Abstract

Machine learning is becoming an increasingly prevalent tool in medical decision making. Propelled by developments in processing power, memory, storage and access to large amounts of data, computers are being asked to tackle increasingly complex learning tasks within medicine. Using machine learning, we can train models on patient data from thousands of patients to make objective decisions.

A startup out of Yale, Spring Care is using machine learning and technology to help find patients relief form mental illness faster. For my project, I worked with Spring Care to develop an app-based questionnaire to help improve treatment following first-episode psychosis. Machine learning techniques have been recently applied to psychosis treatment selection, where there is a wealth of potential baseline predictors and outcomes of interest, and has been published in Lancet Psychiatry (Koutsouleris et al., 2016). Results of the study demonstrated the feasibility of using pre-treatment clinical information to determine generalizable, individual predictions of treatment outcomes. Using the machine learning algorithm resulted in a 40% increase in prognostic certainty compared with pre-test outcome probabilities in our population. Development of the algorithm is proof of concept that machine learning can aid in this specific treatment selection.

The algorithm is only the first step for effective use of machine learning in a clinical setting. A key factor in solving this discrepancy is successful implementation—ease of use, integration with the clinical workflow and resulting meaningful information received for both the clinician and the patient. My project was to implement a questionnaire for clinicians to use to give them access to the model as a tool to aid their treatment selection.

Background

The effects of mental illness on America are enormous. According to the National Institute of Mental health, in 2015 there were an estimated 43.4 million adults in the United States, 17.9%, with a mental illness within the past year.¹ These numbers are staggering, and yet treatment of mental illness is flawed. Adults in the U.S. living with serious mental illness die, on average, 25 years earlier than others, largely due to treatable medical conditions.² The prevalence of mental illness has consequences not just for the quality of life of those individuals, but also more broadly for America. Serious mental illness costs American $193.2 billion in lost earnings per year.³

My project is focused on improving the choice of treatment following first-episode psychosis. The term ‘psychosis’ is used to describe conditions that affect the mind where there
has been some loss of contact with reality. This is a broad term which can include hallucinations (seeing, hearing, smelling, tasting or feeling something that is not real) and paranoia, or delusions (believing in something that is not real even when presented with the facts). People experiencing psychosis may exhibit personality changes, bizarre behavior, difficulty with social interaction and impairment carrying out daily activities.  

First-episode psychosis refers to when a person first shows signs of beginning to lose contact with reality. The exact causes of psychosis are unknown; however, it is believed to be caused by a combination of genetic factors, biological factors and environmental stressors. Improving treatment following first-episode psychosis is especially important because the less time that passes between the onset of psychosis and initiation of appropriate treatment, the greater the patient’s odds of recovery.  

Currently, following first-episode psychosis, up to 50% of patients, develop unfavorable disease outcomes. This occurs despite established treatments for schizophrenia and awareness about multiple outcome predictors. These prognostic factors, which have been identified in group-level studies, include sociodemographic factors, disease course variables, treatment adherence and response, psychiatric comorbidity and functional and cognitive deficits.  

Clinicians lack the tools to dictate how and which of these factors should be combined for prediction. Currently, clinicians make treatment decisions by personally weighing a varying host of diagnostic tools with their own previous experience. This process is limited, subjective and inherently flawed. To try and improve the important task of treating first-episode psychosis, researchers have developed an algorithm that uses machine learning to determine the top ten predictors for a poor vs. good treatment outcome. This has the potential to simplify, standardize and ultimately improve treatment following first-episode psychosis.

The Study

In a study published by Lancet entitled, Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach, Koutsouleris et al. demonstrate the first evidence for the feasibility of generalizable, individual-patient prediction of treatment outcomes in first-episode psychosis using pre-treatment clinical information. In the study, the researchers applied machine learning to sociodemographic, clinical and neurocognitive variables from The European First-Episode Schizophrenia Trial (EUFEST) database to assess the predictability of outcomes, the generalizability of prognostic models and whether differential antipsychotic treatment efficacy exists in patients with poor versus good 1-year prognoses. The researchers set up a rigorously cross-validated learning strategy that autonomously identified the most accurate and parsimonious predictor patterns.  

This machine learning strategy, further detailed in the cited paper, addressed generalizability to new patients by repeated, nested cross-validation. To examine the predictive use of different variable sets, the respective variable blocks were discarded and model was retrained. To measure geographical generalizability, the researchers used a leave-site-out validation approach. To explore the effect of reducing the variables used, a variable pruning step was added to this leave-site-out analysis. To test validity of these reduced variable sets, each prediction model was retrained with its respectively condensed variable set. This is all to illustrate the thorough measures that were carried out. This resulted in the 10 most predictive variables.
This is the first study to show accurate and generalizable individual-patient outcome predictions after 4 and 52 weeks of treatment for first-episode psychosis. This approach might provide a more than 40% gain in correctly predicting outcomes across geographically distinct health-care settings and patient populations. Importantly, the models revealed that the efficacy differences previously found among the five antipsychotics showed substantially different efficacy only in the at-risk group of patients with poor outcome prognoses. The performance of this model exceeded previously published algorithms trained on the EUFEST datasets.7

Project Motivation

As stated above, the findings of this study have significant implications for improvements in treatment of first-episode psychosis. This, however, relies on a clinician’s ability to use this predictive model in the process of assigning treatment. The algorithm is only the first step for effective use of machine learning in a clinical setting. Large enough medical data sets and adequate learning algorithms have been available for many decades, and although there have been many papers applying machine learning algorithms to medical data, very few have contributed meaningfully to clinical care. A key factor in solving this discrepancy is successful implementation—ease of use, integration with the clinical workflow and resulting meaningful information received for both the clinician and the patient.8 My project was to solve this imparity between the existence of and the usability of a machine learning algorithm to improve first-episode psychosis treatment.
Design Considerations

The goal for this project was to implement the questionnaire in a usable way for a typical clinician setting.

Web App vs iOS App?

The first consideration was web app vs iOS app. Ultimately, I decided on an iOS application to make it more usable in a clinical setting. It seems likely that medicine will be trending to, if it’s not there yet, the use of mobile technology tools during patient appointments. Building this as an app properly positions it as another tool available to the clinician to aid and inform decision making. Additionally, the choice to make it an app ensures that it will be developed with a mobile-first mentality.

Device?

Spring Care is currently developing a tool to help clinicians select the right antidepressant for patients. This tool is being designed for an iPad, so in-line with this previous design choice, I settled on an iPad as the main device for the Psychosis Questionnaire app. Further support for this choice came from the larger screen of an iPad. This characteristic of the iPad enabled larger question formats, which are well-suited to the Psychosis Questionnaire app, as some of the questions involved are complex.

Questionnaire Length?

The length of the questionnaire is not something I had complete control over as the Koutsouleris et al. study dictated the necessary questions. However, from a usability standpoint, I wanted to ensure that it could be administered in a typical appointment setting. I set a rough goal of less than an hour for the administration of the questionnaire.

Method

My application builds upon the existing Spring Care application. The first step in the implementation was JSON-ifying the various questionnaires. Rather than hand typing each question into the appropriate JSON format, which would be tedious and mindless, I broke each questionnaire into lists of reusable pieces and ran python scripts to craft these pieces into a final complete JSON format. These files are then served by the ruby on rails code to the iOS app to be rendered as questions. The model relies on answers to sections of:

- Background Questions
- Camberwell Assessment of Need Short Appraisal Schedule (CANSAS)
- Mini International Neuropsychiatric Interview Plus
- Positive and Negative Syndrome Scale
- Global Assessment of Functioning
Using Ruby on Rails as the backend, I built out the routes and controllers to handle serving, receiving, and pre-processing the different questionnaires. The first challenge was translating these pen and paper questionnaires into intuitive and uncomplicated touch questions. The most complex part was implementing the conditional logic that, depending on certain answers, automatically enabled the user to skip over unnecessary chunks of the questionnaire. This was difficult because it required understanding the intricate flow of the questionnaires while still satisfying all of the requirements that the Koutsouleris et al. model expects. It also required processing certain portions of the questionnaire at different times rather than all at once upon its submission. Thus, the portion of the questionnaire processed first could dictate what questions followed, leaving out ones that might no longer be necessary to ask.

**The Application**

![Image of Spring app](image1.png)

**Figure 3:** Screen shots from the application. The first shows the clinician selection where the spring app can have multiple clinicians and each will record patients under their own profile. Next is the patient information screen. The third is the overall explanation screen. The fourth is an example background question.
Testing

I deployed a staging version of Psychosis Questionnaire to an iPad and tested it on two different cohorts of people: my friends and graduate students in the clinical psychology department. My goals for testing was to catch errors, determine what areas needed more information, test flow and question order, get approximate questionnaire administration times and explore the feasibility of this fitting into a clinician setting.

Friends

I tested my app with 2 different friends. I gave them a rough overview of the goal of the app and then had them administer it to me, where they acted as the clinician and I acted as a patient. From these tests I learned that the initial order of the questionnaires could be improved. Some questionnaires depend on the patient’s responses to questions, whereas others required only the clinician’s judgement. Initially, I had these mixed together (i.e. a questionnaire involving patient responses would be followed by a clinician only section and then back to a patient question section). This caused some awkwardness in the administration of the questionnaire. Further, the Positive and Negative Syndrome Scale (PANSS) involves the clinician ranking the patient on certain dimensions following an interview. Originally, the PANSS was the second questionnaire item. However, following my friends’ testing of the application, I moved it to be the penultimate item.

Clinical Psychology Graduate Students

I tested my app with 2 different graduate students both in the clinical psychology department. I had them test it with each other where one acted as the clinician and the other as the patient. As such, I was able to observe closely and note any questions that came up. This was particularly helpful because they had knowledge of and experience with some of the questionnaires involved. Thus, when they asked a question or signified that something was unclear, it was a good sign that I needed to add more explanation. From this test, it became apparent that one of the sections was unintuitive and lengthy.

Figure 4: Testing the application with graduate students. On the left is the graduate student acting as the clinician and recording responses on the app. On the right is the graduate student
acting as the patient.

Future

Following the tests above, the next improvement of the app would be simplifying different components of the questionnaires. To ensure that the validity of the questionnaire remains after the simplification of questions, a side-by-side comparison must be conducted. The next testing step would be to have clinicians test the app. Even with grad students, there are still questions about how much experience clinicians have with administering the different components. Because the application is ultimately meant to be administered by clinicians, maximizing the usability for these target users is the goal. In the future, this app will be another, more objective tool for clinicians to use in the complex process of selecting treatment for patients following first-episode psychosis.

Acknowledgements

Ruzica Piskac, thank you for eagerly agreeing to be my advisor and imparting your excitement about software solutions to real world problems. Adam Chekroud, thank you for your long term vision of the project and your continued support throughout the entire process. Abhishek Chandra, thank you for answering my questions at all hours of the day and being an excellent mentor. Nikos Koutsouleris, thank you for your work developing the machine learning model and then your time answering all of my questions about the details of it.

References