RESEARCH ARTICLE

Beyond high hopes: A scoping review of the 2019–2021 scientific discourse on machine learning in medical imaging

Vasileios Nittas[1,2], **Paola Daniore**[3,4], **Constantin Landers**[1], **Felix Gille**[3,4], **Julia Amann**[1], **Shannon Hubbs**[1], **Milo Alan Puhan**[2], **Effy Vayena**[1], **Alessandro Blasimme**[1] *

1 Health Ethics and Policy Lab, Department of Health Sciences and Technology, Swiss Federal Institute of Technology (ETH Zurich), Zurich, Switzerland, 2 Epidemiology, Biostatistics and Prevention Institute, Faculty of Medicine, Faculty of Science, University of Zurich, Zurich, Switzerland, 3 Institute for Implementation Science in Health Care, Faculty of Medicine, University of Zurich, Switzerland, 4 Digital Society Initiative, University of Zurich, Switzerland

* alessandro.blasimme@hest.ethz.ch

Abstract

Machine learning has become a key driver of the digital health revolution. That comes with a fair share of high hopes and hype. We conducted a scoping review on machine learning in medical imaging, providing a comprehensive outlook of the field's potential, limitations, and future directions. Most reported strengths and promises included: improved (a) analytic power, (b) efficiency (c) decision making, and (d) equity. Most reported challenges included: (a) structural barriers and imaging heterogeneity, (b) scarcity of well-annotated, representative and interconnected imaging datasets (c) validity and performance limitations, including bias and equity issues, and (d) the still missing clinical integration. The boundaries between strengths and challenges, with cross-cutting ethical and regulatory implications, remain blurred. The literature emphasizes explainability and trustworthiness, with a largely missing discussion about the specific technical and regulatory challenges surrounding these concepts. Future trends are expected to shift towards multi-source models, combining imaging with an array of other data, in a more open access, and explainable manner.

Author summary

Machine learning is becoming an important part of digital health and medical imaging. Many believe that it is the solution to some of the challenges our medical systems currently face. In this study, we reviewed the literature to explore this topic, focusing on the promises, challenges, and future developments. The literature emphasises that machine learning allows us to use medical images in ways that are more reliable and precise, requiring less time and resources. That can lead to better decision-making, as well as allow for more people to access affordable image-based care. Some of the mentioned challenges include the large differences between images and imaging techniques, the difficulty in accessing enough high-quality images, the costs and infrastructure associated with that and the resulting geographic inequalities. In addition, the literature emphasizes that is difficult to understand how machine learning works, as well to assess how valid and reliable

it is in analysing medical images. Equally difficult is its regulation and integration in everyday clinical work. The future is expected to bring machine learning models that will be able to analyse different types of images and other clinical data at once, in ways that are more transparent and understandable.

Introduction

Amidst the so-called digital health revolution, the Lancet and Financial Times Commission on governing health futures 2030 called upon decision-makers and experts to consider digital technologies as modern health determinants [1]. Although the digitalization of healthcare lags behind other industries, the amount of healthcare data generated electronically is now larger than ever [2]. Harnessing the power of that data leaves no alternative besides the use of technology [2]. Technologies such as artificial intelligence (AI) and machine learning (ML) have rapidly become the epicenter of that revolution, as well as of political, scholarly, and public attention [2,3].

Much of the hype around AI and ML is rooted in mainstream media, with claims about the limitless capabilities of data-driven algorithms to achieve disproportionate reach [4–7]. Headlines with terms such as "transforming healthcare" and "reshaping medicine" generate unrealistic expectations, and inevitably skepticism by healthcare professionals and patients alike [7,8]. AI is widely used as a broad umbrella term, rooted in computer science and describing computational methods that allow computers to perform tasks that would traditionally require human intelligence [9,10]. While often used interchangeably with AI, ML is a subset of AI which allows computers to self-sufficiently learn to perform a pre-defined task and possibly improve over time [11].

Medical ML undeniably achieved remarkable progress, particularly in the field of visual pattern recognition [2,12]. ML and its subsets such as deep learning (DL) have been extensively applied in the field of medical imaging, accounting for about 40% of all ML-related publications in the health sciences domain over the past five years. With the amount of imaging data skyrocketing, ML offers efficient ways to use this data for clinical decision-making, ranging from computer-aided diagnosis to radiomics and image-guided therapy [13]. Considering these successes, together with the yet-undefined boundaries of the discipline, uncertainty about the clinical impact of AI is expected [5,7].

It is thus appropriate to ask: what are the prospects of ML to become clinically integrated, and what impediments can slow down or stifle innovation? To go from proof-of-concept models built on retrospective data, to ML systems capable of improving healthcare, the nature of the expected impact of AI in healthcare must be clarified. Moreover, we need to identify perceived bottlenecks and impediments to the successful integration of ML in routine clinical medical imaging practice. Such information is key to inform policy development around a clear vision of AI-driven transformation in the medical imaging sector and to foster the emergence of efficient technological and clinical validation standards.

To this aim, this review explores and critically discusses the latest (2019–2021) scientific discourse around ML in clinical imaging, highlighting the field's promise, challenges, ethical and regulatory implications, and future directions. We focused on the big five chronic conditions, including cardiovascular diseases, diabetes, stroke, chronic respiratory diseases, and cancer. The rationale behind this review's focus is the following. First, we believe that the big five chronic conditions constitute a significant amount of available scientific output, and thus, provide a representative picture of current trends. Second, focusing on a selected number of medical areas ensures that this review's output remains manageable. Third, these five chronic

conditions are responsible for a considerable share of healthy years lost, globally. These areas are therefore some of the most relevant for ML application and impact. To the best of our knowledge, this is the first systematic approach to exploring the scientific discussion around ML in image-based care, with a comprehensive outlook encompassing scientific and socio-technical aspects.

Methods

We conducted a scoping review of narrative reviews and editorials. These two types of publications typically reflect expert opinion about the latest trends and prospects of any given scientific field. Therefore, they are likely to provide a broad picture of current scientific discourse, being published in a shorter timeframe not constrained by complex and lengthy methodologies. Our approach was guided by Arksey and O'Malley's framework, as well as Levac, Colquhoun, and O'Brien's conceptual extensions [14,15].

Search strategy and selection criteria

Guided by the study's aims, a specialized librarian developed a comprehensive search strategy that was applied on six electronic databases, including Medline, Embase, CINAHL, PsycInfo, Scopus, and Web of Science. Searches were run on August 12, 2020, and updated on December 5, 2021, to cover publications until August 31, 2021. We used multiple variations (and synonyms) of the following search terms: machine learning, artificial intelligence, and medical imaging. We only included narrative reviews, editorials and commentaries, written in English. An example of our search strategy (Web of Science) is provided in S1 File. In addition, within the same time frame, we hand-searched four of the most prominent journals in the field, including the Journal of Medical Internet Research, npj Digital Medicine, Lancet Digital Health, and Nature Machine Intelligence.

We conducted title, abstract, and full-text screening in duplicate and independently, guided by a set of predefined exclusion criteria (see Table 1) and excluding papers if any of the below criteria were true. Any disagreements were resolved by a third reviewer. For title and abstract screening, we utilized two web-based systematic reviewing tools, DistillerSR (Evidence Partners) and Rayyan [16,17].

Data extraction

We extracted (single reviewer) data with a predefined data extraction sheet that captured the study's aims, including (a) strengths and promises of ML use in clinical imaging, (b) the

Table 1. Selection Criteria.

[1] Addressing ML (and sub-fields, e.g., deep learning) AND Imaging. Studies using the term "AI" interchangeably with ML are included.
  ○ ML & Imaging should be addressed either in (a) title and/or (b) abstract AND throughout the main body
[2] A narrative review, commentary or editorial
[3] Focused on one (or multiple) of the five chronic conditions:
  - Cardiovascular diseases
  - Cancer
  - Chronic Respiratory diseases
  - Diabetes
  - Stroke
[4] Not a technical paper (e.g., describing algorithms without conveying a scientific opinion, neither discussing the broader clinical, social, ethical, or regulatory implications of ML)
[5] Published after the first of August 2019 and before the first of September 2021
[6] Written in English

https://doi.org/10.1371/journal.pdig.0000189.t001

challenges (including weaknesses, barriers), (c) the potential solutions to these challenges, and (d) the field's future trends. We initially tested the form on a sample of 15 studies, which allowed for further refinement and adjustment. Finally, we validated the data extraction sheet through multiple rounds of internal reviews.

Data synthesis and reporting

We synthesized our findings following a qualitative and iterative thematic approach, conducted by one reviewer and quality-checked by a second reviewer [15]. After data familiarization, we generated initial codes inductively. We then clustered emergent themes and synthesized these in conceptual maps that provided an overall picture of our findings. The reporting of our results was based on the Preferred Reporting Items Extension for Scoping Reviews (PRISMA-ScR) statement [18].

Results

Our database searches yielded a total of 8079 references, and our hand searches 23. Of these, 7249 were excluded at the title and abstract screening. The full texts of 673 papers were assessed, leading to 561 further exclusions. We finally included 112 publications, listed in S2 File. Fig 1 provides the PRISMA flow chart for our screening and study inclusion process [19,20].

Most included publications (published between August 2019 to August 2021) were narrative reviews (n = 92, 82%), followed by editorials and short opinion-based papers (n = 20, 18%). First authorship was affiliated with institutions in the European region in 48 (43%) publications, followed by 38 (34%) from North America, 23 (20%) from Asia, and three (3%) from Australia. Most publications addressed the use of image-based ML in oncology (n = 51, 45%), followed by cardiology (n = 39, 35%), respiratory diseases (n = 5, 4%), diabetes (n = 3, 3%) and stroke (n = 3, 3%). Eleven (10%) studies had a multi-disease focus. The following paragraphs provide our findings around: (1) the strengths of ML use in medical imaging, (2) the challenges of ML use in medical imaging, with a focus on ethical and regulatory challenges, (3) the potential solutions to these challenges and (4) future trends.

About 60% (n = 66) of all included publications primarily focused on a single and 40% on multiple imaging modalities (n = 46). Among these, computed tomography (CT) was the most commonly addressed (n = 46, 41%), followed by magnetic resonance imaging (MRI) (n = 33, 29%), ultrasound imaging (n = 20, 18%), nuclear imaging (n = 22, 20%), digital pathology (n = 11, 10%), X-ray imaging (n = 8, 7%), and retinal fundus imaging (n = 3, 3%). About 20% (n = 22) of included publications primarily focused on the use of ML for classification tasks, most often employing artificial neural networks, convolutional neural networks and support vector machines. Among those, most publications focused on cancer (n = 10), followed by cardiovascular diseases (n = 6), pulmonary diseases (n = 4) and diabetes (n = 2). About 14% (n = 16) of publications focused solely on prediction and regression, all of them addressing either cancer (n = 10) or cardiovascular diseases (n = 6). The remaining 66% (n = 74) of publications simultaneously addressed both classification and prediction.

Strengths/Promises

Almost all publications (n = 110, 98%) reported at least one expected strength of ML applications in imaging (with two not mentioning any). We identified four closely related but distinct domains: (a) analytic power, (b) efficiency, (c) clinical impact, and (d) equity. Each of these contained subdomains, described in Table 2.
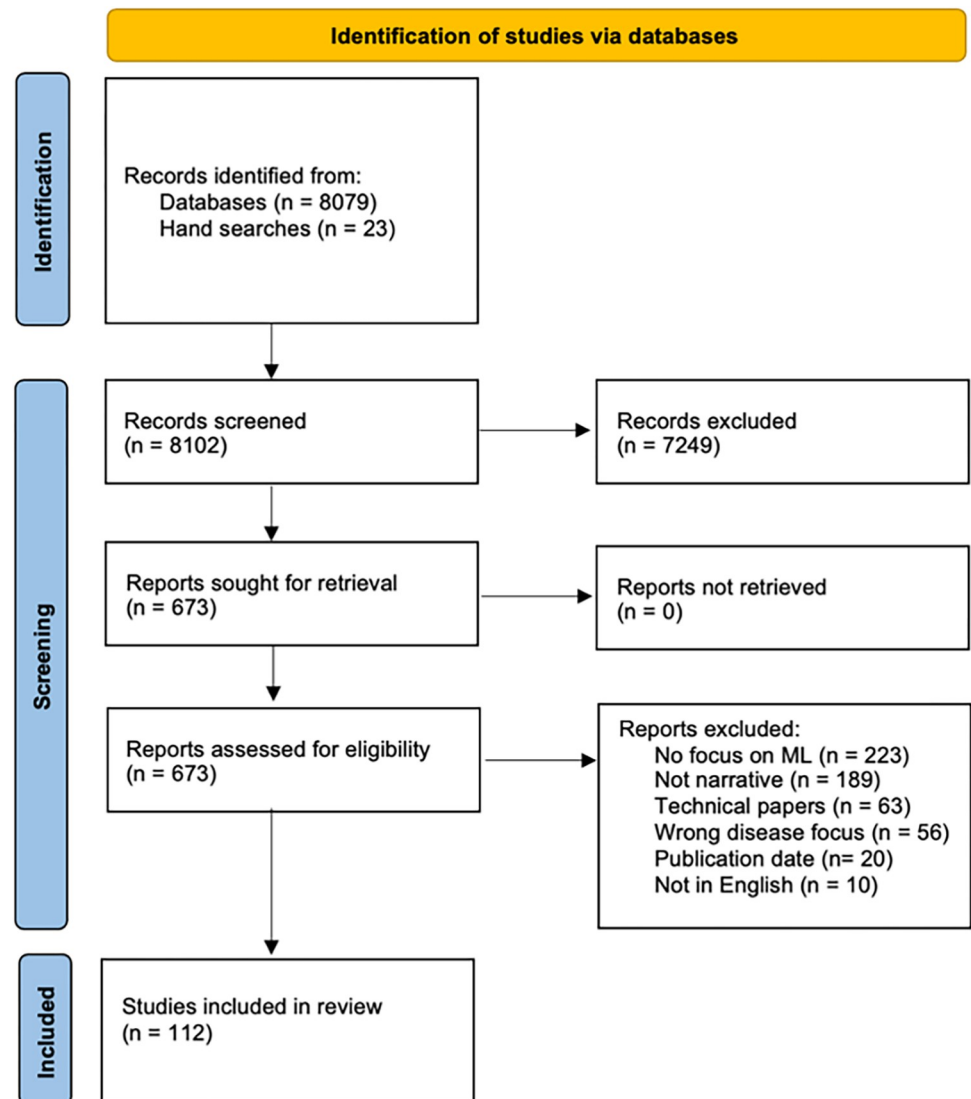
Fig 1. PRISMA flowchart

https://doi.org/10.1371/journal.pdig.0000189.g001

Challenges/Impediments

Most publications (n = 102, 92%) reported at least one challenge of ML applications in medical imaging (with ten not mentioning any). We distinguish between those challenges mostly related to imaging, and those generally pertinent to ML. The first are summarized in Table 3, the second in Table 4. The ML challenges were further divided into three distinct domains, including (a) structural barriers, (b) validity and performance, and (c) clinical relevance.

The challenge of image heterogeneity was mostly discussed in the context of cancer and cardiovascular diseases [9,21–29]. This is particularly applicable to imaging modalities with no standardized pixel values or agreed format, such MRI and digital pathology imaging, or those with high noise and low resolution, such as intravascular ultrasounds [23,30–32]. In both cases, interobserver reliability is high and feature extraction challenging [31,33]. Besides the inevitable heterogeneity arising from the use of different image generation, reconstruction and preprocessing techniques, there is a large impact by confounders, such as noise due to artifacts

Table 2. Reported (expected) strengths and promises of image-based machine learning.

Domain	Subdomain(s)	Descriptions
Analytic power	accuracy	*ability to classify images as well as experts, ensuring diagnostic and prognostic accuracy*
	superiority	*ability to differentiate images better than an expert, identifying patterns not always visible to the human eye*
	objectivity	*lack of human subjective biases and errors, reducing variability and improving comparability and reproducibility*
	big data use	*ability to handle and analyse large amounts of data and tackle the challenges of big data*
Efficiency	time efficiency	*ability to speed up analysis and clinical translation through automation of otherwise time-consuming manual tasks*
	cost efficiency	*ability to lower direct and indirect costs through time and diagnostic efficiency, automation, and enhanced workflows*
Clinical impact	workflow improvements	*ability to optimize clinical workflows through integrating, automating, streamlining, and structuring processes*
	decision making	*supporting faster, cheaper, more accurate and higher-level decision making and clinical interpretations*
	personalization	*facilitate personalized care, though higher analytic power, efficiency and improved clinical workflows*
Equity	reach, access, and affordability	*promise of increased geographic reach of and better access to affordable image-based healthcare*

https://doi.org/10.1371/journal.pdig.0000189.t002

(e.g. implants in CT images) or motion, which is especially relevant to heart imaging [22,29,34]. Authors mentioned that heterogeneity increases as the dimensionality, number and complexity of images increases, making it hard for features to be extracted and used for ML training purposes [28,35]. Modalities such as echocardiography create large datasets, with varying operator-depending image quality, yielding high complexity and interrater variability [36,37]. This is applicable to cardiac imaging in general, which is dynamic and generated while the heart is in constant motion, posing particular challenges for ML [36,37] Variation, noise, and complexity ultimately hinder the extraction of information from images, as well as the determination of ground truth, which is essential during training and validation of ML algorithms [9,27,38].

The next four challenges of image and annotation scarcity, class imbalance, image silos and cost all fall under the category of missing resources and infrastructures in medical imaging. Image scarcity was primarily mentioned in the context of neuro-oncology, particularly with

Table 3. Image-related challenges of machine learning, as reported in the literature.

Domain(s)	Descriptions
Image heterogeneity	*high image variation due to non-standardized acquisition, reconstruction and preprocessing protocols*
Image scarcity	*image scarcity, especially in complex fields such as neuro-oncology and rare diseases*
Annotation scarcity	*lack of properly annotated image datasets, and if annotated, subject to high inter-rated variability*
Class imbalance	*datasets often include too few cases (images indicating disease)*
Image silos	*large volumes of imaging data exist in silos, with low interconnectivity*
Cost	*access to medical images often costly, requiring adequate training and infrastructural support*
Subjectivity	*image selection, segmentation and annotation conducted by human experts and not entirely free of subjective biases*
Inequities / bias	*most advanced imaging modalities mostly accessible in high-income countries and to high-income population subgroups, leaving many underrepresented*

https://doi.org/10.1371/journal.pdig.0000189.t003

Table 4. General machine learning challenges, as reported in the literature.

Domain	Subdomain(s)	Descriptions
Structural barriers	assessment	*challenging validation, evaluation, and prediction of algorithm performance*
	regulation	*lacking regulatory and reimbursement frameworks, inhibiting standardization and creating liability uncertainties*
	resources	*algorithm development requires extensive time, financial, and human resources*
Validity and performance	generalizability	*training with small samples and lacking external validation, limiting broader applicability*
	reliability	*training with small samples and lack of testing, leading to often overfitted algorithms, uncertainty, and poor performance*
	bias	*prone to systematic biases (e.g., learning bias, social biases)*
	narrowness (task)	*limited to solving isolated tasks, unlike human intelligence and expertise, which can grasp clinical complexities in full*
	narrowness (data)	*most algorithms not capable of integrating multiple data sources simultaneously (e.g., images and health records)*
Clinical relevance	explainability	*inherently complex, with data-driven, non-intuitive features, limited understanding of the logic behind algorithms' decisions*
	integration	*lacking clinical integration due to structural barriers, uncertainty, missing training, low explainability and trust*
	human factors	*human factors (e.g., cognition, biases, demographics, experience) mostly overlooked in the design and development of ML algorithms*
	clinical performance	*most current studies retrospective and "offline", not providing evidence on the clinical performance of algorithms in multidimensional and dynamic settings (e.g., real-world conditions)*

https://doi.org/10.1371/journal.pdig.0000189.t004

very heterogeneous tumours like Glioblastoma Multiforme, as well as in the context of nuclear imaging [21,39]. Authors suggested that where images are abundant (e.g., in areas such as breast imaging), they often lack adequate annotations, as these are often not required for regular clinical workflows [40,41]. On the opposite, detailed enough annotations depend on scarce resources such as time, effort and training [40,41]. Some authors mentioned that a plethora of images are stored in silos, across different institutions and platforms, making access and interconnectivity difficult [42,43]. Within these silos, the number of images from healthy individuals is often significantly larger than the number of images indicating disease [41,44]. In addition, when disease is indicated (e.g. malignant nodule) the occupied region of interest is often relatively small and hard to detect [24]. This class imbalance makes a dataset less valuable for ML training and was mentioned in the context of cancer and diabetic retinopathy [41,44]. Primarily mentioned in the context of stroke, images alone are often not powerful enough for creating clinically useful ML tools. Predicting or classifying conditions like stroke depends on an array of contextual data, which images alone do not provide [45]. Finally, accessing medical images to train ML algorithms can be a costly endeavour, requiring large capital costs, skilled and trained personnel and adequate infrastructures [46]. This is particularly relevant to the generation and use of digital pathology images, as the digitalization of pathology workflows and the transition to whole-slide imaging comes with particularly high costs [47].

The two final mentioned challenges are those of subjectivity and inequities. Considering that images are often selected, segmented, interpreted and annotated by individuals, subjectivity and bias cannot be entirely ruled out [48]. In addition to that, advanced imaging modalities are mostly accessible in high-income countries and among higher income population subgroups. This generates image datasets that underrepresent low- and middle-income countries and population subgroups from lower socio-economic strata [49]. A strong example of such inequities is in the field of cardiac imaging, with newest imaging technologies often being geographically and socio-demographically limited [49].

Many of the challenges listed in Table 4 have an ethical and/or regulatory dimension. Authors argue that lack of regulatory frameworks limits fairness and increases the risk of biased algorithms and widening inequities [4,37,50–54]. Such inequities are further exacerbated by ML's costs (e.g. technology, training, staff), which limit its use to high-resource settings [55]. Liability was mentioned as a further challenge. The main question posed was: in

case of harm (e.g., due to misdiagnosis), who is responsible? The physician, institution, or algorithm developer? [56–59]. Some argued that ML applications are merely assistive, and thus liability falls with the treating physician [49,58,60]. Others emphasized that many factors (e.g., integration of a tool within an institution or how it changes with use) make liability difficult to assign, further limiting trust and hindering clinical uptake [40,49,61].

The data-dependency of ML algorithms comes with concerns around privacy, security, and data ownership [25,46,50,59,62,63], including secure storage and sharing of data, and protection from data breaches and attacks [54,64]. Authors argue that ML's need for large volumes of images for training and validation poses security and consent challenges. Individualized consent is rarely feasible, creating the need for alternative solutions, such as opt-out consent or broad consents, in which individuals consent to multiple secondary uses of their data, balancing between data availability and privacy [49,54,65]. Even after consent is obtained, the anonymization of images (to which identifiers are often attached) poses a significant processing burden and risk for re-identification [4]. Finally, the question arises of whether patients should be given the option to not include ML in their care [58,66].

Authors also argue that low explainability could reduce trust, in turn hindering clinical acceptance and uptake [9,32,49]. That may potentially affect the patient-physician relationship, interfering in decision-making and eventually leading to unease or confusion [42,56,67]. An additional challenge is the perception of competition between ML and human intelligence, with healthcare providers fearing gradual replacement, although this was often debunked as unlikely [27,59,68,69]. As overreliance on ML can lead to automation bias, increase mistakes, cause avoidable harm, and deskill physicians [54], clinicians should view ML tools as systems to support rather than replace their decision making.

Suggested solutions for identified challenges

Around 80% of publications (n = 81) provided potential solutions to these challenges. Some argued for facilitating access to high-quality data and existing algorithms, pooling of resources, and dialogue between stakeholders, while ensuring that privacy, ethics, and commercial interests are aligned and safeguarded [30,50,70–73]. Others called for more standardization, including well-defined imaging protocols and research methods (e.g. image acquisition), as well as open access and easily available algorithm codes, to enable reproducibility [9,21,51,74,75]. Standardizing how and what information is extracted from images for ML purposes is another proposed solution [22]. Authors mentioned the Image Biomarker Standardization Initiative, which provides guidelines on how that is to be achieved, yet these remain to be adjusted and adapted to the needs of more challenging imaging areas, such as cardiac imaging [9,71]. The exclusion of image features that are too sensitive to variation has been another suggested solution towards increased standardization [9]. For generalizability, authors emphasized the need for more and better-performed external validation, using multicentric imaging datasets obtained under real-world conditions and representative of the target patient group [4,9,11,22,28,46,72,76–78]. Transfer learning, crowd-sourced, and open-access image repositories, and mathematically generated datasets are alternatives if data scarcity cannot be solved [41,71,79–81].

Beyond enhancing data availability and quality, Radakovich and colleagues highlight the risk of overfitting, which can be regulated through penalization of very complex models or image augmenting techniques that increase heterogeneity (e.g. adding image noise) in datasets with which ML algorithms are trained [80]. A further approach to improving reliability involves training algorithms away from strict outcomes (e.g. disease is/is not present) towards outcomes that communicate uncertainty, a third "grey zone" category [82]. Class imbalance

issues should be addressed with more research into advanced deep learning algorithms that can deal with a low number of cases (images indicating disease) [41]. Some authors pointed to regulatory frameworks as essential for ensuring safe and fair ML; yet they lag behind innovation, and are challenged by the continuously evolving nature of ML algorithms [4,49,54,60,83]. Malpractice insurance must be adjusted to the unique liability challenges posed by AI and ML, complemented by comprehensive risk reduction plans [54].

Acknowledging the black box nature of ML, some authors argued that data science and advanced statistical and computational methods should be integrated into the education of medical professionals [58,59,84–87]. Similarly, ML developers should be educated on clinical needs and challenges, including clinical workflows, so that products avoid unnecessarily delaying clinical practice, overburdening practitioners, or negatively affecting the patient-physician relationship [50,54]. Engagement was repeatedly mentioned, including meaningful inclusion of clinicians, patients, and other stakeholders in the development of ML algorithms [22,39,49,57,88,89]. Finally, some authors argue that ML should be accompanied by understandable feedback on why a decision or classification was made [34]. For example, saliency heat maps visually indicate which parts of an image contribute to an ML decision, confidence intervals communicate uncertainty, and interactive dashboards help physicians better understand a system [12,24,32,40,42,61,90].

Clinical utility and integration

Some publications included in this review highlighted that clinical integration has not yet been achieved, and thus the clinical utility of ML is limited [9,29,91]. The authors suggested potential reasons for this. First, most ML studies fail to compare their algorithms to current gold standards and instead retain a technical, proof-of-concept focus [9,30]. Others argued that the performance of most ML algorithms is not assessed based on their impact on disease outcomes, quality of care, or cost-effectiveness, with most studies adopting rather non-transparent workflows that do not resemble real-word settings [24,36,38,42,50,76,87,92]. Thus, it remains unknown how many of these tools perform under complex and unpredictable conditions, or how they affect the broader planning and provision of care [40,42,59].

To address this, some authors suggest large, multi-center prospective trials and quasi-experimental studies, embedded in clinical settings and focusing on downstream impact, including patient outcomes, safety, efficacy, and acceptance [36,50,63,78,89,93]. These shall be complemented by economic evaluations, evaluating cost-effectiveness relative to standard care [40,50]. Furthermore, it is essential to account for the human factor when developing and evaluating ML tools, understanding the impact of algorithms on clinical behavior, interactions between physician and machine, and effects of human traits (e.g., cognition, bias, demographics, experience) on algorithm performance [53,64,94]. Finally, authors raised concerns that ML studies are not driven by clinical relevance but rather by the availability of datasets, with little effort to address the black box nature and bias-related challenges of ML algorithms [35,42,84].

Future directions

One often discussed future trend was the capability of ML algorithms to combine multiple data sources [21,71,81,85,86,95–97], including biological and biochemical data, genomic information, socio-demographics, life-style risks, linkages to electronic health records, and other clinical information [9,22,23,34,68]. Multi-data ML (or multi-omic ML) is expected to inform and improve clinical workflows [21,22,36,42,68,98,99]. Mobile and wearables devices and sensors generate large volumes of yet-untapped health information, and are expected to

democratize and contribute significantly to multi-omic ML, reducing costs and accessibility [35,57,63,69,96,100]. Yet, the actual application of multi-omic ML remains limited, with first efforts being reported in the field of oncology and cardiac imaging (e.g., combining echocardiographic and other clinical data for heart failure diagnosis; combining radiological tumour data with an array of physiological and genomic data) [69,95].

While advocating for multi-data ML, the literature does not provide insights into which types of image modalities yield highest compatibility with physiological and other clinical data. In fact, most publications did not focus on a single imaging modality but imaging in general. This was followed by publications on CT, MRI and echocardiography. On the other hand, the need to combine images with other data might arise from the area of application, instead of the imaging modality per se. Using ML for stroke detection is a good example of that. Decision support in stroke strongly depends on subtle clinical findings and multiple contextual data. For an ML tool to provide value in stroke care, it must be trained with such rich contextual and clinical data, which includes images and physiological data [45]. Similally, to fully and better understand the pathological mechanisms of certain tumors might require ML models that take into account an array of data, combined with images [22].

Another desired trend is that of higher standardization of image generation and processing, as well as multistakeholder engagement [11]. The future should see an emphasis on collaboration between academia, industry, and healthcare institutions to facilitate knowledge exchange and data sharing around AI and ML [11]. For ML tools to be clinically useful, their future use must be rooted in clinical need, while clinicians of the future will, in turn, need to understand the basics of ML and data science [36,45,57,101]. Shared, open-access databases for ML training and comparison are essential to the field's future [31,60].

Finally, the development of robust and less labour-intensive ML, shifting from hand-engineered input to algorithms that automatically extract required information, was another major trend [32,78]. Future ML will be more robust to noise, more capable of learning from smaller or imbalanced datasets, and designed to shift away from binary (yes, no) outcomes to also communicate uncertainty [41,82,102].

Discussion

The scientific discourse around the use of ML in medical imaging evolves rapidly. While the promises and potential of ML in imaging has been prominently defined, the current scientific discourse highlights multiple remaining challenges. Undoubtedly, medical images differ significantly from any other type of images. They are highly heterogenous, multidimensional, often complex and subject to high interrater variability. For some medical fields they are scarce and difficult to access. Datasets are often incomplete, of low quality, imbalanced, inadequately labeled or annotated, and contain an abundance of images that are of lower value for ML. Current ecosystems are divided in silos, in which large amounts of images are stored, fully separated and with no interconnectivity channels, further increasing heterogeneity and access costs. In the context of ML, all these challenges above present major hurdles, as they limit access to and the use of images for training and validation purposes.

Many of these challenges identified in the literature go well beyond imaging and broadly apply to medical ML in general. However, with imaging being the largest subdomain of medical ML, it can serve as a good indicator for the state of the broader field. Our findings suggest the line between the strengths and challenges of image-based AI is often blurred, with equity as one example. While some scholars argue that ML will increase access and affordability of healthcare, allowing broader geographic reach, others highlight an algorithmic divide, due for instance to bias in training datasets [49,57,103]. Another example is time efficiency. Despite

the emphasis on the time-saving nature of ML, actual clinical benefits for physicians and patients are debatable. Does the time gained allow for less tedious work and more time available for consultation, or will it rather lead to higher workloads, with a higher number of patients seen in less time [3]?

Trustworthiness, perceived as a driver of acceptance and adoption, is explicitly and implicitly present in the discussion around ML, related to model accuracy, reproducibility, generalizability, and explainability [56,67,104–106]. Each of these aspects is ultimately linked to trust, in line with the emphasis placed on trust by policy and regulatory bodies. The European Commission's Joint Research Centre, for instance, calls for an ethics and trust framework that will enable the safe use of AI and ML, while the US Department of Health and Human Services vows to support AI systems that inspire trust and respect privacy and security [107,108]. The High-level Expert Group on Artificial Intelligence established by the European Commission has issued Ethical Guidelines for Trustworthy AI [109]. While not specific to health-related uses of AI, such guidelines emphasize the centrality of trustworthiness. Yet, some scholars challenge the notion of trust in machines, characterizing hype around ML trustworthiness as inappropriate and ineffective. Trust is something given voluntarily, with the trustor vulnerable and reliant on the goodwill of the trustee and cannot be transferred to a relationship between a human and machine. Instead of developing abstract guidelines that foster trust, it may be wiser to focus on developing reliable algorithms, underpinned by policy and regulation [110].

Our findings confirm a strong emphasis on explainability. Explainability is a major discussion topic in the field of medical AI and is not limited to AI for medical imaging. This corresponds with accumulating policy-level pressure. The European Union's General Data Protection Regulation references transparency requirements and data subjects' right to "meaningful information about the logic behind automated decisions using their data" [106,111–113]. Similarly, in its proposed AI Regulation, the European Commission prescribes that high-risk AI systems (which include clinical-grade ML) "shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately" [113]. Yet, there is wider scepticism about the value and feasibility of explainability [114,115]. Ghassemi and colleagues refer to explainability as false hope, as most models are too complex and multi-dimensional to be explained accurately [106]. While explanation efforts (e.g., heat maps in imaging) may provide broad indications of how an algorithm operates, their contribution to understanding individual decisions is superficial, and in some instances misleading [106]. A further challenge is that such explainability methods lack performance indications and add additional risk for error [106]. For example, heat maps indicate how image regions contribute to an algorithm's decision, but not whether these regions are truly clinically relevant [106]. This discussion raises questions over the requirement of explainability of ML in the clinical setting, with important issues of whether and in which circumstances AI should be explained to patients remaining unsettled [114,115].

While explainability remains an important topic, greater attention should be paid to the assessment of ML algorithms, as proposed by other scholars. Most evaluation methods fail to capture what matters most to patients, i.e., the impact of an algorithm on their care and outcomes [116,117]. Comprehensive evaluation frameworks are needed to capture the performance of an algorithm in the clinical setting, and the care-related actions it triggers [116]. In the diagnostic testing field, such evaluations are classified as phase 5 and 6 studies, primarily consisting of randomized controlled trials [118]. Economic evaluations should be an integral part of such frameworks, assessing cost-effectiveness relative to the standard care [40,50]. Cost-saving is an often-mentioned benefit of ML, with little clinical evidence to support this [87].

The predicted evolution of medical ML towards increasingly complex models also warrants consideration. As more data types are integrated into image recognition tasks, the scope of medical ML is likely to expand from simple binary classification in a diagnostic context, to more subtle forms of prognostic assessment and treatment decision support tools. Performance evaluation, validation, and clinical implementation will likely require more granular guidance, improved standards, and best practices. Integrating large-scale omic data may require *ad hoc* adaptations to account for clinical and regulatory uncertainties linked to its use.

Broader regulatory implications for medical ML

To keep pace with the rapid progress of ML, we identified an urgent need for comprehensive regulatory and governance models that go beyond abstract principles [60,65,70,105,119]. Abstract regulatory and ethical principles can provide a false sense of safety, and bear the risk of being misused to increase trust and acceptability [119]. Concrete guidelines, sector-specific standards, and clear regulatory requirements are needed to promote responsible innovation in the medical AI [119]. As suggested by Reddy and colleagues as well as the WHO Ethics and Governance of Artificial Intelligence for Health Guidance, such frameworks should be comprehensive and target equity, transparency, trustworthiness, accountability, and openness [105,120]. Interdisciplinarity should take center stage, fostering a dialogue between law and ethics experts, computer scientists, healthcare professionals, and patients [3,119].

The dynamic nature of ML also requires a degree of regulatory flexibility, one that acknowledges uncertainty and embraces change and adaptation [121,122]. In 2019, the US Food and Drug Administration (FDA) proposed a "Predetermined Change Protocol", through which manufacturers can report expected algorithm changes, how these changes will occur, and how continued safety will be ensured [61,123]. The FDA subsequently published the AI/ML Software as a Medical Device Action Plan, vowing to develop the proposed regulatory framework and contribute to good ML practice [124]. Its Pre-Certification Pilot program, targeting the regulation and oversight of software-based medical devices, was developed to cover AI/ML-based products and is an important response to the constantly evolving AI/ML landscape.

Such regulatory approaches are an important paradigm shift towards embracing the dynamic nature of ML without discounting safety or timely technology access [61]. Inevitably, such frameworks will depend on a proactive approach to ethical ML development, such as through wider implementation of ethics by design [119], which ensures that ethical challenges are acknowledged and addressed at each stage of ML development [119]

Limitations

The findings of the present review should be viewed and interpreted with the following limitations in mind. For the purpose of keeping the volume of eligible studies within a manageable range, we decided to focus on a fraction of the existing literature, mainly addressing the big five chronic conditions. This decision was based on the consideration that the big five chronic conditions cover a large and important fraction of currently existing literature in the field. Nonetheless, that focus provides an incomplete picture of published work.

Conclusion

Our work systematically and comprehensively captured the latest scientific discourse around the use of ML in medical imaging across the continuum of care for the big five chronic diseases. We found a broad range of reported strengths and limitations, ethical and regulatory challenges, as well as pointers for future trends. Medical images, and the ecosystems they are embedded in, present unique challenges. High heterogeneity, complexity, variations in image

quality, scarcity of well-annotated datasets and access hurdles all limit their use for ML training and validation. Among the more broadly applicable ML challenges, we found that the boundaries between strengths and limitations are often blurred, placing a strong positive emphasis on explainability and trustworthiness, with a largely missing discussion about the specific technical and regulatory challenges surrounding these concepts. The uncertainties around the effective translation of image-based ML tools from proof-of-concept phases to clinical integration are equally emphasized. Further efforts are needed to promote this translation, including multistakeholder engagement that takes into consideration the dynamic and complex ecosystem of technical, ethical, and regulatory constraints. Finally, the future of image-based ML is expected to shift towards multi-source models, combining imaging with an array of other data, in a more open access, and explainable manner.

Supporting information

S1 File. Search strategy (Web of Science)
(DOCX)

S2 File. List of in included studies
(XLSX)

S3 File. Data extraction sheet
(XLSX)

Author Contributions

Conceptualization: Vasileios Nittas, Milo Alan Puhan, Effy Vayena, Alessandro Blasimme.

Data curation: Vasileios Nittas, Paola Daniore, Constantin Landers, Felix Gille, Julia Amann, Shannon Hubbs.

Formal analysis: Vasileios Nittas.

Funding acquisition: Effy Vayena, Alessandro Blasimme.

Investigation: Alessandro Blasimme.

Methodology: Vasileios Nittas, Milo Alan Puhan, Effy Vayena, Alessandro Blasimme.

Project administration: Effy Vayena, Alessandro Blasimme.

Resources: Effy Vayena, Alessandro Blasimme.

Supervision: Milo Alan Puhan, Effy Vayena, Alessandro Blasimme.

Validation: Felix Gille, Julia Amann, Alessandro Blasimme.

Writing – original draft: Vasileios Nittas, Shannon Hubbs.

Writing – review & editing: Vasileios Nittas, Paola Daniore, Constantin Landers, Felix Gille, Julia Amann, Shannon Hubbs, Milo Alan Puhan, Effy Vayena, Alessandro Blasimme.

References

1. Kickbusch I, Piselli D, Agrawal A, Balicer R, Banner O, Adelhardt M, et al. The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world. Lancet. 2021; 398: 1727–1776. https://doi.org/10.1016/S0140-6736(21)01824-9 PMID: 34706260

2. Emanuel E, Wachter R. Artificial Intelligence in Health Care Will the Value Match the Hype? N Engl J Med. 2003; 348: 2635–2645. https://doi.org/10.1056/nejmsa022615

3. Murphy A, Liszewski B. Artificial Intelligence and the Medical Radiation Profession: How Our Advocacy Must Inform Future Practice. J Med Imaging Radiat Sci. 2019; 50: S15–S19. https://doi.org/10.1016/j.jmir.2019.09.001 PMID: 31611013

4. Angehrn Z, Haldna L, Zandvliet AS, Gil Berglund E, Zeeuw J, Amzal B, et al. Artificial Intelligence and Machine Learning Applied at the Point of Care. Front Pharmacol. 2020; 11: 1–12. https://doi.org/10.3389/fphar.2020.00759 PMID: 32625083

5. Del Vento D, Fanfarillo A. Traps, pitfalls and misconceptions of machine learning applied to scientific disciplines. ACM Int Conf Proceeding Ser. 2019. https://doi.org/10.1145/3332186.3332209

6. Floridi L. AI and Its New Winter: from Myths to Realities. Philos Technol. 2020; 33: 1–3. https://doi.org/10.1007/s13347-020-00396-6

7. Leufer D. Why We Need to Bust Some Myths about AI. Patterns. 2020; 1: 100124. https://doi.org/10.1016/j.patter.2020.100124 PMID: 33205143

8. Shortliffe EH. Artificial Intelligence in Medicine: Weighing the Accomplishments, Hype, and Promise. Yearb Med Inform. 2019; 28: 257–262. https://doi.org/10.1055/s-0039-1677891 PMID: 31022745

9. Christie JR, Lang P, Zelko LM, Palma DA, Abdelrazek M, Mattonen SA. Artificial Intelligence in Lung Cancer: Bridging the Gap Between Computational Power and Clinical Decision-Making. Can Assoc Radiol J. 2021; 72: 86–97. https://doi.org/10.1177/0846537120941434 PMID: 32735493

10. Yan Y, Zhang JW, Zang GY, Pu J. The primary use of artificial intelligence in cardiovascular diseases: What kind of potential role does artificial intelligence play in future medicine? J Geriatr Cardiol. 2019; 16: 585–591. https://doi.org/10.11909/j.issn.1671-5411.2019.08.010 PMID: 31555325

11. Lin A, Kolossváry M, Išgum I, Maurovich-Horvat P, Slomka PJ, Dey D. Artificial intelligence: improving the efficiency of cardiovascular imaging. Expert Rev Med Devices. 2020; 17: 565–577. https://doi.org/10.1080/17434440.2020.1777855 PMID: 32510252

12. El Naqa I, Haider MA, Giger ML, Ten Haken RK. Artificial intelligence: Reshaping the practice of radiological sciences in the 21st century. Br J Radiol. 2020;93. https://doi.org/10.1259/bjr.20190855 PMID: 31965813

13. Suzuki K, Zhou L, Wang Q. Machine learning in medical imaging. Pattern Recognit. 2017; 63: 465–467. https://doi.org/10.1016/j.patcog.2016.10.020

14. Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. Int J Soc Res Methodol Theory Pract. 2005; 8: 19–32. https://doi.org/10.1080/1364557032000119616

15. Levac D, Colquhoun H, O'Brien K. Scoping studies: advancing the methodology. Implement Sci. 2010; 5: 1–9. https://doi.org/10.1029/2003JD004173.Aires

16. Rayyan Systems Inc. Rayyan Intelligent Systematic Review. 2021.

17. Evidence Partners. DistillerSR.

18. Tricco AC, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018;169.

19. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Rev Esp Nutr Humana y Diet. 2016; 20: 148–160. https://doi.org/10.1186/2046-4053-4-1 PMID: 25554246

20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372. https://doi.org/10.1136/bmj.n71 PMID: 33782057

21. Lohmann P, Galldiks N, Kocher M, Heinzel A, Filss CP, Stegmayr C, et al. Radiomics in neuro-oncology: Basics, workflow, and applications. Methods. 2021; 188: 112–121. https://doi.org/10.1016/j.ymeth.2020.06.003 PMID: 32522530

22. Chinnery T, Arifin A, Tay KY, Leung A, Nichols AC, Palma DA, et al. Utilizing Artificial Intelligence for Head and Neck Cancer Outcomes Prediction From Imaging. Can Assoc Radiol J. 2021; 72: 73–85. https://doi.org/10.1177/0846537120942134 PMID: 32735452

23. Mri M, Bardis MD, Houshyar R, Chang PD, Ushinsky A, Glavis-bloom J, et al. Applications of Artificial Intelligence to Prostate Multiparametric MRI (mpMRI): Current and Emerging Trends. 2020.

24. Geras KJ, Mann RM, Moy L. Artificial intelligence for mammography and digital breast tomosynthesis: Current concepts and future perspectives. Radiology. 2019; 293: 246–259. https://doi.org/10.1148/radiol.2019182627 PMID: 31549948

25. Frix AN, Cousin F, Refaee T, Bottari F, Vaidyanathan A, Desir C, et al. Radiomics in lung diseases imaging: State-of-the-art for clinicians. J Pers Med. 2021;11. https://doi.org/10.3390/jpm11070602 PMID: 34202096

26. Zhou Y, Xu X, Song L, Wang C, Guo J, Yi Z, et al. The application of artificial intelligence and radiomics in lung cancer. Precis Clin Med. 2020; 3: 214–227. https://doi.org/10.1093/pcmedi/pbaa028 PMID: 35694416

27. Oikonomou EK, Siddique M, Antoniades C. Artificial intelligence in medical imaging: A radiomic guide to precision phenotyping of cardiovascular disease. Cardiovasc Res. 2020; 116: 2040–2054. https://doi.org/10.1093/cvr/cvaa021 PMID: 32090243

28. Xu B, Kocyigit D, Grimm R, Griffin BP, Cheng F. Applications of artificial intelligence in multimodality cardiovascular imaging: A state-of-the-art review. Prog Cardiovasc Dis. 2020; 63: 367–376. https://doi.org/10.1016/j.pcad.2020.03.003 PMID: 32201286

29. Lee H, Martin S, Burt JR, Bagherzadeh PS, Rapaka S, Gray HN, et al. Machine Learning and Coronary Artery Calcium Scoring. Curr Cardiol Rep. 2020; 22: 1–6. https://doi.org/10.1007/s11886-020-01337-7 PMID: 32647932

30. Park JE, Kickingereder P, Kim HS. Radiomics and deep learning from research to clinical workflow: Neuro-oncologic imaging. Korean J Radiol. 2020; 21: 1126–1137. https://doi.org/10.3348/kjr.2019.0847 PMID: 32729271

31. Fedewa R, Puri R, Fleischman E, Lee J, Prabhu D, Wilson DL, et al. Artificial Intelligence in Intracoronary Imaging. Curr Cardiol Rep. 2020;22. https://doi.org/10.1007/s11886-020-01299-w PMID: 32472189

32. Ibrahim A, Gamble P, Jaroensri R, Abdelsamea MM, Mermel CH, Chen PHC, et al. Artificial intelligence in digital breast pathology: Techniques and applications. Breast. 2020; 49: 267–273. https://doi.org/10.1016/j.breast.2019.12.007 PMID: 31935669

33. Siegersma KR, Leiner T, Chew DP, Appelman Y, Hofstra L, Verjans JW. Artificial intelligence in cardiovascular imaging: state of the art and implications for the imaging cardiologist. Netherlands Hear J. 2019; 27: 403–413. https://doi.org/10.1007/s12471-019-01311-1 PMID: 31399886

34. O'Regan DP. Putting machine learning into motion: applications in cardiovascular imaging. Clin Radiol. 2020; 75: 33–37. https://doi.org/10.1016/j.crad.2019.04.008 PMID: 31079952

35. Martin-Isla C, Campello VM, Izquierdo C, Raisi-Estabragh Z, Baeßler B, Petersen SE, et al. Image-Based Cardiac Diagnosis With Machine Learning: A Review. Front Cardiovasc Med. 2020; 7: 1–19. https://doi.org/10.3389/fcvm.2020.00001 PMID: 32039241

36. Gahungu N, Trueick R, Bhat S, Sengupta PP, Dwivedi G. Current Challenges and Recent Updates in Artificial Intelligence and Echocardiography. Curr Cardiovasc Imaging Rep. 2020;13. https://doi.org/10.1007/s12410-020-9529-x

37. Day TG, Kainz B, Hajnal J, Razavi R, Simpson JM. Artificial intelligence, fetal echocardiography, and congenital heart disease. Prenat Diagn. 2021; 41: 733–742. https://doi.org/10.1002/pd.5892

38. Lohmann P, Bousabarah K, Hoevels M, Treuer H. Radiomics in radiation oncology—basics, methods, and limitations. Strahlentherapie und Onkol. 2020; 196: 848–855. https://doi.org/10.1007/s00066-020-01663-3 PMID: 32647917

39. Hu LS, Swanson KR. Roadmap for the clinical integration of radiomics in neuro-oncology. Neuro Oncol. 2020; 22: 743–745. https://doi.org/10.1093/neuonc/noaa078 PMID: 32227184

40. Lim G, Bellemo V, Xie Y, Lee XQ, Yip MYT, Ting DSW. Different fundus imaging modalities and technical factors in AI screening for diabetic retinopathy: a review. Eye Vis. 2020; 7: 1–13. https://doi.org/10.1186/s40662-020-00182-7 PMID: 32313813

41. Asiri N, Hussain M, Al Adel F, Alzaidi N. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. Artif Intell Med. 2019; 99: 101701. https://doi.org/10.1016/j.artmed.2019.07.009 PMID: 31606116

42. Russak AJ, Chaudhry F, De Freitas JK, Baron G, Chaudhry FF, Bienstock S, et al. Machine Learning in Cardiology—Ensuring Clinical Impact Lives Up to the Hype. J Cardiovasc Pharmacol Ther. 2020; 25: 379–390. https://doi.org/10.1177/1074248420928651 PMID: 32495652

43. Bivard A, Churilov L, Parsons M. Artificial intelligence for decision support in acute stroke—current roles and potential. Nat Rev Neurol. 2020; 16: 575–585. https://doi.org/10.1038/s41582-020-0390-y PMID: 32839584

44. Schreuder A, Scholten ET, van Ginneken B, Jacobs C. Artificial intelligence for detection and characterization of pulmonary nodules in lung cancer CT screening: ready for practice? Transl Lung Cancer Res. 2021; 10: 2378–2388. https://doi.org/10.21037/tlcr-2020-lcs-06 PMID: 34164285

45. Liebeskind DS, Scalzo F. Editorial: Machine Learning and Decision Support in Stroke. Front Neurol. 2020; 11: 10–11. https://doi.org/10.3389/fneur.2020.00486 PMID: 32547484

46. de Marvao A, Dawes TJW, O'Regan DP. Artificial Intelligence for Cardiac Imaging-Genetics Research. Front Cardiovasc Med. 2020; 6: 1–10. https://doi.org/10.3389/fcvm.2019.00195 PMID: 32039240

47. Sakamoto T, Furukawa T, Lami K, Pham HHN, Uegami W, Kuroda K, et al. A narrative review of digital pathology and artificial intelligence: focusing on lung cancer. Transl lung cancer Res. 2020; 9: 2255–2276. https://doi.org/10.21037/tlcr-20-591 PMID: 33209648

48. Thomas J, Ledger GA, Mamillapalli CK. Use of artificial intelligence and machine learning for estimating malignancy risk of thyroid nodules. Curr Opin Endocrinol Diabetes Obes. 2020; 27: 345–350. https://doi.org/10.1097/MED.0000000000000557 PMID: 32740044

49. Fenech ME, Buston O. AI in Cardiac Imaging: A UK-Based Perspective on Addressing the Ethical, Social, and Political Challenges. Front Cardiovasc Med. 2020; 7: 1–8. https://doi.org/10.3389/fcvm.2020.00054 PMID: 32351974

50. Petersen SE, Abdulkareem M, Leiner T. Artificial Intelligence Will Transform Cardiac Imaging—Opportunities and Challenges. Front Cardiovasc Med. 2019; 6: 1–6. https://doi.org/10.3389/fcvm.2019.00133 PMID: 31552275

51. Cho B, Bang CS. Artificial Intelligence for the Determination of a Management Strategy for Diminutive Colorectal Polyps : Hype, Hope, or Help. 2020; 115: 70–72.

52. Hosein S, Reitblat CR, Cone EB, Trinh QD. Clinical applications of artificial intelligence in urologic oncology. Curr Opin Urol. 2020; 30: 748–753. https://doi.org/10.1097/MOU.0000000000000819 PMID: 32941255

53. Hickman SE, Baxter GC, Gilbert FJ. Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. Br J Cancer. 2021; 125: 15–22. https://doi.org/10.1038/s41416-021-01333-w PMID: 33772149

54. Shung DL, Byrne MF. How Artificial Intelligence Will Impact Colonoscopy and Colorectal Screening. Gastrointest Endosc Clin N Am. 2020; 30: 585–595. https://doi.org/10.1016/j.giec.2020.02.010 PMID: 32439090

55. Grzybowski A, Brona P, Lim G, Ruamviboonsuk P, Tan GSW, Abramoff M, et al. Artificial intelligence for diabetic retinopathy screening: a review. Eye (Lond). 2020; 34: 604. https://doi.org/10.1038/s41433-019-0728-0 PMID: 31822855

56. Keser N. Role of Artificial Intelligence in Cardiovascular Imaging. Anatol J Cardiol. 2019; 22: 10–12. https://doi.org/10.14744/AnatolJCardiol.2019.93727 PMID: 31670722

57. Sevakula RK, Au-Yeung WTM, Singh JP, Heist EK, Isselbacher EM, Armoundas AA. State-of-the-Art Machine Learning Techniques Aiming to Improve Patient Outcomes Pertaining to the Cardiovascular System. J Am Heart Assoc. 2020; 9: 1–15. https://doi.org/10.1161/JAHA.119.013924 PMID: 32067584

58. Chiwome L, Okojie OM, Rahman AKMJ, Javed F, Hamid P. Artificial Intelligence: Is It Armageddon for Breast Radiologists? Cureus. 2020; 12: 1–10. https://doi.org/10.7759/cureus.8923 PMID: 32760624

59. French J, Chen L. Preparing for Artificial Intelligence: Systems-Level Implications for the Medical Imaging and Radiation Therapy Professions. J Med Imaging Radiat Sci. 2019; 50: S20–S23. https://doi.org/10.1016/j.jmir.2019.09.002 PMID: 31648963

60. van Assen M, Cornelissen LJ. Artificial Intelligence: From Scientific Curiosity to Clinical Precocity? JACC Cardiovasc Imaging. 2020; 13: 1172–1174. https://doi.org/10.1016/j.jcmg.2019.09.008 PMID: 31734210

61. Retson TA, Eghtedari M. Computer-Aided Detection/Diagnosis in Breast Imaging: A Focus on the Evolving FDA Regulations for Using Software as a Medical Device. Curr Radiol Rep. 2020; 8: 1–7. https://doi.org/10.1007/s40134-020-00350-6

62. Dorado-Díaz PI, Sampedro-Gómez J, Vicente-Palacios V, Sánchez PL. Applications of Artificial Intelligence in Cardiology. The Future is Already Here. Rev Española Cardiol (English Ed. 2019; 72: 1065–1075. https://doi.org/10.1016/j.rec.2019.05.014 PMID: 31611150

63. Haq IU, Haq I, Xu B. Artificial intelligence in personalized cardiovascular medicine and cardiovascular imaging. Cardiovasc Diagn Ther. 2021; 11: 911–923. https://doi.org/10.21037/cdt.2020.03.09 PMID: 34295713

64. Ou WC, Polat D, Dogan BE. Deep learning in breast radiology: current progress and future directions. Eur Radiol. 2021; 31: 4872–4885. https://doi.org/10.1007/s00330-020-07640-9 PMID: 33449174

65. Xue P, Ng MTA, Qiao Y. The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence. BMC Med. 2020; 18: 1–7. https://doi.org/10.1186/s12916-020-01613-x PMID: 32493320

66. Daisy PS, Anitha TS. Can artificial intelligence overtake human intelligence on the bumpy road towards glioma therapy? Med Oncol. 2021;38. https://doi.org/10.1007/s12032-021-01500-2 PMID: 33811540

67. Francolini G, Desideri I, Stocchi G, Salvestrini V, Ciccone LP, Garlatti P, et al. Artificial Intelligence in radiotherapy: state of the art and future directions. Med Oncol. 2020; 37: 1–9. https://doi.org/10.1007/s12032-020-01374-w PMID: 32323066

68. Davis A, Billick K, Horton K, Jankowski M, Knoll P, Marshall JE, et al. Artificial Intelligence and Echo-cardiography: A Primer for Cardiac Sonographers. J Am Soc Echocardiogr. 2020; 33: 1061–1066. https://doi.org/10.1016/j.echo.2020.04.025 PMID: 32536431

69. Romiti S, Vinciguerra M, Saade W, Anso Cortajarena I, Greco E. Artificial Intelligence (AI) and Cardio-vascular Diseases: An Unexpected Alliance. Cardiol Res Pract. 2020;2020. https://doi.org/10.1155/2020/4972346 PMID: 32676206

70. Davies R, Babu-Narayan S V. Deep learning in congenital heart disease imaging: Hope but not haste. Heart. 2020; 106: 960–961. https://doi.org/10.1136/heartjnl-2019-316496 PMID: 32341135

71. Tandon YK, Bartholmai BJ, Koo CW. Putting artificial intelligence (AI) on the spot: Machine learning evaluation of pulmonary nodules. J Thorac Dis. 2020; 12: 6954–6965. https://doi.org/10.21037/jtd-2019-cptn-03 PMID: 33282401

72. Patil S, Moafa IH, Mosa Alfaifi M, Abdu AM, Jafer MA, Raju K L, et al. Reviewing the Role of Artificial Intelligence in Cancer. Asian Pacific J Cancer Biol. 2020; 5: 189–199. https://doi.org/10.31557/apjcb.2020.5.4.189–199

73. Sollini M, Bartoli F, Marciano A, Zanca R, Slart RHJA, Erba PA. Artificial intelligence and hybrid imaging: the best match for personalized medicine in oncology. Eur J Hybrid Imaging. 2020;4. https://doi.org/10.1186/s41824-020-00094-8 PMID: 34191197

74. Hu W, Yang H, Xu H, Mao Y. Radiomics based on artificial intelligence in liver diseases: Where we are? Gastroenterol Rep. 2020; 8: 90–97. https://doi.org/10.1093/gastro/goaa011 PMID: 32280468

75. Tran WT, Jerzak K, Lu FI, Klein J, Tabbarah S, Lagree A, et al. Personalized Breast Cancer Treatments Using Artificial Intelligence in Radiomics and Pathomics. J Med Imaging Radiat Sci. 2019; 50: S32–S41. https://doi.org/10.1016/j.jmir.2019.07.010 PMID: 31447230

76. Sotoudeh H, Shafaat O, Bernstock JD, Brooks MD, Elsayed GA, Chen JA, et al. Artificial intelligence in the management of glioma: Era of personalized medicine. Front Oncol. 2019; 9: 1–11. https://doi.org/10.3389/fonc.2019.00768 PMID: 31475111

77. Arribas Anta J, Dinis-Ribeiro M. Early gastric cancer and Artificial Intelligence: Is it time for population screening? Best Pract Res Clin Gastroenterol. 2021; 52–53: 101710. https://doi.org/10.1016/j.bpg.2020.101710 PMID: 34172244

78. Schoepf UJ, Tesche C. Oracle of Our Time: Machine learning for predicting cardiovascular events. Radiology. 2019; 292: 363–364. https://doi.org/10.1148/radiol.2019191165 PMID: 31237811

79. Sermesant M, Delingette H, Cochet H, Jaïs P, Ayache N. Applications of artificial intelligence in cardio-vascular imaging. Nat Rev Cardiol. 2021; 18: 600–609. https://doi.org/10.1038/s41569-021-00527-2 PMID: 33712806

80. Radakovich N, Nagy M, Nazha A. Machine learning in haematological malignancies. Lancet Haematol. 2020; 7: e541–e550. https://doi.org/10.1016/S2352-3026(20)30121-6 PMID: 32589980

81. Martín Noguerol T, Paulano-Godino F, Martín-Valdivia MT, Menias CO, Luna A. Strengths, Weaknesses, Opportunities, and Threats Analysis of Artificial Intelligence and Machine Learning Applications in Radiology. J Am Coll Radiol. 2019; 16: 1239–1247. https://doi.org/10.1016/j.jacr.2019.05.047 PMID: 31492401

82. Adamson A, et al. Machine Learning and the Cancer-Diagnosis Problem—No Gold Standard. 2019; 2285–2287.

83. P. S, eta al. Artificial Intelligence in Diagnostic Imaging Status Quo, Challenges, and Future Opportunities. Ugeskr Laeger. 2020; 182: 11–16. https://doi.org/10.1097/RTI.0000000000000499 PMID: 32205816

84. Huynh E, Hosny A, Guthier C, Bitterman DS, Petit SF, Haas-Kogan DA, et al. Artificial intelligence in radiation oncology. Nat Rev Clin Oncol. 2020; 17: 771–781. https://doi.org/10.1038/s41571-020-0417-8 PMID: 32843739

85. Shimizu H, Nakayama KI. Artificial intelligence in oncology. Cancer Sci. 2020; 111: 1452–1460. https://doi.org/10.1111/cas.14377 PMID: 32133724

86. Chang A. Artificial intelligence in pediatric cardiology and cardiac surgery: Irrational hype or paradigm shift? Ann Pediatr Cardiol. 2019; 12: 191–194. https://doi.org/10.4103/apc.APC_55_19 PMID: 31516273

87. Hwang EJ, Park CM. Clinical implementation of deep learning in thoracic radiology: Potential applications and challenges. Korean J Radiol. 2020; 21: 511–525. https://doi.org/10.3348/kjr.2019.0821 PMID: 32323497

88. Theek B, Magnuska Z, Gremse F, Hahn H, Schulz V, Kiessling F. Automation of data analysis in molecular cancer imaging and its potential impact on future clinical practice. Methods. 2021; 188: 30–36. https://doi.org/10.1016/j.ymeth.2020.06.019 PMID: 32615232

89. Lovejoy CA, Phillips E, Maruthappu M. Application of artificial intelligence in respiratory medicine: Has the time arrived? Respirology. 2019; 24: 1136–1137. https://doi.org/10.1111/resp.13676 PMID: 31407427

90. Barone-Rochette G. Will artificial intelligence change the job of the cardiac imaging specialist? Arch Cardiovasc Dis. 2020; 113: 1–4. https://doi.org/10.1016/j.acvd.2019.11.002 PMID: 31899110

91. Ladefoged CN, Andersen FL, Højgaard L. Artificial intelligence for diagnostic imaging. Ugeskr Laeger. 2020; 182: 11–16. https://doi.org/10.1097/rti.0000000000000499 PMID: 32285782

92. Jarrett D, Stride E, Vallis K, Gooding MJ. Applications and limitations of machine learning in radiation oncology. Br J Radiol. 2019; 92: 1–12. https://doi.org/10.1259/bjr.20190001 PMID: 31112393

93. Martini ML, Oermann EK. Intraoperative brain tumour identification with deep learning. Nat Rev Clin Oncol. 2020; 17: 200–201. https://doi.org/10.1038/s41571-020-0343-9 PMID: 32099093

94. Felmingham CM, Adler NR, Ge Z, Morton RL, Janda M, Mar VJ. The Importance of Incorporating Human Factors in the Design and Implementation of Artificial Intelligence for Skin Cancer Diagnosis in the Real World. Am J Clin Dermatol. 2021; 22: 233–242. https://doi.org/10.1007/s40257-020-00574-4 PMID: 33354741

95. Wang X, Bin Li B. Deep Learning in Head and Neck Tumor Multiomics Diagnosis and Analysis: Review of the Literature. Front Genet. 2021;12. https://doi.org/10.3389/fgene.2021.624820 PMID: 33643386

96. Seetharam K, Brito D, Farjo PD, Sengupta PP. The Role of Artificial Intelligence in Cardiovascular Imaging: State of the Art Review. Front Cardiovasc Med. 2020;7. https://doi.org/10.3389/FCVM.2020.618849 PMID: 33426010

97. Vogrin M, Trojner T, Kelc R. Artificial intelligence in musculoskeletal oncological radiology. Radiol Oncol. 2020; 55: 1–6. https://doi.org/10.2478/raon-2020-0068 PMID: 33885240

98. Gil D, Rosell A. Advances in Artificial Intelligence—How Lung Cancer CT Screening Will Progress? J Thorac Oncol. 2019; 14: S33. https://doi.org/10.1016/j.jtho.2019.08.103

99. Keshava SN, Kalva S. Artificial Intelligence in Interventional Radiology. J Clin Interv Radiol ISVIR. 2019; 03: 071–071. https://doi.org/10.1055/s-0039-1694309

100. Seetharam K, Raina S, Sengupta PP. The Role of Artificial Intelligence in Echocardiography. Curr Cardiol Rep. 2020; 22: 1–8. https://doi.org/10.1007/s11886-020-01329-7 PMID: 32728829

101. Cazacu I, Udristoiu A, Gruionu L, Iacob A, Gruionu G, Saftoiu A. Artificial intelligence in pancreatic cancer: Toward precision diagnosis. Endosc Ultrasound. 2019; 8: 357–359. https://doi.org/10.4103/eus.eus_76_19 PMID: 31854344

102. Saxena S, Gyanchandani M. Machine Learning Methods for Computer-Aided Breast Cancer Diagnosis Using Histopathology: A Narrative Review. J Med Imaging Radiat Sci. 2020; 51: 182–193. https://doi.org/10.1016/j.jmir.2019.11.001 PMID: 31884065

103. Yu PK. The Algorithmic Divide and Equality in the Age of Artificial Intelligence. Fla Law Rev. 2020; 72: 19–44.

104. Toreini E, Aitken M, Coopamootoo K, Elliott K, Zelaya CG, van Moorsel A. The relationship between trust in AI and trustworthy machine learning technologies. FAT* 2020—Proc 2020 Conf Fairness, Accountability, Transpar. 2020; 272–283. https://doi.org/10.1145/3351095.3372834

105. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Informatics Assoc. 2020; 27: 491–497. https://doi.org/10.1093/jamia/ocz192 PMID: 31682262

106. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Heal. 2021; 3: e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9 PMID: 34711379

107. Craglia M (Ed., Annoni A, Benczur P, Bertoldi P, Delipetrev P, De Prato G, et al. Artificial Intelligence—A European perspective. 2018. https://doi.org/10.2760/11251

108. Holland CP. BGL Group: Artificial Intelligence (AI) Strategy. SSRN Electron J. 2021. https://doi.org/10.2139/ssrn.3841656

109. High-level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. Eur Comm. 2019. https://doi.org/10.1017/9781108936040.022

110. Kerasidou C (Xaroula), Kerasidou A, Buscher M, Wilkinson S. Before and beyond trust: reliance in medical AI. J Med Ethics. 2021; medethics-2020-107095. https://doi.org/10.1136/medethics-2020-107095 PMID: 34426519

111. EU. EU general data protection regulation. Eur Comm. 2018; 2014: 45–62. https://doi.org/10.1308/rcsfdj.2018.54

112. Ferretti A, Schneider M, Blasimme A. Machine Learning in Medicine: Opening the New Data Protection Black Box. Eur Data Prot Law Rev. 2018; 4: 320–332. https://doi.org/10.21552/edpl/2018/3/10

113. European Commission. LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. Eur Comm. 2021; 0106: 1–108.

114. Amann J, Vetter D, Blomberg SN, Christensen HC, Coffee M, Gerke S, et al. To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems. PLOS Digit Heal. 2022; 1: e0000016. https://doi.org/10.1371/journal.pdig.0000016

115. Babic BB, Gerke S, Evgeniou T, Glenn Cohen I. Beware explanations from AI in health care the benefits of explainable artificial intelligence are not what they appear. Science (80-). 2021; 373: 284–286. https://doi.org/10.1126/science.abg1834 PMID: 34437144

116. Shah NH, Milstein A, Bagley SC. Making Machine Learning Models Clinically Useful. N Engl J Med. 2018; 378: 981–983. https://doi.org/10.1056/nejmp1714229

117. Janssens ACJW Martens FK. Reflection on modern methods: Revisiting the area under the ROC Curve. Int J Epidemiol. 2020; 49: 1397–1403. https://doi.org/10.1093/ije/dyz274 PMID: 31967640

118. Bossuyt PMM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. Med Decis Mak. 2009; 29: 30–38. https://doi.org/10.1177/0272989X09347013 PMID: 19726782

119. Coeckelbergh M. Ethics of artificial intelligence: Some ethical issues and regulatory challenges. Technol Regul. 2019; 31–34.

120. WHO. ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH.

121. Vayena E, Blasimme A. Towards Adaptive Governance in Big Data Health Research. The Cambridge Handbook of Health Research Regulation. Cambridge University Press; 2021.

122. Vayena E, Blasimme A. Health research with big data: Time for systemic oversight. J Law, Med Ethics. 2018; 46: 119–129. https://doi.org/10.1177/1073110518766026 PMID: 30034208

123. United States Food & Drug Administration. Proposed Regulatory Framework for Modifications to Artificial Intelligence / Machine Learning (AI / ML) -Based Software as a Medical Device (SaMD) —Discussion Paper and Request for Feedback. US Food Drug Adm. 2019; 1–20.

124. Ahmad MA, Overman S, Allen C, Kumar V, Teredesai A, Eckert C. Artificial Intelligence/Machine Learning (AI/ML)-Based- Software as a Medical Device (SaMD) Action Plan. 2021; 4023–4024. https://doi.org/10.1145/3447548.3470823