A Clinical Perspective on Big Data in Mental Health

3

John Torous, Nikan Namiri, and Matcheri Keshavan

While the word *analysis* holds special meaning in psychiatry from a psychodynamic therapy perspective, our lives are also constantly being analyzed by machines. Whether we realize it or not, computers have been fully integrated into our lives and devices, ranging from the smartphone we use for phone calls, the cars we use to drive, and the internet we use to communicate across. All of these computers contain algorithms that seek to analyze and understand our behaviors or intentions: the smartphone to remind of appointments and recommend navigation routes, the car to automatically brake if a child jumps in the road, the search engine to offer website links to answer a question. The same algorithms that make today's computers useful are not only restricted to increasing efficiency, ease, and comfort. They can also be, and already are, used to study, predict, and improve mental health. In this chapter we explore the rapidly expanding field of digital psychiatry with a focus on the synergy between data and algorithms that hold the potential to transform the mental health field.

As discussed in other chapters, the accessibility of new technologies, like smartphones, and access to the data they generate have paved new roads for innovation and discovery in many fields. Among them, mental health has received

J. Torous (✉)
Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
e-mail: jtorous@bidmc.harvard.edu

N. Namiri
Department of Bioengineering, University of California Los Angeles, Los Angeles, CA, USA

M. Keshavan
Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

I. C. Passos et al., *Personalized Psychiatry*,
https://doi.org/10.1007/978-3-030-03553-2_3

some of the most prominent advances. Consider for a moment the vast amount of information a smartphone can provide relevant to behavior and mental health. Geolocation data can provide objective measures of exercise and activities, phone call and text message logs measurement of social engagement, voice samples clues to mood, error rates in typing a window into cognition and mental state, and so on. There is so much data generated from smartphones alone that there is a need for collaboration with data science fields in order to help make sense of these myriads. Understanding this collaboration and work, along with the intersection of psychiatry and data science, offers an exciting window into the new world of big data.

To understand this new world of data and algorithm, it is first necessary to lay a groundwork in the concepts of big data and machine learning. While these two terms are often used broadly and their exact definitions are beyond the scope of this chapter—understanding their meaning in the context of clinical psychiatry is an important first step.

Big data is characterized by three principles: velocity, volume, and variety, together known as the three V's of big data. Smartphones utilized for mental health offer an example of high velocity data, as data streams such as geolocation, keyboard strokes, and phone call logs are constantly flowing through from devices and into computers where the data can be stored. Smartphones in mental health are also a paradigm for high volume data, as smartphones can provide a constant data stream from features, such as accelerometer and CPU, which provide millions of data points in a matter hours. In addition to the velocity and volume, smartphones are an example of the variety in big data. Consider the wide range of data types a smartphone can collect that is relevant to mental health, ranging from geolocation to weather data, call and text logs to light sensors, voice data to keyboard strokes, and more. Thus, when considering the velocity, volume, and variety of smartphone data for mental health, it is easy to see why this big data is unlike most other data streams currently utilized in clinical psychiatry.

An issue arising from the mass quantities of big data is creating effective means of analyzing and drawing accurate conclusions from the data, which is precisely where machine learning comes in. Other primary issues include the ethics, patient perspective, security, and appropriate clinical utilization of this data, which are covered in the upcoming sections, as well as later chapters.

The analysis of these big datasets and further extrapolation into feasible application is the crux of machine learning. Machine learning enables physicians and researchers alike to analyze patient data using methods novel to the clinic. The nature of big data means that we need computers to assist in finding meaning or patterns in the data. This does not mean that a psychiatrist allows the computer to make clinical judgments, but rather that he/she allows the computer to suggest potentially useful information garnered from a sea of big data from, for example, a patient's smartphone. Perhaps the machine learning algorithm noted a pattern that when the patient does not leave the home or exercise, mood worsens. This is information the psychiatrist can now use to inquire more and start a discussion with the patient. To find these patterns in the data, the machine accesses mass amounts of data points and organizes them using statistical learning methods.

Statistical learning in machine learning consists of three major subsets: supervised, unsupervised, and semi-supervised learning.

Supervised learning requires a predetermined learning algorithm for the machine, which includes two essential parts: features and outcomes. The features (i.e. time spent at home), which are the predictors of the outcome (i.e. severity of depressive symptoms), are given to the machine as variables for it to then construct models for the most predictive outcomes.

Unsupervised learning, the second method, is similar to supervised learning in the sense that the machine is tasked with categorizing patients based on data. However, unsupervised learning does not possess programmed predictors; instead, the machine sifts through datasets in order to find its own parameters from which to then group patients. This process, known as clustering, requires the machine to perform dimensionality reduction, by which unlikely predictors are eliminated while the remaining ones are used to form relationships with patient outcomes. For example, no one may have programmed the computer to find a relationship between outgoing text messages and manic episodes through supervised learning, but in unsupervised learning the computer is able identify this unseen relationship. Of course many of these new relationships may not be useful in the clinic, as discussed later in this chapter. The psychiatrist must be wary that statistical significance is not the same as clinical significance.

The third and final type is semi-supervised learning, which combines the methods of supervised and unsupervised learning. In semi-supervised learning, only a small subset of the patients have a known outcome, and the rest of the patients are used to corroborate or change the initial relationship.

However, the brief above descriptions of machine learning and big data makes one critical assumption. In Desjardins the clinical world there is always missing or messy data. A patient may not recall how he reacted to a medication, may forget the name of his prior prescriber, is unsure if he was ever diagnosed with bipolar disorder, and so on. Likewise, big data itself is not perfect and often is messy and rife with missingness. Perhaps the geolocation sensor on the phone was not perfectly calibrated, turned off to save battery, or there was a mistake in the app recording that data. Thus prior to inputting into the machine, data may undergo cleaning, a process that removes subjects, or at least part of their data, from the dataset if their data is too messy or has too much missing. While superficially harmless, removing subjects from datasets has the potential to skew analysis, particularly if the removed subjects or data points are from the same group. Consider the simple example of patients with depression turning off their smartphone because they may not want to be contacted by others. This simple turning of an on/off switch means that no data is gathered and much is missing, when these data points could have provided valuable insight into the patient's symptoms.

As an alternative to cleaning messy data, missing data may be filled in through approximation using classical statistics, or statistical learning methods. The general linear regression model (GLM) is the simplest of statistical learning methods. GLM utilizes regression models to develop correlation coefficients between features and outcomes; however, this leads to the issue of overfitting in the case of high-dimensional datasets. Overfitting occurs when modeling of specific parameters fit too closely with a given dataset. Using a larger sample size combats this overfitting, by minimizing the effects of outliers and data that may be merely noise. Although increasing the volume of data will eliminate overfitting, the problem still lingers in high-dimensional research studies, in which the number of parameters is far greater than the number of observations.

Other techniques include elastic net models, a further extension of GLM, which use a large set of features to predict outcomes. Elastic nets will then filter through and select only the highest correlated predictors to incorporate into the final model. This is a manifestation of data reduction: the elimination of particular parameters in order to provide a highly correlated, accurate, and simplistic model for big datasets. Naïve-Bayes and Classification and Regression Trees (CART) are two additional methods of statistical learning. Naïve-Bayes is essentially an application of Bayes' Theorem, in that it classifies the likelihood of an event based on the value of one known variable. This variable is assumed to be independent of other parameters. CART, on the other hand, maps complex relationships between variables using a methodology similar to a flowchart. Data is first split up into categories, each represented as a leaf on the flowchart. The leaves are then connected to outcomes as well as other leaves, depending on the leaf's predictive capabilities.

Mental health research can produce significantly more powerful results when datasets from multiple sources are compiled into one and analyzed as a single dataset. Such analyses require large computational power, but are feasible, as demonstrated by several recent analysis into adolescent alcohol misuse for predicting current and future patterns (Whelan et al. 2014). In this study, an elastic net model was utilized to select for only the most impactful predictors of adolescent overconsumption. The resulting parameters included life experiences, neurobiological nuances, and overall personality of the adolescent. Moreover, the model provided regression values for each predictor, and based on these values, the model was able to remain accurate after application to a new data set. This dataset of new adolescents served to test the model, while the initial set of adolescents were used to first train and create the model. Typically, a dataset of K samples is subdivided, and all but one sample (K-1) is used to properly train and configure the model. Once the model is developed, the last sample is used for a test run, which hopefully results in a low prediction error. This process is repeated K times, each time resulting in a new set of K-1 subgroups for model training, while leaving the final subgroup for testing. This process is referred to as K-fold cross validation, and is widely utilized, including by studies presented later in this chapter.

3.1 Examples of Machine Learning Today in Psychiatry: Medication Selection

Despite tremendous recent increases in psychiatric knowledge of psychopharmacology, in today's world, finding the right medication for a patient can still be a process of trial and error. It can be hard to know a priori which patients will respond well to an antidepressant, and which may find the side effects too hard to bear or may simply not have an adequate response. While clinical experience is crucial in these decisions, machine learning offers both the patient and psychiatrist new information that may augment medication selection.

Matching the right antidepressant medication to the right patient is not trivial. Considering even a simplified definition of depression—meeting five of nine symptoms listed in the DSM-5 for 2 weeks—there are, in mathematical terms, nine choose five combinations of presenting symptoms, which is a total of 126. Biological evidence also suggests that there are subtypes of depression and that different types of depression respond better to certain medications than others. Machine learning can cluster patient symptoms into predictive subsets, from which psychiatrists can then prescribe the optimal prescriptions, targeted for a specific symptom within the patient's general depression.

The following examples (Chekroud et al. 2016, 2017) offers a model based on complete and prior collected data in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial, meaning the challenges of missingness and messiness are not addressed. This study used machine learning to create models to help identify whether a patient will benefit from a particular medication based entirely on the patient's unique background and clinical symptoms. Data from STAR*D (1949 usable patients) was used to construct a 25-predictor model to accurately estimate patient remission from the antidepressant citalopram, a serotonin reuptake inhibitor.

The three most predictive factors of non-remission were baseline depression severity (0.07793), restlessness during the last 7 days (0.06929), and lowered energy level over the last 7 days (0.05893). The most predictive characteristics for remission were having a job (−0.06946), years of education (−0.04712), and loss of insight of the depressive symptoms (−0.04625). The model was internally validated using the STAR*D dataset, resulting in an accuracy of 64.4%, higher in comparison to most predictive clinical models. The model was taken one step further and tried for validation on an external dataset, Combining Medications to Enhance Depression Outcomes (COMED). The COMED patient data was divided into three groups: escitalopram plus placebo, escitalopram plus buproprion, and venlafaxine plus mirtazapine. The predictive accuracy for each group was 59.6%, 59.7%, and 51.4%, respectively. Although the latter treatment group did not create statistically significant results ($p = 0.53$), the other two groups were significant, suggesting this model as promising for predicting medications that would best serve a patient.

The point of such a model is not to replace a psychiatrist, but rather to offer a new tool that may be useful in informed decision making regarding medication selection. Of course before any model can become widely adopted for clinical use, it also must be validated in real world conditions with real world data—which is often messy and missing to some degree. Such research efforts are currently underway and will continue to refine the field's knowledge about matching the right medication to the right patient. Chapter 4 will further discuss this topic.

3.2 Examples of Machine Learning Today in Psychiatry: Suicide Prediction

In the United States, suicide rates have risen to a 30 year high, tragically making suicide one of the top ten causes of death among those aged 10–64 (Curtin et al. 2016). Despite suicide awareness and outreach, this represents a 24% increase since 1999 (Tavernise 2016), and serves as an urgent call to action. While universal screening for suicide is a goal, it is not yet the standard, as implementation serves as the chief barrier. Patients and healthcare providers alike need a simple, yet effective means of quickly identifying risk factors for potential suicidal patients during preliminary evaluations. The grave disparity among research advances and current suicide rates has opened the door for machines learning and big data.

There is an urgent need for new tools to assist in predicting and preventing suicide. As alluded to above, while many area of health such as cancer and infectious diseases have experienced remarkable decreases in mortality rate as well as diagnostic and preventative advancements, suicide rates have increased. Current models to predict suicide risk have only little to moderate predictive utility, deeming previous suicidal attempts as the most common risk factor. Yet the fact that 60% of suicides are performed by those who have never made prior attempts reveals the weakness of these current models (Christensen et al. 2016).

New data and algorithms offer the potential to improve suicide prevention by extending monitoring beyond the clinic, with the ability to even respond to that data in real time. Interfacing with social media also provides machines a mechanism for identifying those at risk in real-time. In November 2017, Facebook announced it will be using artificial intelligence to monitor user's feeds in an attempt to predict who may be at risk (Zuckerberg 2017). While Facebook has not yet revealed what data they utilize and what algorithms they use, social media is becoming an active area of machine learning and mental health research. Other social media platforms are important targets as well for machine learning efforts. Machines can detect tweets and the changes within them that raise flags for suicide. However, further data mining must be performed in order to better characterize profiles of those at risk, and may soon include facial and voice characteristics as markers. By combining big data analysis by machines with individually gathered data streams, short-term risk factors can be quantified and identified almost immediately to provide needed support.

Medical records themselves also provide a source of data for machine learning techniques to offer new information relevant to suicide prevention. A case in point is a study from Montpellier University Hospital, where the records of 1009 hospitalized suicide attempters were analyzed in terms of several clinically-relevant parameters, including impulsiveness, mental disorders, and childhood trauma (Lopez-Castroman et al. 2016). This data was used for a hierarchical ascendant classification to create three homogeneous phenotypic clusters. The first cluster, labeled impulse ambivalent ($n = 604$), contained patients who were characterized by relatively non-lethal means of attempts and planning. The second cluster, well-planned ($n = 365$), had carefully planned attempts, more alcohol or drug abuse prior to the attempt, and had patients who employed more precautions to avoid interruptions. The third group, called frequent ($n = 40$), was the smallest, and included patients with more total attempts, being more serious and violent, and childhood abuse.

There were significant differences between each cluster for all analyzed variables ($p < 0.001$). Of the three clusters, clusters 1 and 2 were the most similar in terms of patient phenotype, so multivariate analysis with CART was performed on these two clusters. Cluster 3, on the other hand, was relatively distinct, possessing a female majority and a prevalent number of tobacco smokers, 90.0% and 80.6% of the cluster, respectively. This cluster was also prevalent in patients with anorexia nervosa (91.7%) and anxiety disorder (23.5%). Clustering is important as it offers clinically relevant and actionable insights that can be used to help clinicians identify those at high risk today. As more research continues, these models will continue to improve.

Clustering is not the only machine learning method useful for suicide prediction. Research groups across the world are actively investigating new data streams as well as new methods. For example, one group explored a neural network model for risk assessment of emergency room patients. The researchers created a software screening tool that 91% of patients found easy to complete, taking an average of 0:56 min, compared to nearly 8 min for a psychiatrist's brief evaluation (Desjardins et al. 2016). In preliminary testing, the neural network model was very accurate in predicting these new datasets, displaying a 91% accuracy in predicting psychiatrist's risk assessment and 89% for assessment of psychiatric intervention. This model provides the initial steps towards emulating the gold standard in evaluating suicide risk, but like all results, this model will need to be re-produced and run with new data to demonstrate its true clinical potential.

Related to suicide, non-suicidal self-injury (NSSI), most common among children and young adults, is deliberate self-injuring without suicidal intentions. The typical lifetime prevalence of NSSI in young adults and children is 13.9–21.4%, and the most common manifestation of NSSI is cutting (Plener et al. 2016). The internet is the most frequently used means by which NSSI health information is obtained. This information is sought not only by those who self-injure, but also the individuals who seek ways to help those who self-injure (*i.e.* parents and caregivers).

A recent study looked at the quality of the web resources for non-suicidal self-injury and highlighted the need for both mental health professionals and internet consumers to be cautious with what they read (Lewis et al. 2014). Researchers from the University of Guelph in Ontario, Canada searched 92 terms related to NSSI that resulted in 1000 Google hits or more. The first page of hits from these terms were evaluated, and the quality of health information on each website was evaluated using established guidelines from the Health On Net (HON) Foundation. They found that each of 340 healthcare websites contained an average of 1.44 ± 1.18 (mean $\pm$ SD) myths about NSSI. The most prominent myths were associating NSSI with a mental disorder (49.3%), abuse (40%), or that women are more likely to self-injure (37%). The mean quality of healthcare information in terms of HON criteria was 3.49 ± 1.40, while only one website received a perfect score of 7. Moreover, very few of these websites were credible, as only 9.6% were endorsed by health (*i.e.* hospitals) and/or academic institutes.

These results are concerning for not only patients but also machine learning efforts. Without proper collaborations between psychiatry and data science fields, it is easy to see how incorrect information could easily be accessed and programmed into machine learning algorithms. The advantage of machine learning tools is they can be delivered at scale to the population, but this is likewise their weakness, as incorrect or harmful information can be similarly scaled as well. Chapter 5 will further discuss this topic.

3.3 Examples of Machine Learning Today in Psychiatry: Symptom/Outcome Monitoring

Machine learning methods can do more than predict risk of self-harm or suicide; they can also help guide treatment decisions such as identifying the right medication for the right patient. For example, one third of patients suffering from Major Depressive Disorder (MDD) do not react adequately to treatment. Much effort has been put into characterizing treatment-resistant depression (TRD), defined as an inability to achieve at least 50% reduction in depression (McIntyre 2014). To investigate the potential of machine learning methods, 480 patients with TRD were studied to identify predictors for ineffective treatments (Kautzky et al. 2017). This patient cohort was taken from the Group for the Study of Resistant Depression (GSRD), a multinational European research consortium. A machine learning model was created using 48 predictors from clinical (change of sleep, suicidality), sociodemographic, and psychosocial (marital status, education) patient aspects. A Random Forest algorithm was used for model development, and results demonstrated that using all 48 predictors resulted in an accuracy of 73.7% for resistance and 85.0% for remission. However, single predictors resulted in an odds ratio of only 1.5; even the strongest single predictor, time between first and last depressive episodes, resulted in merely 56% and 60% accuracy for resistance and remission, respectively. Likewise, clinical predictions made by psychiatrists for treatment resistance are not dictated by a single parameter, but rather by considering many factors of the patient. The clinical line of thinking is reflected by this machine, in that more parameters create a better diagnosis, and may help optimize treatments in the clinic.

Machines do not need to rely solely on previously collected data, as they have demonstrated the ability to learn and make accurate predictions from real-time data. Ecological momentary assessment (EMA) is an important tool used by healthcare professionals to evaluate the mental state of patients throughout their daily activities. However, EMA has typically been administered through self-report questionnaires, leading to response bias and subjectivity. In this era of increasingly ubiquitous smartphones, EMA can be easily conducted via phone-based sensors and surveys, which are becoming more prevalent in psychiatry research. With their myriad of sensors, such as GPS, accelerometer, and ambient light, smartphones can provide real-time information about patient environment. The social logs of smartphones, such as call/text logs and social media profiles, also offer clues about social interactions and communication patterns (Torous et al. 2016).

A study by Asselbergs et al. offered new insights into mental health by demonstrating the potential of real-time phone data when combined with machine learning methods (Asselbergs et al. 2016). A mobile phone app was implemented on 27 Dutch university students to monitor their moods through proxies of social activity, physical activity, and general phone activity. The data was used for predictive modeling, including personalized predictive models for each participant based on individual data from their previous days. A regression algorithm selected and weighed variables into subsets to predict self-monitored mood. The eMate mobile app prompted subjects to evaluate their mood at five set points per day. Two-dimensional and one-dimensional mood evaluations were used, the latter of which simply asked the subject to rate his/her mood on a 10-point scale. The two-dimensional scale, however, used two levels of valence: positive and negative affect.

The unobtrusive, real-time data aspect for the study was collected using iYouVU, a faceless mobile app founded on Funf open-sensing framework. This app collects pre-determined sensor data and app logs, which are then sent over Wi-Fi to a central server. Daily averages of EMA, both one and two-dimensional, were averaged and scaled to each subject. The unobtrusive data included total number of times screen was turned on/off, and call and SMS text message frequency to top five contacts.

The personalized mood prediction machines for each student were created using forward stepwise regression (FSR), in which relevant variables for predicting mood are selected sequentially as more data is accumulated. To maximize predictive variables while avoiding overfitting, only eight variables (the number of data points (42) divided by 5) were used in each student's model. The first FSR was stepAIC procedure, which selects variables based on Akaike information criteria (Akaike 1974). The second FSR method was stepCV procedure, by which variables are selected based on their ability to lower cross-validated mean square error between the phone-collected scores and cross-validated predicted scores. Thus a variable is added to the model unless it increases the mean squared error. The cross validation was performed using leave-one-out cross validation (LOOCV) by predicting residual sum of squares for every model run. The predictive performance of both FSR variants was evaluated using LOOCV, comparing the observed mood rating through the mobile phone with that predicted by the personalized FSR models. The result were relatively underwhelming, as the proportion of correct predictions was 55–76% lower compared to two previously published naive models. This result demonstrates that machine learning methods are not always better than simple baseline models.

However, sometimes machine learning does produce results that are not seen with simpler models or clinical observations alone. A case in point is a study involving speech data and schizophrenia (Bedi et al. 2015). Disorganized speech is often an early sign of prodromal schizophrenia, and a novel study analyzed speech data with machine learning in order to accurately predict schizophrenia conversion among youths with prodromal symptoms. Utilizing latent semantic analysis (LSA), an algorithm that utilizes multiple dimensions of associative analysis of semantic speech structure, researchers studied speech data for over 2.5 years in those at risk for becoming schizophrenic. LSA assumes that the meaning of a word is based on its relation to every other word in the language; words that recur together many times in a transcript can then be indexed in terms of their semantic similarity. A machine learning algorithm was trained using the semantic vectors generated from LSA from those who developed psychosis (CHR+) and those who did not (CHR-) upon follow-up. The machine used a cross-validated classifier, analogous to K-fold cross validation, to learn the speech features which differentiated CHR+ from CHR- participants. Results demonstrated 100% accuracy in predicting psychosis for each participant within the sample used to generate the machine. Not surprisingly, this perfect result is significantly greater than the predictive capability of clinical classifiers from the SIPS/SOPS evaluation (79%). However, the machine was not externally validated on a new dataset different from the initial one used for model fabrication. The true predictive capability of the model is likely lower than the apparent perfect accuracy. Although, automated analysis clearly demonstrates the potential to outperform standard clinical ratings for predicting clinical onset, as machines can provide insight on minute semantic difference that the latter cannot sense.

3.4 Next Step and the Future of Machine Learning in Psychiatry

3.4.1 Outsource Simple Tasks to Machines

While machine learning will not replace psychiatrists, it can help make their work more efficient. Machines have the ability to fully automate generic tasks within psychiatry, such as symptom severity screening. At the time of this writing, The National Health Service in the United Kingdom is assessing an artificial intelligence app, developed by the company Babylon, on nearly 1.2 million users in London, England (Burgess 2017). Rather than have citizens call the non-emergency health service phone line, which is typically understaffed and run by non-medically trained individuals, the app provides a promising alternative through a virtual physician evaluation. This app possesses a database of symptoms which is utilized by the app's chatbot to help patients instantly find out the urgency of their health issues. When presented with a serious case, as assessed by the machine, the chatbot connects patients directly to a physician. The app has demonstrated the ability to assess patient illness in a more accurate manner than phone line operators, while also saving government resources.

3.4.2 Population Level Risk Stratification and New Disease Models

Machine learning methods can also help psychiatry with population level risk prediction. Mental health disorders are typically predicted with machines using single time point cross-sectional variables, most often clinical aspects from initial evaluations. These machines may be compromised by their inability to account for the dynamic nature of symptoms. Thus, predictive modeling can benefit by assessing the micro-level (momentarily/daily) and macro-level (monthly/yearly) dynamic factors that impact the course of psychiatric illnesses (Nelson et al. 2017).

The same models can also offer new ways to conceptualize disease. Dynamic Systems Theory proposes that complex systems consist of sub-systems that are interconnected and highly correlative, while other sub-systems possess diverse aspects that are only loosely related. Distinguishing the sub-systems that are correlative has provided a means for researchers to accurately model aspects of mental illness, one of which is through the EMA. As previously mentioned, this assessment evaluates an individual's mood at many points in a day to detect shifts from baseline. Such micro-level assessment lends to correlations between depressive symptoms and subtle changes in emotional state. On the other hand, recording macro-level changes is done through joint modeling of event outcomes and time-dependent predictors.

These complex systems are also the crux of Network Theory. By using Network Theory, we assume mental disorders are a result of complex relationships between the biological, psychological, and social aspect of our lives. Each system is triggered by the other, resulting in an overall system that is characterized by positive feedback, forming a type of loop, whereby the body may be stuck in a continuous cycle of particular symptoms. These symptoms can sometimes be malicious, which can then be classified as states of mental disorder. Similarly, Instability Mechanisms convey that mental disorders are the result of amplifying minor health issues by feedback loops in the body. What initially seems like a commonplace affect, such as disliking of cramped rooms, can exacerbate into claustrophobia for some individuals if the body is continuously running the loops.

3.4.3 Better Use of Medical Records Data

Machine learning can help not only in better characterizing psychiatric illness, but also in improving the delivery of psychiatric care. Though clinical assessment remains the paradigm for patients seeking diagnosis, there is increasing interest in using retrospective patient records as big datasets. Retrospective data has gained popularity due to its ability to simplify and standardize medicine for more precise results. Electronic health records (EHRs) provide a means of retrospectively phenotyping patients, and correlating their characteristics, whether demographic or diagnostic, to treatment outcomes. But using EHR data can be difficult and combining EHR data across multiple clinics and health systems is a serious challenge due to lack of interoperability. The *green button* movement seeks to make it easier to operationalize EHR data and utilize it in novel ways, such as to learn how a particular patient may respond to treatment compared to others with a similar presentation (Longhurst et al. 2014). This process of screening EHRs was used to change the conventional policy for setting alarm alert limits, which is typically age-based. Lucile Packard Children's Hospital of Stanford operationalized 1000 of EHRs to create a novel distribution of alarm limits for children, based on their heart rate distribution rather than age. This nascent implementation of personalized database data has helped provide more accurate care tailored for each of the pediatric patients.

Physicians have also begun to take initiative in promoting collaboration between researchers in the digital health field through secure sharing of health records and data. Dr. Ashish Atreja, Chief Technology Officer at Icahn School of Medicine at Mount Sinai, has facilitated digital health data sharing among physicians through the digital platform NODE Health (Comstock 2017). This initiative allows for secure sharing of clinical data in efforts of providing a wide range of researchers with patient data that would otherwise be unattainable for them. The researchers who take part in NODE Health are able to foster multi-site projects, rather than conduct costly duplicate studies, because the data is readily available for sharing.

3.5 What are the Next Steps to Realize that Future

3.5.1 A Need for High Quality Data

Despite the early successes and continued promises of machine learning methods for mental health, there is also need for caution. One area regards bias that may inadvertently be scaled up by these methods if the wrong types of data are used to build models. For example, collecting and processing information through social media poses a challenge, as the information is highly skewed by search methods. There have been few studies that address search filters, combinations of keywords and search rules, in their entirety. In a similar vein, very few research groups provide the proportion of usable data that is collected by their filters. Bias in search filters can skew data, which precludes generalizable results. The proportion of quality data that results from search filters must be objectified and characterized in relation to a standard benchmark. Such a benchmark has been aimed to be created by a recent study, which aimed to provide standards for retrieval precision and recall (Kim et al. 2016)

Twitter, for example, is one of the most prevalent social media platforms used to gather data, largely in part due to its high volume. When obtaining data from Twitter, researchers must be aware of colloquial slang, abbreviated words (due to the limit on characters per Tweet), and use of hashtags. Experts in the field of study should be utilized for assistance in filter selection. Signal to noise ratio is also imperative and keywords with a low ratio should be excluded. This threshold ratio depends on the study, but one benchmark to discard tweets is those that result in less than ten tweets in a month or return less than 30% of relevant tweets. The search rules can use Boolean operators, such as AND, NOT, OR, as well as data pre-processing techniques like n-grams and proximity operators.

3.5.2 A Need for Good and (New) Study Design

New tools like machine learning may also require new clinical study designs to make the most efficient use of the resulting data. Ensuring that studies are designed to have not only appropriate controls but also appropriate training and testing datasets must be considered when seeking to utilize supervised machine learning methods. When aiming to utilize unstructured methods, it is useful to consider how data may cluster and whether the outcome metric is suitable. Close partnerships with data scientists are critical to ensure that statistical methods are employed correctly and that spurious correlations or findings are avoided (Ioannidis 2016). Health studies can also learn from the software paradigm of agile development, in that iterative and rapid studies may prove of more value to single long studies that are committed to one particular technology or method. This concept, sometimes referred to as Agile Science, offers an early roadmap of a new way to envision and execute clinical studies (Hekler et al. 2016)

3.5.3 A Need to Realize and Plan for Unintended Consequence

Though machine learning demonstrates the ability to improve the medical field through means such as increased predictive accuracy, there are also unintended side effects. When novel technologies are introduced to healthcare, some aspects of medicine can suffer. One major concern is the over reliance on machine learning to detect symptoms and proposed treatments for patients. This can lead to deskilling, decline in performance when a task becomes automated, which can result in drastic deficits if the technology is removed. Mammogram readers, for example, experienced a 14% decrease in sensing diagnostic markers on images with computer-aided detection (Cabitza et al. 2017).

It is also difficult to fully program machines to consider the clinical parameters that may be only detectable by a holistic, human evaluation. The human experience can sense psychological, social, and relational issues, aspects which must then be quantitatively programmed into data that is interpretable by a machine. Evidently, the problem lies in coding these subtle characteristics that only the human senses are conditioned to perceive. This also encompasses fundamental guidelines of healthcare, which can be overlooked in machines because they are merely taught to recognize patterns in data. For example, a risk prediction machine was created for 14,199 patients with pneumonia, and the machine found that those with both asthma and pneumonia had a lower mortality risk than patients solely with pneumonia (Cabitza et al. 2017). Clinicians were surprised that asthma could be a protective agent, and began questioning the legitimacy of the machine. However, the clinicians could not find a problem with the machine, as it had merely done its job as it had been programmed to do. The issue lied in the coded parameters and data. Patients with both asthma and pneumonia were assigned to intensive care units, which resulted in a 50% reduction in mortality risk than patients with solely pneumonia, who were typically not admitted to intensive care. Contextual factors such as the difference in hospital unit are crucial for accurate modeling, though are difficult to recognize and then accurately encode into machines.

3.6 Conclusion

The future is bright for machine learning in mental health. In recent years, researchers have published numerous studies showing the potential of these methods for predicting suicide, matching patients to the right medicine, increasing efficiency of care, and even monitoring patients outside of the hospital with smartphones and sensors. However, it is worth noting that much of this research has yet to be reproduced or deployed at scale in healthcare systems. Given the nascence of machine learning applied towards mental health, compounded by the challenge of quantifying human behavior, it is not surprising that the field is still exploring its role and potential. But given the direct errors as well as unintended consequences, a cautious approach is warranted. Nonetheless, as the diverse methods and applications of this chapter underscores, the field is rapidly progressing and we expect the impact and role of machine learning in mental health to only continue to grow.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19(6):716–723

Asselbergs J et al (2016) Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 18(3):e72

Bedi G et al (2015) Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophrenia 1(1):15030

Burgess M (2017) The NHS is trialling an AI chatbot to answer your medical questions. Wired. Available at http://www.wired.co.uk/article/babylon-nhs-chatbot-app/

Cabitza F et al (2017) Unintended consequences of machine learning in medicine. JAMA 318(6):517–518

Curtin S et al (2016) Increase in suicide in the United States, 1999–2014. National Center for Health Statistics, Hyattsville Brief No. 241. Available online at https://www.cdc.gov/nchs/products/databriefs/db241.htm

Chekroud AM et al (2016) Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 3(3):243–250

Chekroud AM et al (2017) Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. JAMA Psychiat 74(4):370–378

Christensen H et al (2016) Changing the direction of suicide prevention research: a necessity for true population impact. JAMA Psychiat 73(5):435–436

Comstock J (2017) Mount Sinai launches data sharing initiative for digital health pilots. MobiHealthNews, Portland Available at http://mobihealthnews.com/content/mount-sinai-launches-data-sharing-initiative-digital-health-pilots

Desjardins I et al (2016) Suicide risk assessment in hospitals: an expert system-based triage tool. J Clin Psychiatry 77(7):e874–e882

Hekler EB et al (2016) Agile science: creating useful products for behavior change in the real world. Transl Behav Med 6(2):317–328

Ioannidis JP (2016) Why most clinical research is not useful. PLoS Med 13(6):e1002049

Kautzky A et al (2017) A new prediction model for evaluating treatment-resistant depression. J Clin Psychiatry 78(2):215–222

Kim Y et al (2016) Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. J Med Internet Res 18(2):e41

Lewis SP et al (2014) Googling self-injury: the state of health information obtained through online searches for self-injury. JAMA Pediatr 168(5):443–449

Longhurst C et al (2014) A 'green button' for using aggregate patient data at the point of care. Health Aff 33(7):1229–1235

Lopez-Castroman J et al (2016) Clustering suicide attempters: impulsive-ambivalent, well-planned, or frequent. J Clin Psychiatry 77(6):e711–e718

McIntyre RS (2014) Treatment-resistant depression: definitions, review of the evidence, and algorithmic approach. J Affect Disord 156:1–7

Nelson B et al (2017) Moving from static to dynamic models of the onset of mental disorder: a review. JAMA Psychiat 74(5):528–534

Plener PL et al (2016) The prevalence of nonsuicidal self-injury (NSSI) in a representative sample of the German population. BMC Psychiatry 16(1):353

Whelan R et al (2014) Neuropsychosocial profiles of current and future adolescent alcohol misusers. Nature 512(7513):185

Tavernise S (2016) US suicide rate surges to a 30-year high. New York Times, New York. Available online at http://www.nytimes.com/2016/04/22/health/us-suicide-rate-surges-to-a-30-year-high.html?_r=0

Torous J et al (2016) New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Mental Health 3(2):e16

Zuckerberg M (2017) Here's a good use of AI: helping prevent. Suicide. Available online at https://www.facebook.com/zuck/posts/10104242660091961