US emergency departments: Meta-analysis and systematic review. Am J Emerg Med. 2019;37:1770–1777. [PubMed: 31186154]
2. Clerc Liaudat C, Vaucher P, De Francesco T, Jaunin-Stalder N, Herzig L, Verdon F, Favrat B, Locatelli I and Clair C. Sex/gender bias in the management of chest pain in ambulatory care. Womens Health (Lond). 2018;14:1745506518805641. [PubMed: 30370833]
3. Olenski AR, Zimerman A, Coussens S and Jena AB. Behavioral Heuristics in Coronary-Artery Bypass Graft Surgery. N Engl J Med. 2020;382:778–779. [PubMed: 32074429]
4. Hammond MEH, Stehlik J, Drakos SG and Kfoury AG. Bias in medicine: lessons learned and mitigation strategies. JACC Basic Transl Sci. 2021;6:78–85. [PubMed: 33532668]
5. Gong J, Zhang Y, Feng J, Huang Y, Wei Y and Zhang W. Influence of framing on medical decision making. EXCLI journal. 2013;12:20. [PubMed: 27034630]
6. Busby LP, Courtier JL and Glastonbury CM. Bias in Radiology: The How and Why of Misses and Misinterpretations. Radiographics. 2018;38:236–247. [PubMed: 29194009]
7. Marin JR, Rodean J, Hall M, Alpern ER, Aronson PL, Chaudhari PP, Cohen E, Freedman SB, Morse RB, Peltz A, Samuels-Kalow M, Shah SS, Simon HK and Neuman MI. Racial and Ethnic Differences in Emergency Department Diagnostic Imaging at US Children's Hospitals, 2016–2019. JAMA Netw Open. 2021;4:e2033710. [PubMed: 33512517]
8. Ellenberger K, Jeyaprakash P, Sivapathan S, Sangha S, Kitley J, Darshni A, Chen D, Negishi K and Pathan F. The Effect of Obesity on Echocardiographic Image Quality. Heart Lung Circ. 2022;31:207–215. [PubMed: 34373191]
9. Hyland PM, Xu J, Shen C, Markson LJ, Manning WJ and Strom JB. Race, sex and age disparities in echocardiography among Medicare beneficiaries in an integrated healthcare system. Heart. 2022;108:956–963. [PubMed: 34615667]

10. Waite S, Scott J and Colombo D. Narrowing the Gap: Imaging Disparities in Radiology. Radiology. 2021;299:27–35. [PubMed: 33560191]
11. Onega T, Lee CI, Benkeser D, Alford-Teaster J, Haas JS, Tosteson AN, Hill D, Shi X, Henderson LM and Hubbard RA. Travel Burden to Breast MRI and Utilization: Are Risk and Sociodemographics Related? J Am Coll Radiol. 2016;13:611–9. [PubMed: 27026577]
12. Caraballo C, Ndumele CD, Roy B, Lu Y, Riley C, Herrin J and Krumholz HM. Trends in Racial and Ethnic Disparities in Barriers to Timely Medical Care Among Adults in the US, 1999 to 2018. JAMA Health Forum. 2022;3:e223856. [PubMed: 36306118]
13. Khurana A, Patel B and Sharpe R Jr. Geographic Variations in Growth of Radiologists and Medicare Enrollees From 2012 to 2019. J Am Coll Radiol. 2022;19:1006–1014. [PubMed: 35961410]
14. Bruzelius E, Le M, Kenny A, Downey J, Danieletto M, Baum A, Doupe P, Silva B, Landrigan PJ and Singh P. Satellite images and machine learning can identify remote communities to facilitate access to health services. J Am Med Inform Assoc. 2019;26:806–812. [PubMed: 31411691]
15. Bharadiya J. Artificial Intelligence in Transportation Systems A Critical Review. American Journal of Computing and Engineering. 2023;6:34–45.
16. Wu H, Lu X and Wang H. The Application of Artificial Intelligence in Health Care Resource Allocation Before and During the COVID-19 Pandemic: Scoping Review. JMIR AI. 2023;2:e38397.
17. Cavallo JJ, Zhang Y, Staib LH, Lampert R and Weinreb JC. Disparities in care among patients with cardiac implantable electronic devices undergoing MRI. J Am Coll Radiol. 2017;14:1566–1571. [PubMed: 28899705]
18. Lu J, Stewart J, Bennamoun M, Goudie A, Eshraghian J, Ihdayhid A, Sanfilippo F, Small GR, Chow BJ and Dwivedi G. Deep learning model to predict exercise stress test results: Optimizing the diagnostic test selection strategy and reduce wastage in suspected coronary artery disease patients. Comput Methods Programs Biomed. 2023;240:107717. [PubMed: 37454499]
19. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA and Munger TM. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nat Med. 2019;25:70–74. [PubMed: 30617318]
20. Abduljabbar R, Dia H, Liyanage S and Bagloee SA. Applications of artificial intelligence in transport: An overview. Sustainability. 2019;11:189.
21. van Leeuwen KG, Meijer FJ, Schalekamp S, Rutten MJ, van Dijk EJ, van Ginneken B, Govers TM and de Rooij M. Cost-effectiveness of artificial intelligence aided vessel occlusion detection in acute stroke: an early health technology assessment. Insights Imaging. 2021;12:1–9. [PubMed: 33398669]
22. Tejedor M, Herrero A, Castresana C, Mesón R, Taracido JC, Sánchez M and Delgado M. Can artificial intelligence increase the efficiency of referrals from primary to specialized care? Rev Esp Enferm Dig. 2023;115:141–142. [PubMed: 35770554]
23. Dawoodbhoy FM, Delaney J, Cecula P, Yu J, Peacock I, Tan J and Cox B. AI in patient flow: applications of artificial intelligence to improve patient flow in NHS acute mental health inpatient units. Heliyon. 2021;7:e06993. [PubMed: 34036191]
24. Kaushal A, Altman R and Langlotz C. Geographic Distribution of US Cohorts Used to Train Deep Learning Algorithms. JAMA. 2020;324:1212–1213. [PubMed: 32960230]
25. Eng D, Chute C, Khandwala N, Rajpurkar P, Long J, Shleifer S, Khalaf MH, Sandhu AT, Rodriguez F, Maron DJ, Seyyedi S, Marin D, Golub I, Budoff M, Kitamura F, Takahashi MS, Filice RW, Shah R, Mongan J, Kallianos K, Langlotz CP, Lungren MP, Ng AY and Patel BN. Automated coronary calcium scoring using deep learning with multicenter external validation. NPJ Digit Med. 2021;4:88. [PubMed: 34075194]
26. Peng AW, Dudum R, Jain SS, Maron DJ, Patel BN, Khandwala N, Eng D, Chaudhari AS, Sandhu AT and Rodriguez F. Association of Coronary Artery Calcium Detected by Routine Ungated CT Imaging With Cardiovascular Outcomes. J Am Coll Cardiol. 2023;82:1192–1202. [PubMed: 37704309]

27. Ellenberger K, Jeyaprakash P, Sivapathan S, Sangha S, Kitley J, Darshni A, Chen D, Negishi K and Pathan F. The effect of obesity on echocardiographic image quality. Heart Lung Circ. 2022;31:207–215. [PubMed: 34373191]
28. Uppot RN, Sahani DV, Hahn PF, Gervais D and Mueller PR. Impact of obesity on medical imaging and image-guided intervention. AJR Am J Roentgenol. 2007;188:433–440. [PubMed: 17242253]
29. Edalati M, Zheng Y, Watkins MP, Chen J, Liu L, Zhang S, Song Y, Soleymani S, Lenihan DJ and Lanza GM. Implementation and prospective clinical validation of AI-based planning and shimming techniques in cardiac MRI. Med Phys. 2022;49:129–143. [PubMed: 34748660]
30. Schaefferkoetter J, Shah V, Hayden C, Prior JO and Zuehlsdorff S. Correction to: Deep learning for improving PET/CT attenuation correction by elastic registration of anatomical data. Eur J Nucl Med Mol Imaging. 2023;50:2570. [PubMed: 36964404]
31. Aldaas OM, Igata S, Raisinghani A, Kraushaar M and DeMaria AN. Accuracy of left ventricular ejection fraction determined by automated analysis of handheld echocardiograms: A comparison of experienced and novice examiners. Echocardiography. 2019;36:2145–2151. [PubMed: 31786824]
32. Narang A, Bae R, Hong H, Thomas Y, Surette S, Cadieu C, Chaudhry A, Martin RP, McCarthy PM, Rubenson DS, Goldstein S, Little SH, Lang RM, Weissman NJ and Thomas JD. Utility of a Deep-Learning Algorithm to Guide Novices to Acquire Echocardiograms for Limited Diagnostic Use. JAMA Cardiol. 2021;6:624–632. [PubMed: 33599681]
33. McMillan AB and Bradshaw TJ. Artificial Intelligence-Based Data Corrections for Attenuation and Scatter in Position Emission Tomography and Single-Photon Emission Computed Tomography. PET Clin. 2021;16:543–552. [PubMed: 34364816]
34. Leiner T, Rueckert D, Suinesiaputra A, Baessler B, Nezafat R, Isgum I and Young AA. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. J Cardiovasc Magn Reson. 2019;21:61. [PubMed: 31590664]
35. Gyftopoulos S, Kim D, Aaltonen E and Horwitz LI. Patient recall imaging in the ambulatory setting. AJR Am J Roentgenol. 2016;206:787–791. [PubMed: 26866338]
36. Khouri MG, Ky B, Dunn G, Plappert T, Englefield V, Rabineau D, Yow E, Barnhart HX, St John Sutton M and Douglas PS. Echocardiography Core Laboratory Reproducibility of Cardiac Safety Assessments in Cardio-Oncology. J Am Soc Echocardiogr. 2018;31:361–371 e3. [PubMed: 29395626]
37. Douglas PS, DeCara JM, Devereux RB, Duckworth S, Gardin JM, Jaber WA, Morehead AJ, Oh JK, Picard MH, Solomon SD, Wei K, Weissman NJ, American Society of Echocardiography S and American College of Cardiology F. Echocardiographic imaging in clinical trials: American Society of Echocardiography Standards for echocardiography core laboratories: endorsed by the American College of Cardiology Foundation. J Am Soc Echocardiogr. 2009;22:755–65. [PubMed: 19560654]
38. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP and Marwick TH. Artificial Intelligence in Cardiovascular Imaging: JACC State-of-the-Art Review. J Am Coll Cardiol. 2019;73:1317–1335. [PubMed: 30898208]
39. He B, Kwan AC, Cho JH, Yuan N, Pollick C, Shiota T, Ebinger J, Bello NA, Wei J, Josan K, Duffy G, Jujjavarapu M, Siegel R, Cheng S, Zou JY and Ouyang D. Blinded, randomized trial of sonographer versus AI cardiac function assessment. Nature. 2023;616:520–524. [PubMed: 37020027]
40. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA and Zou JY. Video-based AI for beat-to-beat assessment of cardiac function. Nature. 2020;580:252–256. [PubMed: 32269341]
41. Mikail N, Rossi A, Bengs S, Haider A, Stähli BE, Portmann A, Imperiale A, Treyer V, Meisel A, Pazhenkottil AP, Messerli M, Regitz-Zagrosek V, Kaufmann PA, Buechel RR and Gebhard C. Imaging of heart disease in women: review and case presentation. Eur J Nucl Med Mol Imaging. 2022;50:130–159. [PubMed: 35974185]
42. Duffy G, Cheng PP, Yuan N, He B, Kwan AC, Shun-Shin MJ, Alexander KM, Ebinger J, Lungren MP, Rader F, Liang DH, Schnittger I, Ashley EA, Zou JY, Patel J, Witteles R, Cheng S and Ouyang D. High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy With Cardiovascular Deep Learning. JAMA Cardiol. 2022;7:386–395. [PubMed: 35195663]

43. Fernandes J, Loncaric F, Marciniak M, Gilbert A, Smistad E, Lovstakken L, Mcleod K, Sitges M and Lamata P. Automatic measurement of LV wall thickness from 2D cardiac echocardiography. Eur Heart J-Cardiovasc Imaging. 2022;23:jeab289. 150.
44. Tromp J, Bauer D, Claggett BL, Frost M, Iversen MB, Prasad N, Petrie MC, Larson MG, Ezekowitz JA and Solomon SD. A formal validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. Nat Commun. 2022;13:6776. [PubMed: 36351912]
45. Lin A, Pieszko K, Park C, Ignor K, Williams MC, Slomka P and Dey D. Artificial intelligence in cardiovascular imaging: enhancing image analysis and risk stratification. BJR Open. 2023;5:20220021. [PubMed: 37396483]
46. Kwan AC, Tokodi M, Jain I, Chang E, Theurer J, Tang X, Francisco N, Haddad F, Liang D, Yuan N, Merkely B, Siegel R, Cheng S, Kovács A and Ouyang D. Deep Learning-Derived Myocardial Strain. medRxiv. Preprint posted online March 18, 2022. 10.1101/2022.03.16.22272374
47. Konwerski M, Gąsecka A, Opolski G, Grabowski M and Mazurek T. Role of epicardial adipose tissue in cardiovascular diseases: a review. Biology. 2022;11:355. [PubMed: 35336728]
48. West HW, Siddique M, Williams MC, Volpe L, Desai R, Lyasheva M, Thomas S, Dangas K, Kotanidis CP, Tomlins P, Mahon C, Kardos A, Adlam D, Graby J, Rodrigues JCL, Shirodaria C, Deanfield J, Mehta NN, Neubauer S, Channon KM, Desai MY, Nicol ED, Newby DE, Antoniades C and Investigators O. Deep-Learning for Epicardial Adipose Tissue Assessment With Computed Tomography: Implications for Cardiovascular Risk Prediction. JACC Cardiovasc Imaging. 2023;16:800–816. [PubMed: 36881425]
49. Lo Iacono F, Maragna R, Pontone G and Corino VD. A robust radiomic-based machine learning approach to detect cardiac amyloidosis using cardiac computed tomography. Front Radiol. 2023;3:1193046. [PubMed: 37588665]
50. Lin A, Kolossváry M, Cadet S, McElhinney P, Goeller M, Han D, Yuvaraj J, Nerlekar N, Slomka PJ and Marwan M. Radiomics-based precision phenotyping identifies unstable coronary plaques from computed tomography angiography. JACC Cardiovasc Imaging. 2022;15:859–871. [PubMed: 35512957]
51. Kwan A, Demosthenes E, Salto G, Ouyang D, Nguyen T, Nwabuo CC, Luong E, Hoang A, Osypiuk E, Stantchev P, Kim EH, Hiremath P, Li D, Vasan R, Xanthakis V and Cheng S. Cardiac microstructural alterations measured by echocardiography identify sex-specific risk for heart failure. Heart. 2022;108:1800–1806. [PubMed: 35680379]
52. Xu P, Xue Y, Schoepf UJ, Varga-Szemes A, Griffith J, Yacoub B, Zhou F, Zhou C, Yang Y, Xing W and Zhang L. Radiomics. Circ Cardiovasc Imaging. 2021;14:e011747. [PubMed: 33722057]
53. Koçak B, Durmaz E, Ateş E and Kiliçkesmez Ö. Radiomics with artificial intelligence: a practical guide for beginners. Diagn Interv Radiol. 2019;25:485–495. [PubMed: 31650960]
54. Hughes JW, Yuan N, He B, Ouyang J, Ebinger J, Botting P, Lee J, Theurer J, Tooley JE, Nieman K, Lungren MP, Liang DH, Schnittger I, Chen JH, Ashley EA, Cheng S, Ouyang D and Zou JY. Deep learning evaluation of biomarkers from echocardiogram videos. EBioMedicine. 2021;73:103613. [PubMed: 34656880]
55. Yuan N, Kwan AC, Duffy G, Theurer J, Chen JH, Nieman K, Botting P, Dey D, Berman DS, Cheng S and Ouyang D. Prediction of Coronary Artery Calcium Using Deep Learning of Echocardiograms. J Am Soc Echocardiogr. 2023;36:474–481 e3. [PubMed: 36566995]
56. Ghorbani A, Ouyang D, Abid A, He B, Chen JH, Harrington RA, Liang DH, Ashley EA and Zou JY. Deep learning interpretation of echocardiograms. NPJ Digit Med. 2020;3:10. [PubMed: 31993508]
57. Avard E, Shiri I, Hajianfar G, Abdollahi H, Kalantari KR, Houshmand G, Kasani K, Bitarafan-Rajabi A, Deevband MR and Oveisi M. Non-contrast Cine Cardiac Magnetic Resonance image radiomics features and machine learning algorithms for myocardial infarction detection. Comput Biol Med. 2022;141:105145. [PubMed: 34929466]
58. Nayor M, Shen L, Hunninghake GM, Kochunov P, Barr RG, Bluemke DA, Broeckel U, Caravan P, Cheng S, de Vries PS, Hoffmann U, Kolossvary M, Li H, Luo J, McNally EM, Thanassoulis G, Arnett DK and Vasan RS. Progress and Research Priorities in Imaging Genomics for Heart and Lung Disease: Summary of an NHLBI Workshop. Circ Cardiovasc Imaging. 2021;14:e012943. [PubMed: 34387095]

59. Wei L, Niraula D, Gates EDH, Fu J, Luo Y, Nyflot MJ, Bowen SR, El Naqa IM and Cui S. Artificial intelligence (AI) and machine learning (ML) in precision oncology: a review on enhancing discoverability through multiomics integration. Br J Radiol. 2023;96:20230211. [PubMed: 37660402]

60. Gichoya JW, Thomas K, Celi LA, Safdar N, Banerjee I, Banja JD, Seyyed-Kalantari L, Trivedi H and Purkayastha S. AI pitfalls and what not to do: mitigating bias in AI. Br J Radiol. 2023;96:20230023. [PubMed: 37698583]

61. Piña IL, Jimenez S, Lewis EF, Morris AA, Onwuanyi A, Tam E and Ventura HO. Race and ethnicity in heart failure: JACC focus seminar 8/9. J Am Coll Cardiol. 2021;78:2589–2598. [PubMed: 34887145]

62. Puyol-Anton E, Ruijsink B, Mariscal Harana J, Piechnik SK, Neubauer S, Petersen SE, Razavi R, Chowienczyk P and King AP. Fairness in Cardiac Magnetic Resonance Imaging: Assessing Sex and Racial Bias in Deep Learning-Based Segmentation. Front Cardiovasc Med. 2022;9:859310. [PubMed: 35463778]

63. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY and Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med. 2021;27:2176–2182. [PubMed: 34893776]

64. Lin A, Manral N, McElhinney P, Killekar A, Matsumoto H, Kwiecinski J, Pieszko K, Razipour A, Grodecki K, Park C, Otaki Y, Doris M, Kwan AC, Han D, Kuronuma K, Flores Tomasino G, Tzolos E, Shanbhag A, Goeller M, Marwan M, Gransar H, Tamarappoo BK, Cadet S, Achenbach S, Nicholls SJ, Wong DT, Berman DS, Dweck M, Newby DE, Williams MC, Slomka PJ and Dey D. Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study. Lancet Digit Health. 2022;4:e256–e265. [PubMed: 35337643]

65. Holste G, Oikonomou EK, Mortazavi BJ, Coppi A, Faridi KF, Miller EJ, Forrest JK, McNamara RL, Ohno-Machado L, Yuan N, Gupta A, Ouyang D, Krumholz HM, Wang Z and Khera R. Severe aortic stenosis detection by deep learning applied to echocardiography. Eur Heart J. 2023;44:4592–4604. [PubMed: 37611002]

66. Avram R, Olgin JE, Ahmed Z, Verreault-Julien L, Wan A, Barrios J, Abreau S, Wan D, Gonzalez JE, Tardif J-C, So DY, Soni K and Tison GH. CathAI: fully automated coronary angiography interpretation and stenosis estimation. NPJ Digit Med. 2023;6:142. [PubMed: 37568050]

67. Ueda D, Matsumoto T, Ehara S, Yamamoto A, Walston SL, Ito A, Shimono T, Shiba M, Takeshita T and Fukuda D. Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study. Lancet Digit Health. 2023;5:e525–e533. [PubMed: 37422342]

68. Goto S, Mahara K, Beussink-Nelson L, Ikura H, Katsumata Y, Endo J, Gaggin HK, Shah SJ, Itabashi Y and MacRae CA. Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms. Nat Commun. 2021;12:2726. [PubMed: 33976142]

69. Avram R, Barrios JP, Abreau S, Goh CY, Ahmed Z, Chung K, So DY, Olgin JE and Tison GH. Automated Assessment of Cardiac Systolic Function From Coronary Angiograms With Video-Based Artificial Intelligence Algorithms. JAMA Cardiol. 2023;8:586–594. [PubMed: 37163297]

70. Gichoya JW, Banerjee I, Bhimireddy AR, Burns JL, Celi LA, Chen LC, Correa R, Dullerud N, Ghassemi M, Huang SC, Kuo PC, Lungren MP, Palmer LJ, Price BJ, Purkayastha S, Pyrros AT, Oakden-Rayner L, Okechukwu C, Seyyed-Kalantari L, Trivedi H, Wang R, Zaiman Z and Zhang H. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health. 2022;4:e406–e414. [PubMed: 35568690]

71. Duffy G, Clarke SL, Christensen M, He B, Yuan N, Cheng S and Ouyang D. Confounders mediate AI prediction of demographics in medical imaging. NPJ Digit Med. 2022;5:188. [PubMed: 36550271]

72. Zou J, Gichoya JW, Ho DE and Obermeyer Z. Implications of predicting race variables from medical images. Science. 2023;381:149–150. [PubMed: 37440627]

73. Dratsch T, Chen X, Rezazade Mehrizi M, Kloeckner R, Mahringer-Kunz A, Pusken M, Baessler B, Sauer S, Maintz D and Pinto Dos Santos D. Automation Bias in Mammography: The

Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance. Radiology. 2023;307:e222176. [PubMed: 37129490]

74. Rezaei M and Shahidi M. Zero-shot learning and its applications from autonomous vehicles to COVID-19 diagnosis: A review. Intell Based Med. 2020;3:100005. [PubMed: 33043311]
75. Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, Hicklen RS, Moukheiber L, Moukheiber D, Ma H and Mathur P. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health. 2023;2:e0000278. [PubMed: 37347721]
76. Ricci Lara MA, Echeveste R and Ferrante E. Addressing fairness in artificial intelligence for medical imaging. Nat Commun. 2022;13:4581. [PubMed: 35933408]
77. Rouzrokh P, Khosravi B, Faghani S, Moassefi M, Vera Garcia DV, Singh Y, Zhang K, Conte GM and Erickson BJ. Mitigating Bias in Radiology Machine Learning: 1. Data Handling. Radiol Artif Intell. 2022;4:e210290. [PubMed: 36204544]
78. Zhang K, Khosravi B, Vahdati S, Faghani S, Nugen F, Rassoulinejad-Mousavi SM, Moassefi M, Jagtap JMM, Singh Y, Rouzrokh P and Erickson BJ. Mitigating Bias in Radiology Machine Learning: 2. Model Development. Radiol Artif Intell. 2022;4:e220010. [PubMed: 36204532]
79. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, Spitzer E, Raji ID and Gebru T. Model Cards for Model Reporting. Paper presented at: Proceedings of the Conference on Fairness, Accountability, and Transparency; 2019; Atlanta, GA, USA.
80. Gebru T Datasheets for datasets. Communications of the ACM. 64:86–92.
81. Kahn CE Jr. Hitting the Mark: Reducing Bias in AI Systems. Radiol Artif Intell. 2022;4:e220171. [PubMed: 36204534]
82. Askin S, Burkhalter D, Calado G and El Dakrouni S. Artificial Intelligence Applied to clinical trials: opportunities and challenges. Health Technol (Berl). 2023;13:203–213. [PubMed: 36923325]
83. Wang R, Chaudhari P and Davatzikos C. Bias in machine learning models can be significantly mitigated by careful training: Evidence from neuroimaging studies. Proc Natl Acad Sci U S A. 2023;120:e2211613120. [PubMed: 36716365]
84. Elreedy D and Atiya AF. A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. Information Sciences. 2019;505:32–64.
85. Jeong JJ, Tariq A, Adejumo T, Trivedi H, Gichoya JW and Banerjee I. Systematic Review of Generative Adversarial Networks (GANs) for Medical Image Classification and Segmentation. J Digit Imaging. 2022;35:137–152. [PubMed: 35022924]
86. Ricci Lara MA, Echeveste R and Ferrante E. Addressing fairness in artificial intelligence for medical imaging. Nat Commun. 2022;13:4581. [PubMed: 35933408]
87. Paulus JK and Kent DM. Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. NPJ Digit Med. 2020;3:99. [PubMed: 32821854]
88. Carey AN and Wu X. The statistical fairness field guide: perspectives from social and formal sciences. AI and Ethics. 2023;3:1–23.
89. Fletcher RR, Nakeshimana A and Olubeko O. Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. Front Artif Intell. 2020;3:561802. [PubMed: 33981989]
90. Speicher T, Heidari H, Grgic-Hlaca N, Gummadi KP, Singla A, Weller A and Zafar MB. A unified approach to quantifying algorithmic unfairness: Measuring individual &group unfairness via inequality indices. Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 2018:2239–2248.
91. Hardt M, Price E and Srebro N. Equality of opportunity in supervised learning. Advances in neural information processing systems. 2016;29.
92. Li F, Wu P, Ong HH, Peterson JF, Wei WQ and Zhao J. Evaluating and mitigating bias in machine learning models for cardiovascular disease prediction. J Biomed Inform. 2023;138:104294. [PubMed: 36706849]
93. Vokinger KN, Feuerriegel S and Kesselheim AS. Mitigating bias in machine learning for medicine. Communications Medicine. 2021;1:25. [PubMed: 34522916]

94. Ganapathi S, Palmer J, Alderman JE, Calvert M, Espinoza C, Gath J, Ghassemi M, Heller K, Mckay F and Karthikesalingam A. Tackling bias in AI health datasets through the STANDING Together initiative. Nat med. 2022;28:2232–2233. [PubMed: 36163296]

95. Mittermaier M, Raza MM and Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. NPJ Digit Med. 2023;6:113. [PubMed: 37311802]

96. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, Pestrue J, Phillips M, Konye J, Penoza C, Ghous M and Singh K. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med Medicine. 2021;181:1065–1070.

97. Obermeyer Z, Powers B, Vogeli C and Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366:447–453. [PubMed: 31649194]

98. Yu AC, Mohajer B and Eng J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. Radiology: Artificial Intelligence. 2022;4:e210064. [PubMed: 35652114]

99. Van Norman GA. Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs. JACC Basic Transl Sci. 2016;1:170–179. [PubMed: 30167510]

100. Van Norman GA. Drugs, Devices, and the FDA: Part 2: An Overview of Approval Processes: FDA Approval of Medical Devices. JACC Basic Transl Sci. 2016;1:277–287. [PubMed: 30167516]

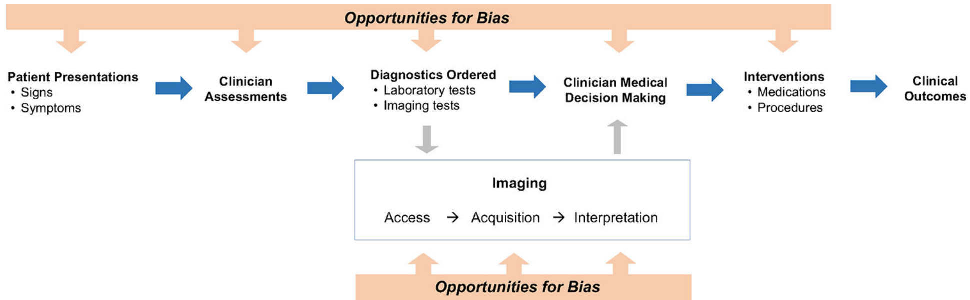

Figure 1. Opportunities for Bias in Clinical Workflows.

In addition to the multiple opportunities for bias that are recognized to exist throughout the clinical workflow, from initial patient presentation to decision making and intervention, there are additional and often overlooked opportunities for bias to be introduced as well as mitigated in the process of obtaining medical imaging.

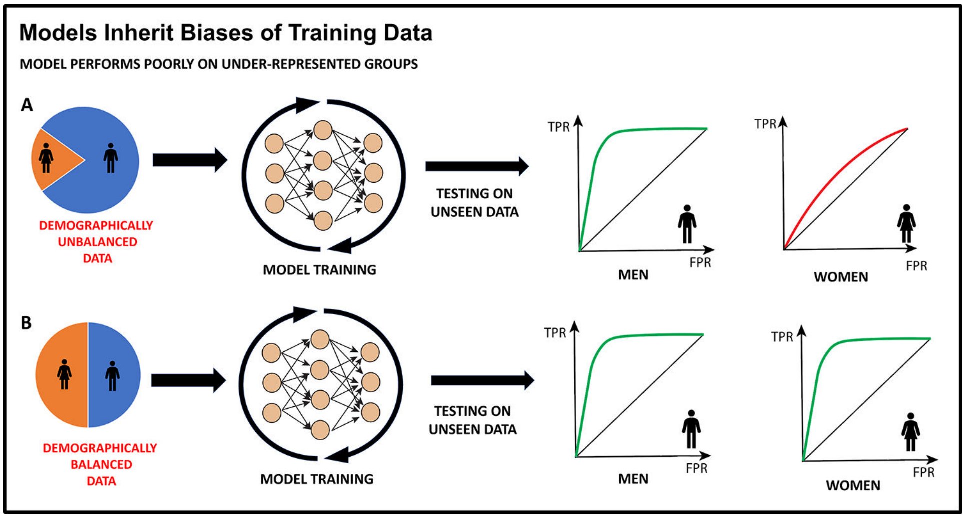

Figure 2. Imbalanced training data can lead to biased models.

Although it is unclear whether any under-representation in training data or under-representation relative to the population proportion of a demographic group is a driver, imbalanced training datasets (as shown in Model A) have led to models that perform worse for select population subgroups compared to models trained where demographic groups are better balanced (Model B). For this reason, investigating bias in model predictions and imbalance in training data is key to creating equitable models.

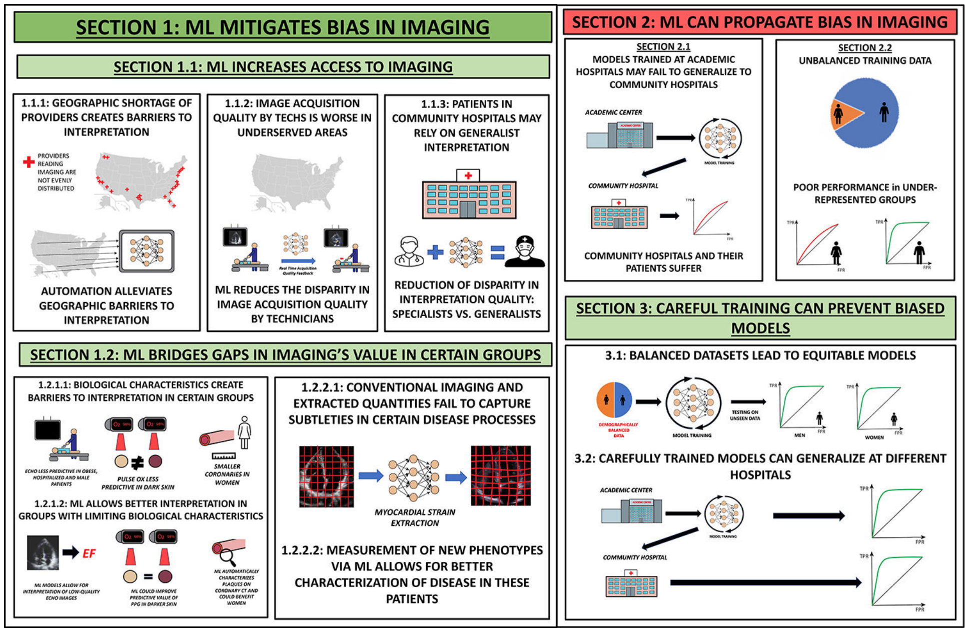

Figure 3. Challenges and Opportunities Inherent to Machine Learning Applications in Imaging.
Machine learning can both mitigate (Section 1) and propagate (Section 2) bias in medical imaging. With appropriate care, providers can create models that are not only clinically useful but that can also be equitable (Section 3).

Table 1.

Types of bias in medical imaging.

Type of Bias	Definition	Areas of Imaging Affected: *Examples*
Anchoring Bias[4]	Tendency to rely on information heard early or first on a topic to decide	**Access:** Cardiologist assumes coronary artery disease not causing chest pain due to patient's history of gastroesophageal reflux disease discussed early in encounter. **Interpretation:** Cardiologist's ejection fraction estimation biased by early comments in fellow's interpretation.
Availability Bias[4]	Tendency to over-weight or over-rely on information readily coming to mind	**Access:** Following a lecture on acute coronary syndrome, an Emergency Department physician fails to consider pulmonary embolism as a cause of chest pain and does not recommend chest vessel imaging. **Interpretation:** Radiology resident concludes pneumothorax is likely primary spontaneous pneumothorax after lecture on spontaneous pneumothorax in young men.
Bias Blind Spot[4]	Tendency to under-estimate one's own biases	**Access:** Primary care physician fails to recognize biases and apply knowledge gained from reading an article on cognitive bias on imaging referral practices. **Interpretation:** Radiologist fails to recognize how cognitive bias influences own errors despite reading article on cognitive bias impacting image interpretation.
Confirmation Bias[4]	Tendency to over-weight or over-value information that only supports a previously held idea	**Access:** ED physician over-weights a patient's history of venous thromboembolism to support possible PE diagnosis despite patient having signs and symptoms concerning for acute coronary syndrome. **Interpretation:** Physician with low suspicion for pulmonary embolism rules it out after patient with pleuritic chest pain displays no McConnell's sign on echo despite low sensitivity.
Framing Effect[5]	Tendency to base decisions and judgement on presentation of information rather than actual information itself	**Access:** Physician taught that 30% of women with chest pain have obstructive coronary artery disease recommends echo more often compared to another physician as taught that 70% of women with chest pain do not have obstructive coronary artery disease. **Interpretation:** Radiologist misses crucial finding on computed tomography imaging due to framing with a less concerning clinical history.
Gender Bias[3]	Tendency to based treatment or diagnosis decisions on gender-based misconceptions	**Access:** Women with chest pain less likely to receive cardiac workup (echo, etc.) for chest pain. **Interpretation:** Cardiologist unconsciously underestimates ejection fraction in sick men given higher rates of cardiomyopathy in men.
Hindsight Bias[3]	Tendency to over-estimate how much an event could have been foreseen or prevented	**Access:** Physician more likely to recommend cardiac workup (echo, etc.) for men presenting with chest pain than women with chest pain. **Interpretation:** Cardiologist unconsciously underestimates ejection fraction in sick men, given higher rates of cardiomyopathy in men.
Premature Closure[3]	Tendency to decide based on limited information without exploring a full range of possibilities	**Access:** Cardiologist suspects aortic stenosis based on patient's family history and does not explore possibility of concomitant coronary artery disease, failing to recommend stress testing or angiography. **Acquisition:** Provider recommends standard echo without considering more advanced cardiac imaging methods for patient with possible microvascular coronary disease. **Interpretation:** Cardiologist sees aortic regurgitation on echocardiogram of hemodynamically unstable patient and concludes the assessment, failing to consider possible aortic dissection.
Status Quo Bias[4]	Tendency to prefer current practices or guidelines while ignoring newer, more effective ones due to a resistance to change	**Access:** Primary care physician uses outdated perioperative assessment guidelines to recommend cardiac diagnostic workup needed for a pre-operative patient. **Acquisition:** Radiologist and technicians favor using a familiar old imaging protocol as opposed to a newer evidence-based protocol.

Table 2.

Potential sources of bias in medical imaging, examples, and potential solutions.

	Sources of Bias	Examples	Potential Solutions
Access	I. Clinician practice patterns	• Women with chest pain are less likely to be referred for cardiac stress testing compared to men with chest pain • Transthoracic echo is ordered less frequently for older, non-White women	• Evidence-based algorithms can guide clinical decision making on cardiac stress test referral for both sexes • AI algorithms that can screen electrocardiograms for low ejection fraction and other cardiac abnormalities that would warrant an echo evaluation
	II. Geography (distance)	• Black women tend to have to travel farther for magnetic resonance imaging • Transportation barriers a more frequent for Black and Hispanic patients than their counterparts	• AI analyses of satellite images can identify disparate distances between neighborhoods and imaging centers • Adapting intelligent transportation systems can be repurposed to close gaps in ability to access imaging centers
	III. Resources (equipment, supplies, personnel)	• Uneven distribution of radiologists, imaging specialists • Uneven distribution of specialized imaging machines across the country	• Automated AI interpretation allows does not rely on expert operators or over-readers • AI/ML can guide healthcare resource allocation
	IV. Cost (insurance, financing)	• Limited or lack of insurance prevent patients from obtaining advanced cardiovascular imaging tests	• Algorithms can reassess evidence-based need for tests • AI algorithms may reduce local or test specific costs
Acquisition	I. Technical variables (equipment, software, personnel)	• Some advanced imaging protocols require technology or software not available at every center • Patients may be incorrectly protocoled, especially in underserved areas	• Modified or AI algorithms can facilitate advanced measures on images acquired using standard methods (e.g. coronary artery calcium) • Automated AI-based protocoling • AI guided image acquisition for minimally trained (e.g. echo)
	II. Patient variables (body or morbidity related variables)	• Suboptimal image quality in patients with obesity or breast or device related attenuation or artefacts • Suboptimal positioning for imaging due to illness	• AI guided acquisition designed to account for presence of obesity, breast, or devices • AI guided patient positioning
	III. Population variables (population prevalence of a disease)	• Protocols may default to optimally characterize diseases more common in a given population • Suboptimal protocoling and image acquisition may be more common in less resourced environments	• AI algorithms can integrate available data to facilitate patient-centered protocoling • AI guided acquisition for minimally trained personnel can allow for more optimal imaging
Interpretation	I. Population variables (population prevalence of a disease)	• Over- or under-representation of uncommon diseases in a given population can cause readers to either over or under-diagnose these conditions	• Computer vision algorithms have been shown to improve accuracy for identifying relatively uncommon diseases such as amyloidosis and hypertrophic cardiomyopathy
	II. Interpreter variables (expertise, experienced)	• Inter- and intra-observer variability is widely prevalent across for human measures of structural and functional features (EF, etc.)	• Automated AI measurement reduces variability for many standard cardiovascular measures (wall thickness, cavity dimensions, ejection fraction, plaque size and composition)
	III. Patient variables (demographics, comorbidities)	• Sociodemographic characteristics can influence human assessments of acquired images • Actual differences based on biological differences (e.g. smaller-caliber coronaries in women) may be overlooked	• Radiomics and automated AI measurements can minimize or remove potential variation due to human bias, and incorporate demographic and clinical information to account for expected biological differences • Fusion AI/multi-omics can permit more personalized interpretation and management plans
	IV. Image variables (image quality, accessibility)	• Lower quality cardiovascular images often lead to non-diagnostic or overtly limited image interpretation	• Deep learning algorithms trained on lower quality images can demonstrate increasingly consistent performance such images when tested in real-world scenarios

Table 3.

Recent published large studies of AI in cardiovascular imaging and any reporting of performance in demographic subsets.

Study Title	Journal	Demographics	Demographic or Clinical Subset Analysis of Performance
Video-based AI for beat-to-beat assessment of cardiac function[40]	*Nature*	Age & Sex	None
High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning[42]	*JAMA Cardiology*	Age, Sex, Race, Ethnicity	BMI
Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study[64]	*Lancet Digital Health*	None	None
Severe aortic stenosis detection by deep learning applied to echocardiography[65]	*European Heart Journal*	Age, Sex, Race, Ethnicity	None
Automated coronary calcium scoring using deep learning with multicenter external validation[25]	*NPJ Digital Medicine*	Age, Sex	None
CathAI: fully automated coronary angiography interpretation and stenosis estimation[66]	*NPJ Digital Medicine*	None	Gender, CAD territory
Deep learning interpretation of echocardiograms[56]	*NPJ Digital Medicine*	Age, Sex, Height, Weight	None
Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study[67]	*Lancet Digital Health*	Age, Sex	None
Video-based deep learning for automated assessment of left ventricular ejection fraction in pediatric patients[67]	*Journal of the American Society of Echocardiography*	Age, Sex	Age, Sex, BSA
Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms[68]	*Nature Communications*	Age, Sex	None
Automated assessment of cardiac systolic function from coronary angiograms with video-based artificial intelligence algorithm[69]	*JAMA Cardiology*	Age, Sex, BMI	Sex
Blinded, randomized trial of sonographer versus AI cardiac function assessment[39]	*Nature*	Age, Sex, BMI, Race, Location, Method of EF Evaluation,	Age, Sex, BMI, Race, Patient Inpatient/Outpatient Location, Method of EF Evaluation
Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram[19]	*Nature Medicine*	Age, Sex	Age and Sex

Table 4.

A framework for considering bias mitigation approaches when developing and implementing AI algorithms in medical imaging.

Framework for Considering Approaches to Mitigating Bias Specific to AI Applications in Medical Imaging
Development Stage *Before training (Phase 0–1)* • Identify all potential present and future sources of bias during model planning and design • Seek to maximize inclusion of all relevant populations and subpopulations for training and testing • Appropriately handle all datasets according to industry standards[60,77] (e.g. involve domain expertise, correct labeling, no data or feature loss) • From exploratory data analysis, consider rebalancing datasets as needed (e.g. based on distribution of key covariates and outcomes) *During training (Phase 2)* • Employ bias mitigation model development strategies[76,78] (e.g. data augmentation, adversarial training) • Anticipate and evaluate for potential algorithmic bias[60] *After training (Phase 3)* • Examine performance overall and across subgroups, ideally in a clinical trial setting if possible • Assess fairness metrics (e.g. demographic parity, equalized odds or opportunity, disparate Impact) with attention to potential effects of demographic proxies (e.g. race proxies include patterns of health care utilization)[60] • Compare performance with and without including additional variables relevant to the outcome, e.g. fusion models including clinical, biomarker, genomic, social determinants of health, and other variables • Consider re-training, based on analysis and interpretation of performance metrics • Generate 'model cards' for algorithms[79] and 'datasheets' for datasets[80]
Implementation Stage*(Phase 4a)* • Upon initial real-world deployment, assess performance overall and across subgroups (with fairness metrics interpreted in context) • Transparently report results from initial deployment[81] • Update 'model cards' for algorithms and 'datasheets' for datasets
Management and Monitoring Stage*(Phase 4b)* • Routinely assess performance overall and across subgroups for applications in practice (with and fairness metrics interpreted in context) • Transparently report serially assessed results of applications in practice[81] • Routinely update 'model cards' for algorithms and 'datasheets' for datasets