Medical AI: Can Artificial Intelligence replace doctors?

STUDENT NO: 220592356

Introduction

This essay aims to discuss how Artificial Intelligence (AI) systems are implemented in the medical field. These systems are used mainly to diagnose patients, recommend courses of treatment, or classify if something is cancerous or not. In the future, these systems will become even more beneficial to patients and doctors as well. Despite their benefits, these systems involve a lot of risks and uncertainties and thus there is still a long way to go for developers and lawmakers/policymakers to successfully integrate these systems into the medical field.

Since my background is in computer science, I was taught machine learning and artificial intelligence. These two new technologies were something completely different, and it was something that fascinated me. For this essay, I wanted to write about something related to these technologies because although they allow the development of autonomous machines they need to be regulated and controlled. Through this project, I wanted to discuss how these technologies could be used today, how to achieve their proper regulation and what are their important ethical issues.

Artificial intelligence has made tremendous progress in recent years, and it is now used in a variety of fields, including medicine. The latter will be discussed in depth in this essay. Patients have long awaited the implementation of systems related to AI because they allow for more freedom and individual-specific treatment; however, physicians do not seem to like Artificial Intelligence and resist its application, because they are unprepared for such a change in clinical practice (Briganti et al., 2020). At the same time, AI has the potential to undermine core values in medicine such as autonomy, privacy, and safety (Martinho et al., 2021). Learning about the perspectives of practitioners on disruptive AI technologies is an important step toward the ethical deployment of such technologies. The ethics of AI implementation in healthcare extends beyond issues of medical practice and career.

It is absurd to believe that autonomous machines can outperform or replace human intelligence, creativity, and responsibility. Humans must be able to anticipate what will occur, because in this case, particularly in the medical field, humans, not machines, are responsible for what occurs. The creators of these machines believe that their creations will do the right thing, even though their machines have flaws such as algorithmic bias (Shneiderman, 2021). These machines are built on the researchers' intuition about what constitutes a "good" explanation (Andersen et al., 2021).

The systems that are used in the medical field are mostly black-box systems and no one, not even the developers, knows how they extract their results. These AI systems should log activity to allow for transparent and retrospective failure reviews. Including activity, logs would ensure appropriate accountability, particularly in applications with significant ramifications for people and organizations (Shneiderman, 2021). For example, if a machine is used to determine whether a mole is cancerous or not, it should show the physician the path it took to classify it as cancerous. Because cancer treatment places a significant financial and emotional strain on the patient, it is critical to ensure that the machine did not misdiagnose the patient or make an incorrect calculation during the process.

Background Literature

Because we are talking about black-box machines that use Artificial Intelligence and Machine Learning, it is important to define the terms. When a device tries to copy cognitive functions such as learning and problem-solving, the term AI is used (Pesapane et al., 2018). AI is a field dedicated to developing systems that perform tasks that otherwise human intelligence would be imperative. Machine Learning (ML) is a term defined by Arthur Samuel in 1959 to describe a subfield of artificial intelligence that encompasses all approaches that enable computers to learn from data without being explicitly programmed; this field has been extensively applied to medical imaging (Pesapane et al., 2018).

There is a lot of material that talks about algorithms that are like a black box, and there has been much debate about it (Topol et al., 2019). Even for experts, unpacking the way an algorithm reaches its output is frequently impossible because it is too difficult or it is protected as a trade secret. "You can't unbundle them in the way that a statistic can be pulled apart and the variables isolated," says Naylor (Vogel et al., 2019).

This obscurity has given rise to demands for explanation. The General Data Protection Regulation of the European Union requires transparency and simplification of algorithms that are classified as black boxes before they can be used to diagnose or treat a patient (Topol et al., 2019).

Another important aspect of these systems is that they use Machine Learning, which introduces the concept of prediction confidence. For example, if a system predicts a cancerous lesion most often the system's output would be the prediction and a confidence percentage. The confidence shows how sure is the system about its prediction. If the systems are difficult to interpret, the clinician must know how to use the system correctly and understand its output. The clinician should also understand the system's confidence in its prediction and whether is reasonable. If the system's confidence is low, the best design practice would be to failsafe and refuse to make any predictions (Challen et al., 2019).

<u>In-depth discussion of challenges</u>

Several studies have been conducted to compare experts in medical fields with algorithms that have been developed for that field.

In one study, the algorithm performed 0.96 in detecting carcinoma and 0.94 in detecting melanoma. At the same time, dermatologists using the same dataset made classifications with 0.76 and 0.86 accuracies, respectively (Topol et al., 2019).

In another study, an algorithm was tested for its accuracy against four expert radiologists, and the conclusion was that the algorithm outperformed the radiologists. The algorithm's accuracy reached 0.63 for pneumonia and 0.87 for heart enlargement or a collapsed lung (Topol et al., 2019).

These algorithms can process and observe millions of numbers of inputs instantaneously. As a result, an AI-powered application can classify suspicious skin areas better than dermatologists. Therefore, AI can help with tasks that radiologists argue on, such as identifying pulmonary tuberculosis on chest radiographs (Buch et al., 2018).

Because these systems are implemented/trained using specific datasets they sometimes come with their challenges. One example was a system that was trained to diagnose skin cancer. This system was trained and used mainly data from white patients. When the same system was used for patients of color, the accuracy dropped significantly.

Another example was IBM's Watson supercomputer (Vogel et al., 2019) which made incorrect recommendations for treating cancer patients. This was revealed from leaked internal documents. These recommendations were not in line with national treatment guidelines. Fortunately, according to the documents, no patients were adversely affected, according to the documents, which would have been detrimental. The reason for the problem was the training of the computer. When training the AI, the doctors and the engineers used hypothetical patient chases rather than real patient data. The training included treatment options for each type of cancer. As we know the dataset that had been given was synthetic, which means that Watson's recommendations were the doctors' treatment preferences and not a machine-learning analysis of real patient cases.

When developing AI systems for healthcare, they must avoid bias in their algorithms and should demonstrate how they reach their results. To avoid bias being included in algorithms of medical AI, both the FDA of the United States and the EMA of the European Union have taken steps to ensure that these systems are properly trained.

The FDA emphasizes the importance of the appropriateness of these systems for a racially diverse data set that represents the patient population for which the algorithm will be used. The EU has imposed an act according to which to train such systems the data used must meet certain quality criteria that are subject to proper data management. The data must be checked so that it doesn't include any potential bias. Additionally, the data must be relevant and representative of the group of patients that the systems that are going to be applied to. Moreover, the data must be error-free and include all patient information (Vokinger et al., 2021).

However, there are some risks associated with this act. The European Union is a region that includes many countries. Each country included has different ethnic characteristics. Today when a medical system is approved by one country of the EU, it may be approved by other countries without having to go through a different approval process. As mentioned earlier each country has different ethnic characteristics, so if country A adopted an algorithm for its population, the same algorithm may not work for country B because the population of country B is not represented in the algorithm. Nevertheless, this algorithm gets approved without a second thought. To get the approval process right, then the developers would have to retrain the algorithm based on the European datasets that represent the society of country B. This process of training an algorithm, again and again, is going to increase costs, delay the authorization process, and harm innovation; none of which is ideal (Vokinger et al., 2021).

When talking about bias, it's important to consider that physicians have a bias as well when it comes to treating patients (Dawson et al., 1987). Confirmatory bias and ego bias are the two most important biases that doctors have. Confirmatory bias refers to the tendency of doctors to only look for evidence that can confirm their hypotheses and not check for anything else. This bias not only causes one to see confirmatory evidence while ignoring all other possibilities, but it also influences how data is interpreted. The ego bias is when doctors, particularly surgeons, believe that the mortality rate of their patients is lower than the mortality rate for the entire service. To summarize, we cannot criticize the use of bias in system design without also acknowledging that doctors treat patients with biases of their own.

Concerning biases, it was briefly documented how physicians have their own biases and thus the systems can never be completely free of bias. It is also important to note that even access to the healthcare system can vary by socioeconomic status, race, and even gender (Vokinger et al., 2021) because these systems are trained using data from previous years. The fact that the data used is historical may result in a bias in the data from previous societies and this may create incorrect training of the system which will lead to incorrect decisions. An example of this is the diagnosis of myocardial infarction in women. This diagnosis is normally presented with atypical symptoms in women. As a result, when training the AI algorithm using historical record data it will not recommend cardiac ischemia testing for women, because they won't have the symptoms that would make the algorithm recommend the test. This will lead to a delay in potentially life-saving treatment for women.

Typically, when a system or product is released, the product is subject to liability that includes injuries that may occur due to its poor design or failure to warn of a potential hazard. However, this liability clause is still unclear on how it can be applied to the technology of AI if the decision-making process is independent of the medical provider. According to Gilke, Minssen, and Cohen, this can be applied by making the developers mention any potential bias that their algorithm includes, so the medical provider can be aware and can consider it when using that system (Marotta, 2022).

Conclusion and future work

Throughout the essay, AI systems in the medical field were introduced, as well as some of the ethical and legal issues that come with them. Furthermore, it was shown how legislative bodies such as the FDA have tried without success to reduce the risks associated with them. As a result, until those issues are resolved, these systems cannot be used autonomously in this context.

As discussed in the essay, the aim of these systems is not to replace doctors. They exist to help physicians by providing them with useful recommendations. Many of the goals of recent research have been to compare solutions involving AI with physicians as if they were competitors (Briganti et al., 2020). Future work should compare clinical trials with AI. Only then will AI be accepted as a supplement to physicians.

My opinion about AI systems is that they can be very useful tools for doctors. The systems can be beneficial for them and make the diagnosis process way faster because they can act as an assistant to the doctor. In order to be useful tools, these systems cannot be black boxes, but they should provide evidence of how they reached that result and evidence to back up their opinion about something. Additionally, AI systems in my opinion can be a substitute for doctors in some scenarios. For example, they can be used in remote places where there are not any doctors at the time or in regions where access to healthcare professionals is limited. However, if a system is used as a replacement for a doctor, it should be checked and validated by experts at regular intervals as calibration.

References

Andersen O., T., Nunes, F., Wilcox, L., Kaziunas, E., Matthiesen, S., & Magrabi, F. (2021, May). Realizing AI in healthcare: challenges appearing in the wild. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-5).

Briganti, G.; Le Moine O. (2020) Artificial Intelligence in Medicine: Today and Tomorrow. Frontiers in Medicine 2020, Vol. 7, Article 27, https://doi.org/10.3389/fmed.2020.00027.

Buch V, Irfan Ahmed I., Maruthappu, M. (2018) Artificial intelligence in medicine: current trends and future possibilities British Journal of General Practice, 68 (668): 143-144. DOI: https://doi.org/10.3399/bjgp18X695213.

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. BMJ Quality & Safety, 28(3), 231-237.

Dawson, N. V., & Arkes, H. R. (1987). Systematic errors in medical decision making judgment limitations. Journal of general internal medicine, 2(3), 183–187. https://doi.org/10.1007/BF02596149.

Marotta, L. (2022). When AI is Wrong: Addressing Liability Challenges in Women's Healthcare.

Martinho A., Kroesen M, Chorus C., (2021) A healthy debate: Exploring the views of medical doctors on the ethics of artificial intelligence, Artificial Intelligence in Medicine, Volume 121, https://doi.org/10.1016/j.artmed.2021.102190.

Pesapane, F., Codari, M., Sardanelli, F. (2018). Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. European radiology experimental, 2(1), 35. https://doi.org/10.1186/s41747-018-0061-6.

Shneiderman, B., (2021) "Human-Centered AI." Issues in Science and Technology 37, no. 2, 56–61.

Steinhubl S., Muse E., Topol E. (2015) The emerging field of mobile health. Sci Trans Med. (2015) 7:283rv3. https://doi.org/10.1126/scitranslmed.aaa3487.

Topol E. .(2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 25:44–56. https://doi.org/10.1038/s41591-018-0300-7.

Vogel, L. (2019). Rise of medical AI poses new legal risks for doctors.

Vokinger, N., Gasser U. (2021). Regulating AI in medicine in the United States and Europe. Nat Mach Intell 3, 738–739 https://doi.org/10.1038/s42256-021-00386-z.