Thesis Project Portfolio

Participatory Design: Data Model Management Application

(Technical Report)

Participatory Design as a Solution to the Ethical and Legal Implications of Algorithmic Diagnosis in Medical Imaging

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Gary Tillman Dean

Spring, 2023

Department of Computer Science

Introduction

Research at Stanford using an AI model to diagnose two types of lung cancer found that the model differentiated between the two types and predicted survival rate more effectively than the human doctors (Conger, 2016). One of the head researchers, Michael Snyder, said "Two highly skilled pathologists assessing the same slide will agree only about 60 percent of the time. This approach replaces this subjectivity with sophisticated, quantitative measurements that we feel are likely to improve patient outcomes" (Conger, 2016). To many, increased accuracy is an indisputable good, and we have seen the results of this sentiment as AI is now being implemented in medical imaging across the country. But using computer algorithms that reach decisions through indecipherable weightings within neural networks has also led to many ethical dilemmas.

Medical imaging is a useful technology that has now become nearly universal. To allow doctors to see internal problems such as bone breaks, tumors, or bleeding, different types of images can be taken using electromagnetic waves or sound waves to bypass the outer layers of skin or soft tissue. Historically, a human radiologist examines a patient's imaging and makes a diagnosis based on features of the image that they have been trained to identify. However, the dawn of image recognition algorithms has led to a new way of analyzing medical imaging.

Image recognition algorithms built with machine learning have become common across many different fields including ecommerce, marketing, and autonomous vehicles. They have become remarkably accurate, often beating human professionals and dramatically improving efficiency. Medical imaging has not remained untouched from these algorithms as many healthcare professionals now use software to diagnose patients based on their MRI, X-ray, or CT scan. However, there are several problems with their use. First, the increased accuracy is not equitably distributed across races and genders. Second, humans often instinctively distrust diagnostic algorithms. Thirdly, the algorithms could create legal and ethical problems with things like data privacy and misdiagnosis.

In this paper, I argue that by applying a participatory design approach to the development of medical imaging analysis algorithms, doctors and patients will have more confidence and trust in them and that some of the issues with bias and accuracy can be solved, but that ultimately some aspects of the algorithms will remain uncontrollable. In my literature review section, I present the current application of machine learning in medical imaging, the benefits gained from this application, and the ethical concerns associated with it. I gather my data from medical studies, academic research, and whitepapers from medical professionals. In my analysis, I find that participatory design can remedy some of the ethical concerns but cannot ultimately address the issue entirely due to the nature of the algorithms. I conclude that more frameworks and solutions should be researched, but that principles of participatory design can be incredibly valuable when applied to this problem.

Literature Review

Machine learning (ML) algorithms are becoming increasingly used in medical imaging diagnosis, and this growth is mostly being driven by their accuracy. Medical imaging as a field covers a variety of different types of imaging and scans, including x-rays, MRI, CT scans, and ultrasound. Some machine learning algorithms have been proven to be more accurate in analyzing medical imaging, such as MRI, than human doctors (Nishida et al., 2022). For example, a study was done in which three machine learning algorithms were trained using ultrasound images of livers with and without tumors, each with a different number of images in the training set: AI model-1 with the least images, AI model-2 with the second-least, and AI model-3 with the most.

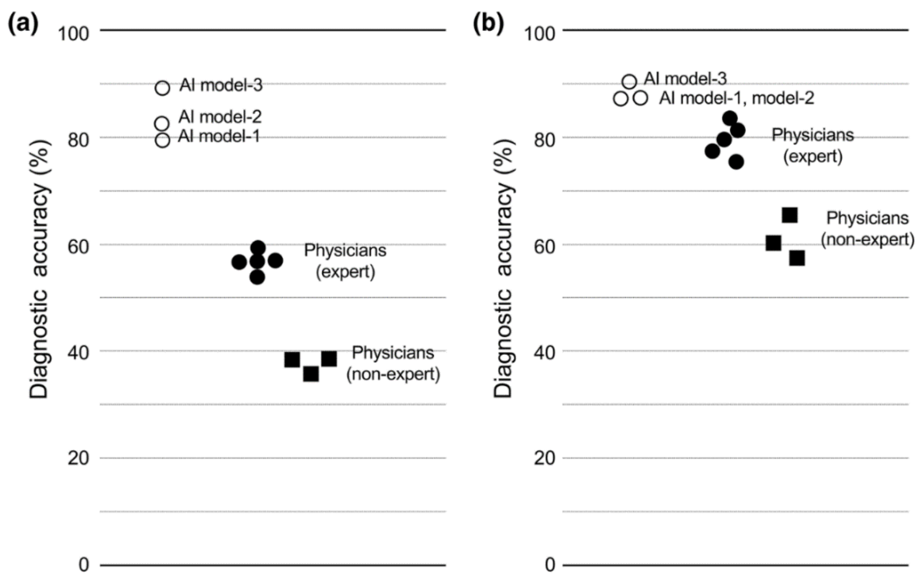

Figure 1: (a) Comparison of accuracies on four-class discrimination (b) comparison of accuracies for the malignant tumor (Nishida et al., 2022, p. 315)

As can be seen in Figure 1, all three algorithms outperformed human analysis, both expert and non-expert physicians, in diagnosing four-class discrimination and malignant liver tumors: “The accuracies of AI models are significantly higher than those of human experts (p = 0.0325 by Wilcoxon rank sum test)” (Nishida et al., 2022, p. 316).

This accuracy extends across other conditions and imaging types as brain MRIs after a stroke to determine likelihood of developing symptomatic intracranial hemorrhage: “Advances in ML and DL have allowed for the development of more accurate models which outperform the traditional SEDAN and HAT scores” (Mainali et al., 2021, p. 4). Yet another example is in breast ultrasounds to identify cancer where scientists found that an algorithm outperformed radiologists in multiple categories: “when considering sensitivity, specificity, and accuracy, S-Detect™ for Breast produced a favorable performance compared to the radiologists, as an empirical result in the study” (O’Connell et al., 2022, p. 103). This increase in accuracy is impossible to ignore when the ultimate goal of medicine is providing the best possible care for patients.

As Bhardwaj (2017) notes, the implementation of machine learning in healthcare is also part of a larger change needed to address the rapid growth of the industry. Bhardwaj (2017) specifically highlights a startup, Enlitic, that is using machine learning "to turn medical data such as lab x-rays and images, patient histories, and physician notes into meaningful insights and patterns" (p. 238). Enlitic recently found that their software was 50% more accurate than a panel of expert radiologists at diagnosing lung cancer nodules (Bhardwaj et al., 2017). Dramatically increased accuracy and efficiency would be welcome in the US, a country where "healthcare costs are rising at rates close to double of our economic growth rate" (Bhardwaj et al., 2017, p. 236). When considered this way, machine learning seems less like an option to potentially increase accuracy, and more like a financial inevitability to ensure quality care for the growing number of patients. However, this increase in accuracy and efficiency does come at a cost.

One concern is that accuracy in image analysis machine learning algorithms has been shown to inequitable across different groups. Misdiagnosis can result in loss of life or other severe health consequences so providing quality care to all should be paramount. In many image recognition algorithms, accuracy has been shown to differ based on race or gender, creating inequity and bias in healthcare (Chen et al., 2021). Many groups such as low-income, transgender and non-binary, and immigrant populations are not well documented in medical datasets. This lack of representation leads to massively imbalanced datasets that skew towards all the majority demographics. Imbalanced datasets are common in medical imaging as many of the conditions that need to be identified are extremely rare, such as brain tumors which occur in only 1% of the population (Zhang et al., 2019).

Another big concern is that training algorithms can result in privacy concerns. As Jaremko (2019) notes in a whitepaper written by a group of Canadian radiologists: "Historically, a patient's medical data was consulted only occasionally and solely in care of that patient, at initial consultation and follow-up, then archived and often deleted over time due to the cost of storage" (p. 108). As AI is introduced into healthcare, the data is also used in another way – to create massive datasets that will train algorithms to diagnose other patients (Jaremko et al., 2019). Patient's data must be used in a whole different way for these algorithms to be effective, and as the doctors note, "Personal information in radiology comprises mainly images, analogous to photos taken by various 'cameras' (x-ray, ultrasound, computed tomography [CT], etc). This data is highly personal and sensitive" (Jaremko, et. al, 2019, p. 109). Massive amounts of data being gathered, stored, and accessed increases the risk of a leak.

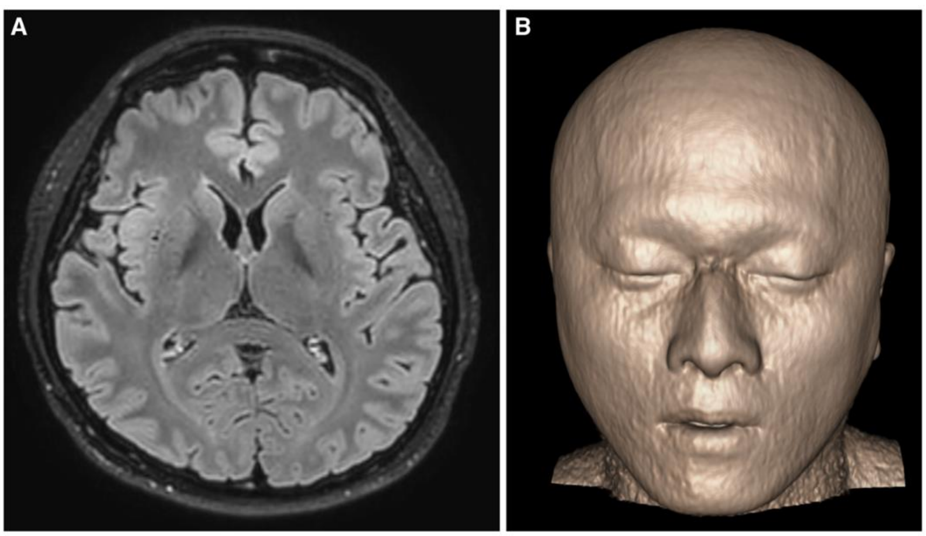

Figure 2: Digital Recreation of Patient's Face from Data (Jaremko et al., 2019, p. 113)

Figure 2 shows how extremely personal information allowing identification of patients can be found in seemingly anonymous data groups. As digital recreation technology continues to advance, data privacy breaches like this will become even more dangerous.

ML algorithms often provoke distrust from doctors and patients because they work very differently than a human analyzing an image. As Karmakar (2021) notes, the decisions made by ML algorithms are fundamentally at odds with medicine and science, as they are made in milliseconds by indecipherable neural networks. In a study done to analyze the patient side of this distrust, results showed that patients trusted AI diagnosis less than human doctors, even when they are shown that the AI is more accurate (Juravle et al. 2020). On the doctor side, the issue is even more complicated. What does a doctor do when their diagnosis is at odds with the result from the algorithm? Algorithms rarely provide reasoning behind their feedback, as the images are simply kicked through thousands of layers of the neural network which then arrives at a conclusion. A situation like this puts the doctor in an impossible dilemma: if they defer to the algorithm they are going against their own professional opinion, but if they go with their own decision and are wrong, they are ignoring a supposedly very accurate algorithm (Grote et al. 2020). Machine learning obscures situations in the field of medicine that were historically clear.

Arguably the largest concern of them all is the ethical liability of errors by the AI. Dai (2022) highlights the potential shift in liability: “Traditionally, adhering to the standard of care shields the physician from liability even when adverse patient outcomes occur. When the physician uses AI, the situation changes” (p. 4445). In their whitepaper, the Association of Canadian Radiologists emphasizes this risk:

> Worse, AI may be harmful. The US Food and Drug Administration states that diagnostic medical devices can be harmful in 5 ways [53]: increasing false-positive results (leading to unnecessary additional procedures), increasing false-negative results (failing to diagnose disease), being applied to inappropriate populations; being misused by human users; and malfunctioning by providing incorrect output. (Jaremko, et. al, 2019, p. 115).

These five ways are meant to apply to all diagnostic medical devices but all of them apply to these algorithms. Failure to diagnose could result in the disease becoming much worse. False-positive results could lead to unnecessary, dangerous, and costly procedures. A simple computer malfunction could lead to a mistake. When mistakes are made responsibility must be assigned.

Yet AI clouds what would be an easy case of liability, depending on the level of autonomy given to the AI:

> Indeed, liability with these level 1 systems continues to rest with the user (radiologist or other clinician). However, as higher levels of autonomy are intended, liability will begin to shift to the AI system and hence to the manufacturer and regulatory body. If the device is intended to autonomously diagnose a certain disease and the system is used as intended by the physician or institution, how could they be accused of malpractice? (Jaremko, et. al, 2019, p. 115).

The Canadian radiologists observe that when the AI replaces the expert human doctor's role and judgment entirely, it should be argued that the liability for mistakes lies on the AI and its developer. The field of medicine, due to its unique ethical dilemmas, can turn a simple image recognition algorithm into a liable agent for drastically altering a human's life. Efforts to mitigate these trust and accuracy issues are often an uphill battle due to the black-box nature of the ML algorithms. ML algorithms typically only provide the result of a classification (either diseased or not) without providing any rationale as to how they made that classification (Komura et al. 2019).

The framework that I attempt to address these issues with is participatory design (PD). PD began in Scandinavia but spread to the U.S. in the 1980s (Gregory, 2003). PD was first used as a strategy "to address workplace transformations brought about by computers" (Donia et al., 2021, p. 2). As a framework, it centers around the idea that "the people destined to use the system play a critical role in designing it" (Gregory, 2003, p. 62). It argues that by allowing all of the future stakeholders of a product, system, or research to be involved in the design decision process more usable, useful, and creative solutions will be proposed and executed. But in participatory design, the solution is not the only potential benefit. As Gregory (2003) notes, other benefits include "improving the knowledge upon which systems are built; enabling people to develop realistic expectations and reducing resistance to change; and increasing workplace democracy by giving the members of an organization the right to participate in decisions that are likely to affect their work" (p. 63). Allowing each user to be involved increases trust and knowledge in the system, allowing them to feel more empowered and encouraged to use it. In the analysis section of this paper, I explore how this powerful framework can be applied to medical imaging analysis algorithms.

Methods

I gathered secondary sources, mostly research on the issues caused by ML algorithms, such as accuracy bias and trust concerns, and on the potential participatory design's application to ML algorithms. I gathered research primarily from medical professionals, STS academics, and computer scientists. In my review of this literature, I looked to see who the stakeholders are in medical imaging analysis and how they could be effectively involved in the design process of the ML algorithms. I explored several different academic opinions on how participatory design can be applied to this field – both for and against it.

Analysis

Participatory design in the most traditional sense is difficult to apply across the machine learning process. The algorithms often train entirely on their own, assigning weights to certain nodes without human interference. "Normally in PD, the designer can to some extent predict effects of the design choices. This is different with AI, where the rationale for the system's operation is hidden both in difficult-to-understand formalisms and difficult-to-know training sets – often too difficult for a nonexpert to see through" (Bratteteig et al., 2018, p. 4).

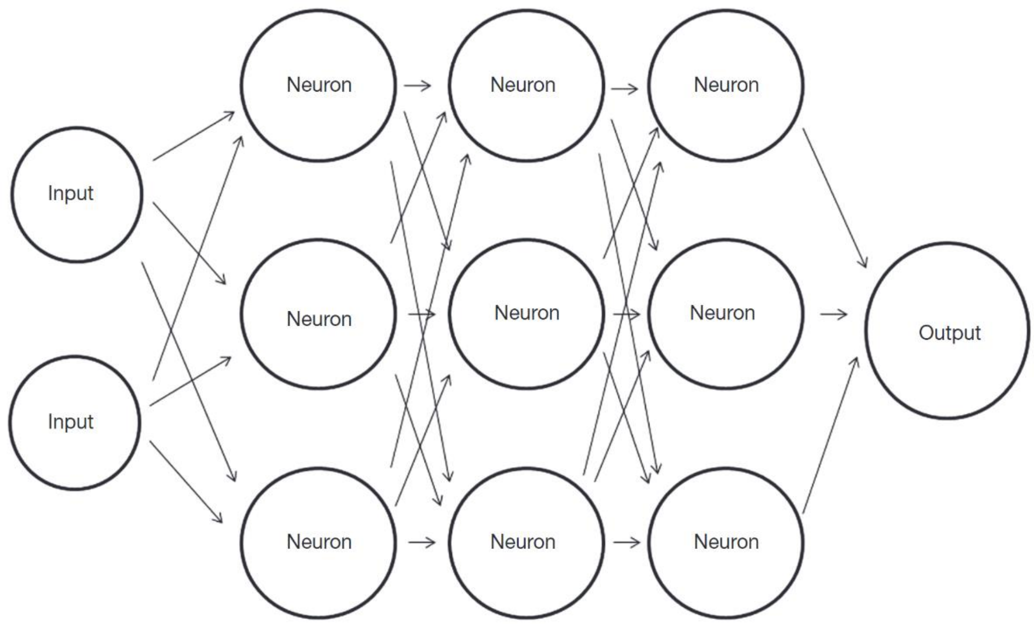

Figure 3: Simple Artificial Neural Network (Klang, 2018, p. 1327)

In Figure 3, a very simple example of a neural network with only three hidden layers is presented to be used in supervised learning. During the training phase, thousands of images are sent through the input and through each layer one by one until the results are combined to give an output. The network is designed to work like a human brain with each neuron having a unique weight and “firing” based on the input received. While training, the weights of each neuron shift based on whether the output is correct, and after receiving thousands of inputs, the accuracy of the network greatly increases. At this point the training phase is complete, and the algorithm is ready to take real input. Over the course of the training process, the developers have no control over how the network changes or assigns weights to the neurons.

As a result, some argue that PD is made obsolete by AI due to the lack of control provided to the stakeholders in the training of the algorithm: “Even if one understands what AI does in principle, it is not possible to foresee how it changes over time and hence how the activity in which the system will be part of will change (maybe even fundamentally). This is problematic if PD should maintain its aim to enhance users' control over their activities" (Bratteteig et al. 2018, p. 3). Sloane (2020) notes how ML systems are often biased and the trend towards applying more "participatory ML systems" but cautions against seeing it as a fix for machine learning. Although ML seems like the perfect candidate for the framework due to the complexity and trust issues causes, applying it gets complicated quickly (Sloane et al., 2020).

However, PD can still provide valuable help in other parts of the design process like bigger picture decision-making and data collection for the algorithms themselves. The first way in which participation can be increased to decrease bias against minority groups is to increase their presence in the data used to train algorithms. As Sloane (2020) observes, our data is being constantly collected to improve algorithms' performance: "Billions of ordinary web users also continually participate in the production and refinement of ML, as their online (and offline) activities produce neatly labeled rows of data on how they click their way around the web, navigate their streets, and engage in any number of other… activities" (p. 2). Although this aspect of participation may seem too obvious, it is arguably the most important part of applying participatory design. One of largest contributors to differences in accuracy is imbalanced training data. Chen (2021) observes: "Recent work on acute kidney injury achieved state-of-the-art prediction performance in a large dataset of 703,782 adult patients… however, the authors noted that model performance was lower in female patients since they make up 6.38% of patients in the training data" (p. 6). Chen (2021) also notes that this bias cannot be avoided by simply using a balanced initial smaller set of data and then training further using the larger imbalanced data set. De Bruijne (2016) points out the dangers of an imbalanced dataset:

> However, depending on the training data, diagnosis decisions could well be driven not by signs of disease, but by signs of a confounding factor that is correlated with disease status in the training set. For instance, if a disease has higher prevalence in men than in women, a complex learning algorithm might decide that the size of certain structures is a good

> indicator for the risk of disease, while in a study covering a large age range, signs of normal aging might be highlighted as strongly suspicious of dementia (p. 3).

One fix is collecting a training set that is balanced for these confounding factors such as gender or age, but a better fix is incorporating possible other predictors in the learning to try to learn the relationship between the confounders and images (de Bruijne et al., 2016). The composition of the training data completely almost completely molds the algorithm, and when given an imbalanced dataset, algorithms designed to identify patterns will identify incorrect signs of disease that skew results towards certain age groups, genders, or races. The best potential solutions seem to be either collecting "balanced comprehensive data" or "creating specialty learning algorithms" (Chen et al., 2021, p. 6). Collecting balanced data can't be a small endeavor or afterthought to be effective – it must be a concerted, total effort.

The idea that there are no technical changes that can be made to machine learning algorithms is a myth. While they are mostly a product of their dataset, there are still design choices in the mathematics behind how they interact with that dataset. Chen (2021) notes that one area that could result in biases such as these is when choosing a loss function for the algorithm: "Recent work has shown that models trained with a surrogate loss may exhibit errors that disproportionately affect undersampled groups in the training data" (p. 12). Seemingly simple, purely technical decisions like this can have a large impact on the level of bias present in the algorithms after they complete their training (Chen et al., 2021). To recognize the importance of design choices like the loss function, it is important to involve those represented by undersampled data.

Another important foundational step to ensure that potential moral and ethical challenges are considered early on is to increase representation of minority groups on the development teams for these algorithms. De Boer (2021) notes the importance of minority representation:

"Ensuring that development teams for ML diagnostic systems are inclusive and diverse—in terms of gender, age, race, culture, socioeconomic background, and so forth—should enable the qualitative moral impacts to be foregrounded early on in the design process" (p. 261). Chen (2021) notes what a lack of diversity in development can lead to by observing some differences in research done and funded by minority groups: "Research shows that scientists from underrepresented racial and gender groups tend to prioritize different research topics. They produce more novel research, but their innovations are taken up at lower rates" (p. 6). Not having these groups represented in the design process simply makes it easier to overlook certain problems. As de Boer (2021) observes, participatory design "equips participants with a repertoire for expressing how they are affected by the introduction of ML and for engaging in a dialogue about whether these effects are desirable" (p. 261). Hiring minority developers empowers them by giving them a platform to voice how the algorithms affect their communities.

In the same way increasing minority representation in the development process can help address issues specific to those groups, involving doctors in the design process of the algorithms and educating them on how the algorithms have been trained and function will allow them to both trust them more and be aware of their limitations. In their whitepaper, the Canadian Radiologists, recommend two things to prepare for potential liability issues:

> 1. CAR to work together with other stakeholders such as provincial Ministries of Health and the Canadian Medical Protection Association to develop guidelines for appropriate deployment of AI assistive tools in hospital departments and radiology groups, seeking to minimize potential harm and institutional liability for malpractice in case of medical error involving AI.
> 2. Radiologists using AI should be aware of its limitations, use AI appropriately within algorithms of care, and not allow AI to replace human expert judgment (Jaremko, et. al, 2019, p. 116).

The importance of being aware of the limitations of AI is the most important part of these recommendations, and the best way to ensure that is by getting radiologists involved in the development process for the algorithms. Obviously, it will be impossible to let every radiologist in the room, but having representation from their expertise will allow several key things. First, it will allow them to voice concerns or suggestions about how to make a product to better suit their needs. Second, it will allow them to take knowledge of how the algorithm works and educate other radiologists by speaking at conferences or writing academic literature on how to safely use and implement the software. As Leimanis (2021) points out "Medical practitioners will have to be able to explain the treatment decisions that will be based on AI recommendations. It will in turn require significantly higher level of understanding by the medical staff of the underlying technology, including AI applications, both from technological, medical and regulatory aspects" (p. 99). The education on the algorithms does not mean that that doctors must become machine learning experts themselves: "Although doctors might not need to know the detailed mathematical calculations used in a machine learning algorithm, they could be educated about the types of data used in making the predictions and the relative weights assigned to each type of data (Ngiam et al. 2019, p. 271)." Knowledge of the algorithms will not only be helpful but almost a requirement to practice medicine and participatory design will ensure it is achieved.

Participatory design cannot solve every issue related to ML medical imaging analysis algorithms, but by applying it to the data collection and training process, increasing minority representation on the development teams, and involving radiologists in the development process, many of the issues can be either fixed or at least partially remedied.

Conclusion

Many parts of our lives are impacted by these black-box algorithms, but when considering the massive consequences of medical imaging diagnosis, we must actively try to better understand and manage them. As I have argued in this paper, participatory design can be a useful framework to apply to medical imaging analysis algorithms, paving the way for a more equitable distribution of accuracy, increase in trust from doctors and patients, and security of patient's data. My hope is that from this paper, healthcare professionals, patients, and programmers can gain a better understanding of how machine learning is being used in medical imaging analysis and the associated concerns. Hopefully, they will be able to also apply some of the insights offered by participatory design to data collection and algorithm design in spite of the difficulties posed. They will also realize that participatory design as a framework cannot be a catch all solution to the issue, and that further research using other frameworks could lead to even more solutions to the problems presented in this paper. Even with some of the risks and ethical issues, ML offers great hope for doctors to be able to identify diseases much quicker than a human could, resulting in potentially many saved lives.

References

Bhardwaj, R., Nambiar, A. R., & Dutta, D. (2017). A Study of Machine Learning in Healthcare. In S. Reisman, S. I. Ahamed, C. Demartini, T. Conte, L. Liu, W. Claycomb, M. Nakamura, E. Tovar, S. Cimato, C. H. Lung, H. Takakura, J. J. Yang, T. Akiyama, Z. Zhang, & K. Hasan (Eds.), *2017 Ieee 41st Annual Computer Software and Applications Conference (compsac), Vol 2* (pp. 236–241). Ieee. https://doi.org/10.1109/COMPSAC.2017.164

Bratteteig, T., & Verne, G. (2018). Does AI make PD obsolete? Exploring challenges from artificial intelligence to participatory design. *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2*, 1–5. https://doi.org/10.1145/3210604.3210646

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021). Ethical Machine Learning in Healthcare. In R. B. Altman (Ed.), *Annual Review of Biomedical Data Science, Vol 4* (Vol. 4, pp. 123–144). Annual Reviews. https://doi.org/10.1146/annurev-biodatasci-092820-114757

Conger, K. (2016, August 16). *Computers trounce pathologists in predicting lung cancer type, severity*. News Center. http://med.stanford.edu/news/all-news/2016/08/computers-trounce-pathologists-in-predicting-lung-cancer-severity.html

de Boer, B., & Kudina, O. (2021). What is morally at stake when using algorithms to make medical diagnoses? Expanding the discussion beyond risks and harms. *Theoretical Medicine and Bioethics*, *42*(5–6), 245–266. https://doi.org/10.1007/s11017-021-09553-0

de Bruijne, M. (2016). Machine learning approaches in medical image analysis: From detection to diagnosis. *Medical Image Analysis*, *33*, 94–97. https://doi.org/10.1016/j.media.2016.06.032

Dai, T., & Tayur, S. (n.d.). Designing AI-augmented healthcare delivery systems for physician buy-in and patient acceptance. *Production and Operations Management*. https://doi.org/10.1111/poms.13850

Donia, J., & Shaw, J. (2021). Co-design and Ethical Artificial Intelligence for Health: Myths and Misconceptions. *Aies '21: Proceedings of the 2021 Aaai/Acm Conference on Ai, Ethics, and Society*, 77–77. https://doi.org/10.1145/3461702.3462537

Gregory, J. (2003). Scandinavian approaches to participatory design. *International Journal of Engineering Education*, *19*(1), 62–74.

Grote, T., & Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, *46*(3), 205–211. https://doi.org/10.1136/medethics-2019-105586

Jaremko, J. L., Azar, M., Bromwich, R., Lum, A., Cheong, L. H. A., Gibert, M., Laviolette, F., Gray, B., Reinhold, C., Cicero, M., Chong, J., Shaw, J., Rybicki, F. J., Hurrell, C., Lee, E., & Tang, A. (2019). Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology. *Canadian Association of Radiologists Journal-Journal De L Association Canadienne Des Radiologistes*, *70*(2), 107–118. https://doi.org/10.1016/j.carj.2019.03.001

Juravle, G., Boudouraki, A., Terziyska, M., & Rezlescu, C. (2020). Trust in artificial intelligence for medical diagnoses. In B. L. Parkin (Ed.), *Real-World Applications in Cognitive Neuroscience* (Vol. 253, pp. 263–282). Elsevier. https://doi.org/10.1016/bs.pbr.2020.06.006

Karmakar, S. (2022). Artificial Intelligence: The future of medicine, or an overhyped and dangerous idea? *Irish Journal of Medical Science*, *191*(5), 1991–1994. https://doi.org/10.1007/s11845-021-02853-3

Klang, E. (2018). Deep learning and medical imaging. *Journal of Thoracic Disease*, *10*(3). https://doi.org/10.21037/jtd.2018.02.76

Komura, D., & Ishikawa, S. (2019). Machine learning approaches for pathologic diagnosis. *Virchows Archiv*, *475*(2), 131–138. https://doi.org/10.1007/s00428-019-02594-w

Leimanis, A., & Palkova, K. (2021). Ethical Guidelines for Artificial Intelligence in Healthcare from the Sustainable Development Perspective. *European Journal of Sustainable Development*, *10*(1), 90–102. https://doi.org/10.14207/ejsd.2021.v10n1p90

Mainali, S., Darsie, M. E., & Smetana, K. S. (2021). Machine Learning in Action: Stroke Diagnosis and Outcome Prediction. *Frontiers in Neurology*, *12*, 734345. https://doi.org/10.3389/fneur.2021.734345

Ngiam, K. Y., & Khor, I. W. (2019). Big data and machine learning algorithms for health-care delivery. *Lancet Oncology*, *20*(5), E262–E273. https://doi.org/10.1016/S1470-2045(19)30149-4

Nishida, N., Yamakawa, M., Shiina, T., Mekada, Y., Nishida, M., Sakamoto, N., Nishimura, T., Iijima, H., Hirai, T., Takahashi, K., Sato, M., Tateishi, R., Ogawa, M., Mori, H., Kitano, M., Toyoda, H., Ogawa, C., & Kudo, M. (2022). Artificial intelligence (AI) models for the ultrasonographic diagnosis of liver tumors and comparison of diagnostic accuracies between AI and human experts. *Journal of Gastroenterology*, *57*(4), 309–321. https://doi.org/10.1007/s00535-022-01849-9

O'Connell, A. M., Bartolotta, T. V., Orlando, A., Jung, S.-H., Baek, J., & Parker, K. J. (2022). Diagnostic Performance of An Artificial Intelligence System in Breast Ultrasound. *Journal of Ultrasound in Medicine*, *41*(1), 97–105. https://doi.org/10.1002/jum.15684

Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2020). Participation is not a Design Fix for Machine Learning. https://doi.org/10.48550/arXiv.2007.02423

Zhang, Z., & Sejdic, E. (2019). Radiological images and machine learning: Trends, perspectives, and prospects. *Computers in Biology and Medicine*, *108*, 354–370. https://doi.org/10.1016/j.compbiomed.2019.02.017