MEDICAL AI: OUTDATED REGULATION DELAYS CARE

A Research Paper submitted to the Department of Engineering and Society
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Matthew Pillari

August 5, 2022

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS
Catherine D. Baritaud and Bryn Seabrook, Department of Engineering and Society

MEDICAL AI TO SOLVE DIAGNOSTIC ERROR

One out of every seven diagnoses are incorrect. These errors kill around 60,000 people per year. This devastating figure still fails to capture the additional tens of thousands of patients who suffer chronic pain and injury due to missed or false positive diagnoses (Rodziewicz et al., 2022). Tragically, the technology to prevent a significant portion of these errors already exists. Computer vision technology and medical data availability has progressed to the point that deep learning powered, computer aided diagnosis (CAD) software is able to outperform radiologists on many tasks. However, very few of this medical computer vision research has made it through the U.S. Food and Drug Administration (FDA) review processes and into clinics (Lee et al., 2013). The FDA's review process does not have any specialized pathway for deep learning software medical devices, and therefore the reviewers of artificial intelligence (AI) and deep learning based medical software are ill-equipped to understand and fairly evaluate deep learning products. These deep learning products get stuck in long, expensive, and convoluted review and response cycles with the FDA. These cycles can last a year or more and prevent the latest advancements in computer vision from ever being able to help patients (Van Norman, 2016).

In order to aid in the fight against diagnostic error, the technical aspect of the Capstone research was to develop a deep learning model to predict bloodstream infection in the adult Intensive Care Unit (ICU). The ongoing collaborative effort is led by University of Virginia (UVA) Professor of Computer Science Rich Nguyen, UVA Professor of Medicine Dr. Randal Moorman, UVA Doctor of Infectious Diseases Dr. Chris Moore, Director of UVA's Intensive Care Unit and 2022 graduation speaker Dr. Taison Bell, University of Pittsburgh Professor of Critical Care Medicine Dr. Gilles Clermont, assisted by UVA Graduate Students Jackson Brandberg and Joy Qiu, as well as my fellow Undergraduate Researchers Louisa Edwards and

Zach Boner This team has gathered medical data from the UVA ICU, University of Pittsburgh ICU, and publicly available data from the "MIMIC" dataset to train and validate a deep learning CAD tool that may be used in the ICU environment.

In 1984, Nel Noddings published "Caring: A Feminine Approach to Ethics and Moral Education". From this paper, the Ethics of Care framework was born. Ethics of care is a normative, virtue ethics-based framework that views "true care" and benevolence as virtues. Rosemarie Tong's 1998 paper "The Ethics of Care: A Feminist Virtue Ethics of Care for Healthcare Practitioners" applies this framework to the medical profession. According to Tong, a medical professional expresses the virtue of "true care" by maintaining "active concern for the good of others and of community with them" (Tong, 1998, p. 134). The ability to express care is essential for ensuring the best outcomes for patients, as Tong argues that a person who truly cares will be naturally compelled to do everything in their power to avoid error. This natural compulsion cannot be mimicked in a practitioner who lacks the ability to truly care.

The focus of the STS part of the capstone project relates to the FDA approval process for deep learning powered computer aided diagnosis devices. The FDA's failure to engage with the technical complexity of AI software leads to a slow and partially ineffective understanding of the safety and operation of the device. In this paper, a review of diagnostic error, CAD, and the FDA process will lead into an Ethics of Care based STS discussion of how a CAD-specific FDA approval pathway could improve the quality of care for all patients.

ETHICS OF CARE FOR ENGINEERS

As engineers develop tools meant for patient care, they must also adopt true care as a virtue. According to Ethics of Care, a developer who does not truly care about the patient who will be impacted by a CAD device is more likely to deploy an inadequate product and

neglectfully hurt patients. Just like medical practitioners, ethical engineers are bound by ethics of care to avoid error at all costs.

RADIOLOGY

The practice of radiology is uniquely tied to technological innovations. Unlike other fields of medicine, radiology was born out of the invention of a new way to view the entire internal structure of the human body, rather than a focus on symptomatic or surgical treatment of a particular body part or system. In 1895, Wilhelm Röntgen discovered X-ray radiation and subsequently the novel ability to look inside the living human body to view internal structures. The medical implications of this discovery were immediately recognized and the field of diagnostic radiology was born (Cho, 2016).

After the discovery of X-ray, CT scanning, Sonography, MRI, and fMRI, have each revolutionized the practice of medicine. Practically no other field of medicine has been reinvented via a new technology with such regularity.

DIAGNOSTIC ERROR

In “Medical Error Reduction and Prevention”, Rodziewicz et al. (2022) identify two categories of medical error: errors of omission and errors of action. Errors of omission include things like forgetting to send a prescription, accidentally skipping a step during a procedure, or missing a diagnosis and similar failures of action that lead to injury. Errors of action include incorrect actions like failure to identify a prescription interaction or an incorrect positive diagnosis.

A retrospective study conducted by Hanna et al. in 2018 found that longer shifts and higher volumes of work were correlated with increased errors in radiological diagnosis. In order to better understand what types of errors radiologists tend to make, especially while fatigued,

further research by Lee (2013) enumerated the most common failures that occur in radiology, including failures of visual perception, fatigue, stress, communication failure, and more. Lee found both cognitive and systematic factors for radiological error. Real world factors such as long shifts, oversized workload, bias, and more also serve as lead-ins for the above listed factors.

In understanding cognitive error, the field of medicine recognizes "Type 1" and "Type 2" decision making (Busby et al., 2018). Type 1 decision making is based on the automatic, pattern matching side of cognition. Type 1 decision making is effortless and does not engage much with logical analysis, but rather relies on "feel" and recognition of patterns that commonly indicate a disease. Type 2 decision making is characterized by effortful, analytical, and deliberate consideration before coming to a conclusion (Busby et al., 2018). While it is normal to engage both types when making decisions, it is believed that Type 1 decision making leads to more errors and less explainability than Type 2 decision making. Lee (2013) claims that fatigue is caused by systematic factors such as long shifts and overloading work, which leads to reduced energy and effortful decision making. Since Type 1 (automatic) decisions require a much lower level of effort, a tired radiologist will be more likely to rely solely on Type 1 decision making and thus be much more prone to errors (Busby et al., 2018).

According to Lee, about 30% of abnormalities on X-rays are missed. Studies of lung and breast cancer have suggested an even higher error rate of up to 90% in some circumstances (Lee et al., 2013). Increased training and peer review of diagnoses are the two of the most straightforward and common interventions to decrease diagnostic error. However, these interventions only eat into the time and productivity of radiologists, who are already spread very thin. Novel research into deep learning powered diagnostic software offers a new, less costly intervention.

DEEP LEARNING POWERED COMPUTER AIDED DIAGNOSIS

Deep learning classification, a subfield of machine learning, uses artificial neural networks to learn patterns from labeled data and then use those learned patterns to classify novel data.

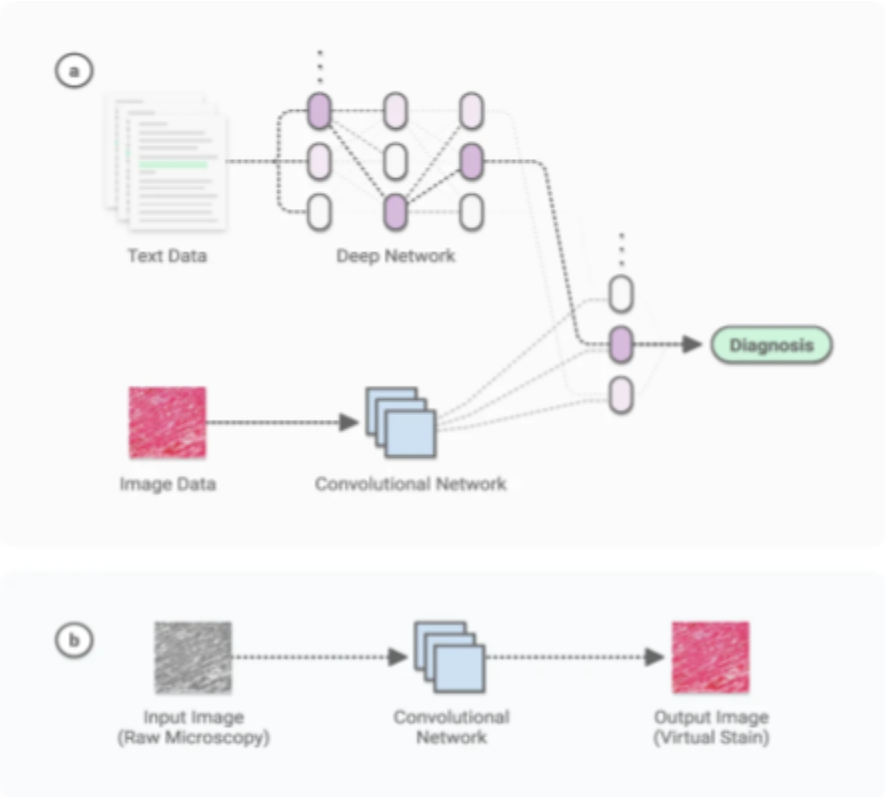

Figure 1: Demonstration of deep learning powered CAD. (a) A system that incorporates written reports with imaging to produce a diagnosis, (b) a system that enhances a raw image to aid the radiologist. (Esteva et al. 2021).

Computer vision is the application of deep learning (See Figure 1) on visual data, such as an X-ray. While performance characteristics and requirements vary widely across implementations, a few core rules of deep learning remain. Deep learning models, particularly computer vision models, need extremely large datasets of labeled training data and plentiful computational resources. Find more information about the technical operation of medical deep learning models in Esteva et al's "Deep learning-enabled medical computer vision" (2021).

Medical imaging classification and diagnosis remained nearly untouched by computational methods until the mid 2010s. Three major factors have led to an ongoing wave of CAD development: the widespread digitization of medical records, the exponential increase of computational power, and the massive research efforts of computer vision researchers across domains. In concert, these factors create a data rich environment, with plentiful computational power, and a mountain of domain-agnostic computer vision algorithms for recognizing patterns and categorizing images (Esteva et al., 2021).

BENEFITS OF COMPUTER AIDED DIAGNOSIS

Automated diagnosis of X-rays via CAD has a few distinct advantages over human radiologists. CAD models, as compared to human radiologists, do not get fatigued, are faster, more precise, more specific, can be deployed anywhere in the world at a low cost, can be trained to eliminate understood biases, and their performance can be evaluated on large sets of data. Also, CAD can be designed to learn from new cases and continually improve accuracy. All of these factors point to CAD being able to reduce error in radiology. Research and commercial institutions alike recognized the scientific and economic value of large scale investment into deep learning powered diagnostic models. Research institutions have continued to publish extremely positive results that prove the efficacy of CAD.

ETHICAL CONSIDERATIONS

By creating a product that is intended to make decisions about the care a patient may or may not receive, there is always the risk that CAD software may make an incorrect assessment and lead to a worse outcome for a patient. While the performance of AI models is often better than radiologists, the impact of an incorrect diagnosis is still just as severe. As such, the

application of AI in medical devices should be approached with ethics of care, and thus a much higher level of caution than other common applications of AI.

The developers of medical AI, more often than not, are researchers and computer scientists that have never taken the Hippocratic oath, and have little to no training in the field of medicine. Meanwhile, the models they create may end up making life and death decisions for millions of patients. While the FDA tries to ensure that any CAD software brought to market is safe, without ethically driven developers, negligence or direct deception during the approval process may lead to the FDA signing off on a product that may not be completely safe for patients (Van Norman, 2016).

The level of extreme caution needed for CAD development is in stark contrast to the attitude of move fast and break things, as Mark Zuckerberg characterized the tech industry. In most other applications of AI, the most severe negative outcome may be less effective social media recommendations or inefficient stock trades. In medical AI, a missed diagnosis could be a death sentence for someone with an emerging and progressive illness such as cancer. This difference in ethical considerations is what leads many experienced AI developers into making mistakes when developing CAD software.

Using the Ethics of Care (Tong, 1998) framework, we can break down the stakeholders, their needs, and how to best deliver true care to everyone involved. True care looks different for each stakeholder. For radiologists and other doctors, CAD should aid them, not cause more work, not increase error, and it should not remove them from the process of care. For patients, true compassionate care means receiving the most accurate results possible, in a timely manner, in a way that makes them feel part of the process. Developers should be able to develop the best tools possible without having their innovation overly limited by regulation. And the FDA should be

able to achieve their mission without being seen as the “enemy” to be deceived, but rather as a collaborator.

Balancing progress and ethics has been an endless effort since the industrial revolution, and the struggle will continue as technology continues to advance and modify every area of life. The role of the government is to keep the citizens protected from negligent and bad faith companies, and as a result has a duty to limit innovation until it is proven safe and effective. Meanwhile, the over limitation of companies, especially in the field of medicine may end up causing more harm than good as patients cannot benefit from technological improvement if the FDA will not allow new AI products to come to market.

HOW THE FOOD AND DRUG ADMINISTRATION EVALUATES DEVICES

The FDA distinguishes CAD software into two categories: Computer Aided Detection (CADe) and Computer Aided Diagnosis (CADx). CADe devices only highlight areas of interest (See figure 2) for further consideration while CADx devices can actual diagnose a specific disease. CADx products, while very useful, have a higher bar of quality control set by the FDA since they take the burden of the decision making out of the hands of the radiologist.

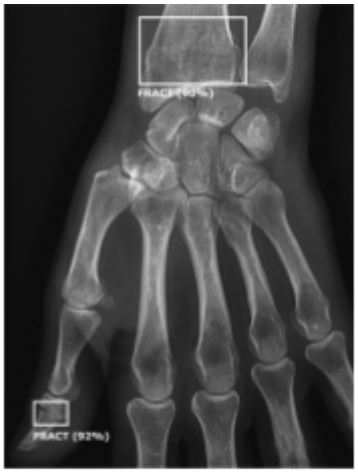

Figure 2: ROI Output of CADe Device. Imagen’s OsteoDetect product. Imagen (2022).

Once a CADe model has been developed, in order to market and use this software in practice, the FDA must first clear the device. In order to achieve FDA approval, two main pathways should be explored. A 510(k) application may be submitted for a product that is believed to be safe, effective, and substantially equivalent to an existing FDA approved product. This 510(k) pathway is useful as it allows for

an expedited process by drawing equivalences to existing products. Since those existing products are proven safe and effective, and this new product is substantially equivalent, therefore the new product must also be safe and effective. The other route to consider is a De Novo application. This route is for products that are deemed too different from any existing product to use the 510(k) pathway.

Figure 3: 510(k) Dependency Graph. Example of a single CAD De Novo approval (Imagen's OsteoDetect) used as the predicate device for a network of 510(k) approved CAD devices. (Adapted by Matthew Pillari (2022) U.S. Food and Drug Administration (2022)).

During the De Novo process, the FDA and the company work together to prove the safety and efficacy of the product. If the FDA deems the product to be safe and effective, the FDA may clear and approve the product for marketing. The De Novo process is longer and more difficult, and as such, companies almost always attempt the 510(k) application first. After either of these

processes, the approved product may then be used as the predicate for any future 510(k) applications (Van Norman, 2016). The dependency graph in Figure 3 is an example of how a single De Novo device (Imagen's OsteoDetect, labeled in red) can be used as the predicate for many further devices. OsteoDetect was the original De Novo approved CADe product that the other products, labeled with blue nodes, used as the predicate device for the more rapid 510(k) approval process. All of the products labeled in blue proved to the FDA that their product was substantially equivalent to OsteoDetect.

CONSIDERATIONS OF RISK

The FDA aims to ensure that all medical devices are safe and effective for patients. In pursuit of this goal, the FDA carefully considers all of the ways in which a device might inflict harm on patients. Sources of risk from a CAD device include false positives and false negatives. A false positive diagnosis may lead to unnecessary work-up and potentially harmful erroneous treatments. A false negative diagnosis may obscure a progressing illness and lead to unnecessary exasperation of a condition, or potentially death if the missed diagnosis is not rectified in time. These considerations are not unique to CAD medical devices; they are the same risks that human radiologists also work under.

FDA SHORTCOMINGS

The FDA does not allow for CAD to continually improve with the new cases it reads in the clinic. Currently, updates to the algorithm, even as small as one extra case of training data, is seen as a substantive change which requires an entire new 510(k) process to be completed. According to FDA documentation, the 510(k) process is supposed to take 90 days. In practice however, the FDA may submit an Additional Information Request and extend the actual length of the process to 200 days or more. With the current pace of machine learning development, a

product with 200 or more days between updates will fall behind the state of the art in a matter of months. This major limitation is currently unavoidable with current FDA regulation.

FUTURE OF FDA CAD REGULATION

Recent FDA documents (U.S. Food and Drug Administration, 2021) indicate that the FDA intends to change how incremental updates to an AI algorithm are regulated. Currently, a new 510(k) is needed whenever such an update is made, but the FDA is exploring new policies that might allow a more streamlined approval process. The FDA has not made any official statements about the future of their AI strategy on a broader scale. However, through observing the release of increasing numbers of CAD devices annually, the FDA is sending the signal that they are open to supporting more widespread use of AI in medicine (Benjamens et al., 2020). As more CAD devices are approved, the options for predicate devices increases, allowing more new CAD products to use the 510(k) process rather than the lengthier De Novo process.

ETHICAL ENGINEERS CAN SAVE LIVES USING AI

Ethics of Care when applied to healthcare professionals dictates that it requires true care to deliver high quality care. In Ethics of Care, a doctor should not require active effort to feel the duty that comes with being trusted with a patient's life. Instead, this care comes naturally and is hard to alienate. Now, as machine learning engineers enter the medical space, it is important that the process of creating healthcare devices is not treated with the same levity as other common engineering fields. Rather, machine learning engineers must subscribe to the Ethics of Care virtue of “true care”. Without this virtue, the engineer may create a negligently crafted product that will endanger patient lives. The virtue of care is especially important to adopt as the FDA is not equipped to fully evaluate deep learning powered CAD software. The slow and frustrating

process of FDA approval may tempt an engineer into detaching themselves from the implications of what they create in pursuit of pushing through the approval process, but this temptation will lead to negligent design and potentially lethal consequences.

An FDA pathway for CAD devices in combination with ethical care from engineers will create an environment where CAD can quickly and directly reduce diagnostic error and save lives.

REFERENCES

Benjamens, S., Dhunnoo, P., & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *Npj Digital Medicine*, *3*(1), 1–8. https://doi.org/10.1038/s41746-020-00324-0

Busby, L. P., Courtier, J. L., & Glastonbury, C. M. (2018). Bias in radiology: The how and why of misses and misinterpretations. *Radiographics*, *38*(1), 236–247. https://doi.org/10.1148/rg.2018170107

Cho, K. W. (2016). Ethical foundations of the radiological protection system. *Annals of the ICRP*, *45*(1), 297–308. https://doi.org/10.1177/0146645316631207

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *Npj Digital Medicine*, *4*(1), 1–9. https://doi.org/10.1038/s41746-020-00376-2

Hanna, T. N., Lamoureux, C., Krupinski, E. A., Weber, S., & Johnson, J.-O. (2018). Effect of shift, schedule, and volume on interpretive accuracy: A retrospective analysis of 2.9 million radiologice examinations. *Radiology*, *287*(1), 205–212. https://doi.org/10.1148/radiol.2017170555

Lee, C. S., Nagy, P. G., Weaver, S. J., & Newman-Toker, D. E. (2013). Cognitive and system factors Contributing to diagnostic errors in radiology. *American Journal of Roentgenology*, *201*(3), 611–617. https://doi.org/10.2214/AJR.12.10375

Pillari, M. (2022). *ROI Output of CADe Device.* [Figure 2]. *STS Research Paper:* Medical AI: Outdated regulation delays care (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Pillari, M. (2022). *510(k) Dependency Graph.* [Figure 3]. *STS Research Paper:* Medical AI: Outdated regulation delays care (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Rodziewicz, T. L., Houseman, B., & Hipskind, J. E. (2022). Medical error reduction and prevention. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK499956/

Sherkow, J. S., & Aboy, M. (2020). The FDA De Novo medical device pathway, patents and anticompetition. *Nature Biotechnology*, *38*(9), 1028–1029. https://doi.org/10.1038/s41587-020-0653-6

Taylor-Phillips, S., & Stinton, C. (2019). Fatigue in radiology: a fertile area for future research. *The British Journal of Radiology*, *92*(1099), 20190043. https://doi.org/10.1259/bjr.20190043

Tong, R. (1998). The Ethics of Care: A feminist virtue ethics of care for healthcare practitioners. *The Journal of Medicine and Philosophy*, *23*(2), 131–152. https://doi.org/10.1076/jmep.23.2.131.8921

U.S. Food and Drug Administration. (2022, August 3). *510(k) premarket notification database.* U.S. Department of Health and Human Services. [Data set] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm

U.S. Food and Drug Administration. (2021, September 22). *Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-Based software as a medical device (SaMD).* U.S. Department of Health and Human Services. https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf

Van Norman, G. A. (2016). Drugs, devices, and the FDA: Part 2. *JACC: Basic to Translational Science*, *1*(4), 277–287. https://doi.org/10.1016/j.jacbts.2016.03.009