Mindful Machine Learning:

Ethical Considerations for Data-Driven Epilepsy Research

Amadeo Cabanela

Practicum: Boston Children's Hospital

"Dad, will I ever get the chance to drive someday?", my son often asks. He has suffered from epilepsy for over 20 years. My primary motivation for joining the Master of Science in Data Science (MSDS) program at the University of San Francisco has been to study machine learning and contribute to neuro research with the hope of enabling my son and the millions of people affected by epilepsy to seek better treatment and to enjoy everyday freedoms using cutting-edge technology like self-driving cars. According to the World Health Organization (WHO), epilepsy affects approximately 50 million people worldwide, making it one of the most common neurological disorders globally. Epilepsy is a condition where people experience seizures due to abnormal brain activity. Seizures result from sudden bursts of electrical activity in the brain, leading to a wide range of manifestations, from brief lapses in consciousness to convulsive movements in grand mal seizures. The unpredictable nature of seizures poses significant challenges for individuals living with epilepsy, impacting their quality of life, social interactions, and mental well-being. It poses significant challenges for individuals affected by it and for the scientific community seeking effective treatments.

Fortunately, innovators are rapidly advancing data science in healthcare and scientific research has made significant strides in unraveling the underlying mechanisms of epilepsy in recent decades. I am fortunate to have been working at Boston Children's Hospital (BCH) for my practicum with my mentor and expert in epilepsy research, Dr. William Bosl. One project I worked on was feature extraction using techniques like tensor factorization to identify latent features relevant to Rolandic Epilepsy in children. Another closely related and major project for the practicum is using machine learning for seizure forecasting. At a high level, this project entails analyzing brain activity data and using these features to give a probability of a Rolandic Seizure happening (which may be more helpful than providing a binary prediction). These projects have been an eye-opening experience and allowed me to see the power of leveraging state-of-the-art machine learning techniques to propel scientific research forward and bring people with epilepsy closer to more effective treatment. While data science and machine learning have the powerful ability to advance neuro research, it is vital to incorporate ethical considerations into the project design. This includes ensuring the responsible use of sensitive medical data and transparency in how developed models operate to ultimately safeguard patient trust and well-being. Having a data ethics framework in place for neuro research projects enables researchers to maximize benefits for patients while minimizing potential risks and harms. The seizure forecasting project has many ethical considerations to think about at different points within the project pipeline. Boston Children's Hospital has controls in place for many of these. There are also potential areas for improvement.

While scientific advancements in epilepsy research hold promise for improving diagnosis and treatment, they also raise important ethical considerations, particularly regarding the use of

patient data, from their general health information to recorded brain activity. The electroencephalogram (EEG) is the most common test used to diagnose epilepsy. It records the electrical activity of the brain using small, metal discs called electrodes attached to the scalp. EEG recordings contain sensitive information about an individual's neurological condition, including the frequency, duration, and severity of seizures. Having unrestricted access to EEG data comes with risks and potential harms. Disclosing sensitive information about a patient's neurological condition can lead to stigmatization and discrimination in several ways. First, the lack of social awareness about neurological conditions has led to misconceptions and stereotypes. Unauthorized disclosure of seizure-related information might reinforce these misconceptions, leading to social stigma in which people with epilepsy are unfairly judged. Furthermore, employers may have misconceptions about the ability of people with neurological conditions to perform job tasks. Unrestricted disclosure of patient data related to their neurological condition may result in job discrimination, limited career opportunities, or even termination of employment. In one of our readings, "Privacy as Civil Right", Alvaro Bedoya explores the concept of privacy at the intersection of civil rights and civil liberties. Towards the end of the lecture, Bedoya described how enslaved individuals during the Underground Railroad era relied on privacy to plan their routes, evade capture and find freedom. In alignment with the idea of privacy for protection, given the historical discrimination and stigmatization of neurological conditions and epilepsy, privacy remains essential for the well-being of people with these conditions. Safeguarding their privacy empowers them to manage their health on their own terms and protect themselves from prejudice. Therefore, the collection, storage, security, analysis, and maintenance of EEG data must be conducted in accordance with rigorous ethical standards to safeguard patient privacy and autonomy. BCH is outstanding when it comes to protecting their patients' EEG data and adhering to data privacy protocols. First, they have a thorough process for onboarding employees and granting system access. Then they have multiple layers of data protection, including Fast Healthcare Interoperability Resources (FHIR), which is a standard set of rules and specifications for exchanging electronic healthcare data efficiently and securely by following a composition approach where exchangeable content is gathered and sent on an as-needed basis.

Related to ensuring data privacy, another major ethical consideration in epilepsy research is the issue of informed consent. Given the invasive nature of EEG monitoring and the potential risks associated with data sharing, researchers must obtain explicit consent from patients before collecting and using their brain data for research purposes. This entails providing patients with clear and comprehensive information about the nature of the research, its potential risks and benefits, and their rights to privacy and confidentiality. The BCH Epilepsy dataset is decoupled from their Electronic Health Record (EHR) system, which is an electronic version of a patient's medical history maintained by their provider over time. It may include all of the key administrative clinical data relevant to that person's care under a particular provider, including demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports. This decoupled design gives Boston Children's Hospital another layer of data protection because patients' EEG data has to be correlated with the EHR system when we want to identify the EEG file belonging to a particular patient. This reduces biases that our machine learning model might draw related to demographic groups and allows the model to focus more on revealing insights more directly related to seizure forecasting like finding patterns in the electrical signals from the brain activity data.

Explainable AI and model transparency have been a long-running challenge and is especially important when it comes to using model predictions to make decisions in epilepsy treatment. Explainable AI is important for the seizure forecasting practicum project for several reasons. First, the model prediction might influence a patient's treatment plan, including their medications which can sometimes have painful side effects. Therefore, patients, their healthcare providers, and people in their support system need to be able to trust the predictions made by the model in order to make justifiable data driven medication choices and treatment decisions. Secondly, explainable AI can help identify errors or biases in the model that we can correct in order to make its predictions more reliable. One area of improvement for the seizure forecasting project is the metrics used to measure the success of the model and overall project. As Goodhart's Law states, "When a measure becomes a target, it ceases to be a good measure." When the seizure forecasting project was previously explored, researchers framed it as a binary classification problem with discrete class labels (seizure = 1, no seizure = 0), using accuracy and ROC-AUC as metrics. Since the goal of this approach is to empower patients to seek appropriate care based on seizure forecast, BCH researchers decided to change their approach and frame it as a probabilistic binary classification (with continuous predictions) instead. With this change, patients can be more informed about the risk of a seizure on a probability scale ranging from 0 to 1. For example, a patient's prediction with a .35 (or 35%) probabilistic risk of having a seizure for a certain day is more informative than the old discrete measurement of 0 because the patient will now have a better idea of the risk and therefore be able to plan their activities for the day accordingly. While this is an improvement, this project still relies on the model metrics for measuring project success. We can take this further by including improvements proposed by Rachel Thomas and David Uminksy in their paper "Reliance on Metrics is a Fundamental Challenge for AI." They propose a framework for a healthier use of metrics, one of the principles being to "Use a slate of metrics to get a fuller picture and reduce gaming [of metrics]" (Thomas et. al.). In accordance with this principle, we can extend the use of sensory EEG data and clinical information like age and sex for seizure forecasting. It might be helpful to integrate other data sources like patient interviews and mood assessments over time to provide a clearer picture of a patient's risk for an upcoming seizure.

In conclusion, while using brain data and machine learning to help people with epilepsy is highly beneficial, it is also a big responsibility that requires careful consideration of its ethical consequences. Working on the seizure forecasting machine learning project has required me and BCH researchers to consider the following to maximize benefits to patients while minimizing harm: data privacy, informed consent, and explainable AI for model transparency. Putting these into action requires layers of protection for handling data carefully and securely, practicing diligence in obtaining patient consent, and carefully considering effective ways of measuring the project's success in addition to model metrics. Moving forward, there are many opportunities for BCH to further integrate ethical frameworks into data projects, some of which are currently underway. For example, one interesting project BCH researchers have looked into is using federated learning with wearable technology to collect physiological signals at scale for enhanced seizure prediction while maintaining patient privacy. All in all, I am so grateful that my experience with the Boston Children's Hospital practicum gave me the opportunity to learn about and contribute to improving patient outcomes for people with epilepsy, a cause I care deeply about. This experience has inspired me to approach future machine learning and data science projects with social responsibility, from ensuring proper security controls are in place to carefully considering effective metrics. I plan to take what I learned from my practicum experience and the

Ethics in Data Science class to empower people to make powerful data driven decisions while maintaining ethical standards to ensure the well-being of all parties involved.

Works Cited

World Health Organization. "Epilepsy." Fact sheet. The White House, Washington, D.C. n.d. Web. 28 Apr. 2024.

Bedoya, Alvaro, Privacy as Civil Right (May 12, 2020). New Mexico Law Review, Vol. 50, No. 3, 2020, Available at SSRN: https://ssrn.com/abstract=3599201

Wikipedia contributors. "Fast Healthcare Interoperability Resources." Wikipedia, The Free Encyclopedia, 3 Apr. 2024. Web. 4 May. 2024.

Vieluf S, Cantley S, Jackson M, Zhang B, Bosl WJ, Loddenkemper T. Development of a Multivariable Seizure Likelihood Assessment Based on Clinical Information and Short Autonomic Activity Recordings for Children With Epilepsy. Pediatr Neurol. 2023 Nov;148:118-127. doi: 10.1016/j.pediatrneurol.2023.07.018. Epub 2023 Jul 28. PMID: 37703656.

Thomas RL, Uminsky D. Reliance on metrics is a fundamental challenge for AI. Patterns (N Y). 2022 May 13;3(5):100476. doi: 10.1016/j.patter.2022.100476. PMID: 35607624; PMCID: PMC9122957.

OpenAI. "ChatGPT 3.5." OpenAI, 2022. https://chatgpt.com/.
        Note: ChatGPT was used for grammar review.