May 2021

Predicting Factors of Re-Hospitalization After Medically Managed Intensive Inpatient Services in Opioid Use Disorder

Brian Kay
University of Wisconsin-Milwaukee

PREDICTING FACTORS OF RE-HOSPITALIZATION AFTER MEDICALLY MANAGED INTENSIVE INPATIENT SERVICES IN OPIOID USE DISORDER

by

Brian Kay

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Biomedical and Health Informatics

at

The University of Wisconsin-Milwaukee

May 2021

ABSTRACT

PREDICTING FACTORS OF RE-HOSPITALIZATION AFTER MEDICALLY MANAGED DETOXIFICATION IN OPIOID USE DISORDER

by

Brian Kay

The University of Wisconsin-Milwaukee, 2021
Under the Supervision of Professor Jake Luo

Introduction
Opioid use disorder has continued to rise in prevalence across the United States, with an estimated 2.5 million Americans ailing from the condition (NIDA, 2020). Medically managed detoxification incurs substantial costs and, when used independently, may not be effective in preventing relapse (Kosten & Baxter, 2019). While numerous studies have focused on predicting the factors of developing opioid use disorder, few have identified predictors of readmission to medically managed inpatient level of care. Utilizing a high-fidelity dataset from a large multi-site behavioral health hospital, these predictors are explored.

Methods
Patients diagnosed with Opioid Use Disorder and hospitalized in the inpatient level of care were analyzed to identify readmission predictors. Factors include patient demographics, patient-reported outcome measures, and post-discharge treatment interventions. Patients re-hospitalized to the inpatient level of care were binary labeled in the dataset, and various machine learning algorithms were tested and evaluated for performance. Methods include random forest, gradient boosting, and deep learning techniques. Evaluation statistics include specificity, accuracy, precision, and Matthew's Coefficient.

Results
Overall, there was a wide variation in correctly predicting the class of patients that would readmit to a medically managed level of inpatient detoxification. Out of the six models evaluated, three of the six did not converge, thus not producing a viable feature ranking. However, of the other three models that did converge, the deep learning model produced almost perfect classification, producing an accuracy of 98%. AdaBoost and the logistic regression model produced an accuracy of 97% and 61%, respectively. Each of these models produced a similar set of features that were important to predicting which patient profile would readmit.

Conclusions
The results indicate that overall reduction in the Quick Inventory of Depressive Symptomology, discharge disposition, age, length of stay, and a patient's total number of diagnoses were important features at predicting readmission. Additionally, deep learning algorithms vastly outperformed other machine learning algorithms and traditional statistical methods.

DEDICATION

To my wife and family for endlessly supporting my endeavors to grow my knowledge.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

COVID-19	Coronavirus 2019
OUD	Opioid Use Disorder
FDA	Food and Drug Association
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
ASAM	American Society of Addiction Medicine
EHR	Electronic Health Record
QLESQ	Quality-of-life Enjoyment and Satisfaction Scale
QIDS	Quick Inventory of Depressive Symptomology
PHQ-9	Patient Health Questionaire- 9
ICD-10	International Classification of Diseases 10^{th} edition
HIPAA	Health Insurance Portability and Accountability Act
ROAS	Rogers Outcome Assessment System
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

LIST OF EQUATIONS

Introduction

Opioid use disorder has continued to rise in prevalence across the United States, with an estimated 2.5 million Americans ailing from the disease in 2020 alone (NIDA, 2020). In addition to the Coronavirus 2019 (COVID-19) pandemic proliferating across the United States, there are indications that there will be a surge of opioid use throughout the pandemic and afterward (Wakeman et al., 2020). With cases continuing to increase, not only may there be a lack of treatment facilities to handle the influx of patients, but an economic burden on individuals and healthcare payers if the cost of care is not effectively managed. With medically managed detoxification incurring substantial costs for a single episode of care, and when used asynchronously, it may not effectively prevent relapse (Kosten & Baxter, 2019). One strategy to reduce these costs would be to reduce the probability that an individual readmits to the same care post-discharge level.

Previously, insights for a whole episode of care were difficult to ascertain; however, it is now more plausible through the increased systematic collection of healthcare data and advanced data analytics. Using machine learning techniques, researchers can better identify factors that precipitate health outcome states leading to better long-term health conditions. In addictions research, a myriad studies have focused on predicting the characteristics of developing opioid use disorder; however, few have identified predictors of readmission to medically managed withdrawal at a medically managed inpatient level of care. The lack of information represents a gap in the literature and valuable insights into individuals suffering from opioid use disorder while enhancing long-term positive outcomes.

Utilizing a high-fidelity dataset from a large multi-site behavioral health hospital, these predictors are explored. This work seeks to apply advanced data analytic techniques such as

supervised machine learning to effectively predict which factors may lead to more intensive levels of care in individuals with opioid use disorder. Furthermore, these predictions can serve as practical decision support to reduce the overall spend associated with opioid use disorder. Finally, this work seeks to evaluate different machine learning types and advanced analytics to garner insight into predictors of readmission. Further evaluating the use and application of machine learning algorithms in healthcare, specifically mental and behavioral health.

Epidemiological perspective of opioid use disorder

Opioids are natural or synthetic chemicals that interact with particular parts of one's brain chemistry. Opioids are known for reducing the feeling of pain in one's body, hence the use traditionally as pain relievers (Rosenblum et al., 2008). Opioids, in their chemical form, have been used extensively post-surgery, through cancer treatment, or in palliative care. While effective for reducing pain in many individuals, opioids can produce many side effects, such as nausea, euphoria, and drowsiness. Many times, these side effects are managed using added prescription medications. Opioids as a drug class are extensive and varied, with many opioids having differences in their chemical composition. The chemical composition differences can also affect the potency of the opioid, producing a range of effects— with one of the most potent opioids being Fentanyl. Overarchingly, opioids fall into the Schedule II drug class as characterized by the Federal Drug Administration (FDA). Meaning they have a high potential of abuse or physical dependence.

Even though opioids can be used therapeutically in the aforementioned cases, opioids have highly addictive properties. Many individuals find themselves addicted to opioids after being legally prescribed opioids for therapeutic intent. When an individual uses opioids in a

problematic fashion, they may be diagnosed with opioid use disorder. Opioid use disorder is typically a lifelong chronic disorder and is managed using treatment, medications, and behavioral therapy. According to the Diagnostic and Statistical Manual of Mental Health Disorder 5th Edition (DSM-5), the gold standard for diagnosing mental and behavioral disorders (American Psychiatric Association, 2013). An individual need to meet two of the following criteria in order to be diagnosed with opioid use disorder:

- Taking larger amounts or taking drugs over a longer period than intended.
- Persistent desire or unsuccessful efforts to cut down or control opioid use.
- Spending a great deal of time obtaining or using the opioid or recovering from its effects.
- Craving, or a strong desire or urge to use opioids
- Problems fulfilling obligations at work, school, or home.
- Continued opioid use despite having recurring social or interpersonal problems.
- Giving up or reducing activities because of opioid use.
- Using opioids in physically hazardous situations.
- Continued opioid use despite ongoing physical or psychological problem likely to have been caused or worsened by opioids.
- Tolerance (i.e., need for increased amounts or diminished effect with continued use of the same amount)
- Experiencing withdrawal (opioid withdrawal syndrome) or taking opioids or a closely related substance) to relieve or avoid withdrawal symptoms.

The DSM-5 allows for clear criteria of individuals who meet the criteria for opioid use disorder. Furthermore, these diagnostic criteria are considered the gold standard, which allows a diagnosis to be consistent between location and provider.

Treatment of Opioid Use Disorder

The treatment of opioid use disorder (OUD) differs by level of treatment intensity and may include different levels of care. The American Society of Addiction Medicine (ASAM) has defined criteria for which level of care an individual should be placed to manage their treatment of opioid use disorder. This criterion is referred to ASAM criteria. ASAM criteria are multi-dimensional which seek to have a holistic assessment of the intensity of services required by an individual. The ASAM criteria are composed of six distinct dimensions, which include the following, acute intoxication and/ withdrawal potential, biomedical conditions and complications, emotional, behavioral, or cognitive conditions, readiness to change, relapse, continued use, or continued problem potential, and recovery living environment. The severity and ranking of each of these levels drive treatment intensity, and individuals should be placed. With “level 1" being outpatient services and “level 4" medically managed intensive inpatient services (*About the ASAM Criteria*).

When an individual is classified at “level 4" for ASAM criteria, they need the most intensive addiction treatment. ASAM “level 4” is a medically managed intensive inpatient services are a 24-hour treatment that offers nursing care, a daily session with a medical provider, and at least 16 hours of counseling. The intensity of this level of care is intended to help manage much of the medical components associated with detoxification from opioids. The length of stay varies due to the severity of one's symptoms but typically ranges from three days to seven days.

Given that an individual is in a monitored facility, there is a multitude of information collected through their treatment, including a comprehensive psychosocial assessment, a thorough nursing assessment, and a social and familial assessment. These assessments supply a base of information for the treatment team to properly treat all aspects of one's condition.

This includes comprehensive demographic information, diagnosis following the DSM-5, medication history, current medications, and socioeconomic traits. Additionally, much of the data is temporal, with values changing throughout their treatment in the intensive level of care.

The cost of medically managed detoxification may incur high costs throughout the course of one's treatment. In an ASAM “level 4" inpatient, costs per stay are significant, with the average stay totaling approximately $4,500 (Substance Abuse and Mental Health Services Administration 2006). With a high rate of relapse for OUD, it is possible that over the course of one's lifetime, they accumulate significant costs associated with readmitting to medically managed detoxification.

The use of FDA-approved recovery medications is also a significant next step in the long-term treatment of OUD and is becoming a critical component of medically managed detoxification. Currently, in the United States, there are three FDA approved recovery medications, Methadone, buprenorphine, and naltrexone. All of the medications work by targeting the μ-opioid receptor. Both Methadone and buprenorphine are considered agonists in the treatment of OUD, which reduce cravings for use and are typically prescribed for long-term use. Naltrexone, on the other hand, is an antagonist, extinguishing the effects of opioids rapidly. Due to this, naltrexone is typically one of the first responses to overdose, as it is broadly used in emergency services. With Methadone and buprenorphine increasing in utilization for long-term

treatment of OUD, many medically managed detoxification programs are initiating these medications during a patient's treatment (Volkow, 2018).

Measurement-based care in Behavioral Health

An essential step at improving the efficacy of behavioral health treatment and increasing data collection in behavioral health is through the use of measurement-based care. The use of measurement-based care in the psychiatric setting is considered a best practice but is limited in the field due to implementation challenges and providers' willingness to incorporate measurement-based care in their workflow. It is estimated that only 17.9% of psychiatrists and 11.1% of psychologists in the United States routinely administer symptom rating scales to their patients (Zimmerman & McGlinchey, 2008).

Measurement-based care is the systematic administration of symptom rating scales to drive clinical decision-making at the level of the individual patient. It is ultimately constructed to increase efficiency, accuracy, and consistency of symptom assessment and allow the provider to detect signals of non-response better. These clinical measures are well-validated and psychometrically sound, allowing the identification of behavioral health conditions and tracking treatment process over time. These measures may also be employed to assess if treatment interventions effectively treat a patient's symptoms. Providers of behavioral health treatment have constructed protocols to administer these measures systematically, with patients receiving a battery of assessments at admission, at defined intervals within treatment, and discharge. This systematic remeasurement is critical for guiding interventions, and the measures utilized have been designed to be clinically actionable. This provides direct benefit to patients who are completing these measures as it guides the interventions in their treatment. It also allows clinicians to detect better if a patient is experiencing worsening symptoms; typically, mental

health providers only detect symptom regression 21.4% in their patients without the aid of clinical assessments (Hatfield et al., 2010). Additionally, measurement-based care sets the foundation for robust outcome reporting, the ability to aggregate patients' treatment response not only over a single episode of time but aggregated across a location or treatment service.

Typically, measures administered in a psychiatric population may include a quality-of-life measure and depression inventories. Given that there is a high comorbidity of depression with opioid use disorder, many inpatient psychiatric providers choose to measure this construct over treatment systematically. The quality-of-life enjoyment and satisfaction scale (QLESQ) is a 16-item measure that quantifies an individual's quality of life in 16 different domains. The measure is self-report, psychometrically valid, and short of administering (Endicott et al., 1993). The quick inventory of depressive symptomology (QIDS) is a 16-item depression inventory that is self-reported. The inventory allows the patient to endorse symptoms of hopelessness, weight loss, fatigue, sleep disturbance, and suicidal ideation. This assessment is psychometrically sound, easy to administer, and as well published in the field. Additionally, the assessment allows an individual to rate their depression symptoms over time, quantifying depression symptomology over time (Rush et al., 2003).

Data analytics in healthcare

With the increase of data being collected through the use of EHR's and the ability for more advanced data analysis to better predict disease, outcomes, and the optimal treatments, the use of data analytics has proliferated in healthcare. Additionally, the amount of data within the healthcare setting has continued to increase exponentially year over year. In 2013, it was estimated that 153 exabytes of health care data were collected, with 2021 collecting more than

2,314 exabytes (Banks, 2020). Furthermore, it is anticipated that in 2022, there will be a zettabyte of healthcare data collected. One zettabyte of data is equivalent to 152 million years of Ultra High Definition 8k video. This the sheer volume of data collected and large datasets amassing, healthcare is primed for applying advanced analytic techniques to aid in the analysis and uncover new insights at a rate that has not been encountered before.

The application of data analytics can be grounded in the types of analytics. There are three major types of analytic categories: descriptive analytics, predictive analytics, and prescriptive analytics. Where descriptive analytics is inferencing information from a sample or a dataset, these inferences typically summarize the data through means, medians, or mode to provide a representation of the sample or a larger population. Predictive analytics is the extraction of information from a dataset and predicting trends and behavioral patterns. This helps understand the historical patterns and perspectives that lie in datasets and how they can predict future events. Predictive analytics can use various techniques from regression analysis, machine learning, and time series forecasting. Finally, prescriptive analytics aim to predict and determine the optimal decision based on a set of business rules. Prescriptive analytics is considered a category of advanced analytics given the level of sophistication and application.

The aforementioned analytic categories are prevalent in healthcare, however, in different frequencies. Where descriptive analytics is widespread, through day-to-day healthcare operations, predictive and prescriptive analytics are seen less commonly. However, there has been an increase in prescriptive analytics, particularly in the radiological space, in recent years to aid in the identification of diseases (Choy et al., 2018). Despite the increase of these analytics in medical-surgical applications in the healthcare sector, there have been limited applications in the mental and behavioral. This may be due to less structured information, or data points with less

fidelity, unlike lab values and biometric data. Despite these limitations, the ability for high fidelity datasets in mental and behavioral health is becoming more of a reality due to the treatment becoming more structured, evidence-based, and utilizing measurement-based care.

Machine learning in Behavioral Health

Despite the increase use of machine learning techniques in healthcare, there has been minimal use of supervised machine learning techniques in the behavioral setting. With most applications seeking to either predict outcome state or quantify the risk of a rare event in an individual. In behavioral health, some of the applications of supervised machine learning techniques have been applied to predict suicide risk (Roy et al., 2020) and predict the onset of opioid use disorder (Ellis et al., 2019). Many of these applications follow the same methods of supervised learning. The "learning" ascribed to machine learning techniques allows the computer to take part of the dataset and learn patterns, and then the learned patterns are validated on another part of the data to understand accuracy, sensitivity, and specificity. Typically, the original dataset is split into a 70% and 30% ratio, where 70% of the cases are training, and 30% of the cases are validation, where the output of the learned algorithm can be checked for validity. To reduce potential bias in the dataset, cross-fold validation is applied. Cross-fold validation is using the dataset and splitting it in diverse ways to find generalizable variables. Each time the dataset is split, iteratively identifies generalizable variables, thus reducing bias incorporated into a potential decision tree. The times that the cross-folds are performed varies from 10 times up to 24 times (Kohavi, 1995). Differences from each of the runs are then generalized to produce an ensemble-based model. Typical outputs from machine learning models are called decision trees. Decision trees consolidate the patterns from the multiple learning runs into a set of rules. An

added advantage to producing decision tree-based models is they can identify linear relationships. In healthcare, relationships between outcome variables are often non-linear in fashion, allowing machine learning to take advantage of these relationships. Utilizing these advanced analytics, this study seeks to understand the predictive factors associated with readmission to medically managed inpatient services for the treatment of opioid use disorder by leveraging decision trees that allow an end-user to understand the associated factors with the outcome state.

For example, Hatton et al. (2019) attempted to predict depression outcomes in a geriatric population utilizing the Patient Health Questionnaire (PHQ-9) (Kroenke et al., 2001). Utilizing a sample of 200 individuals who were "Older adults," the researchers utilized patient-reported outcome measures to classify if individuals would maintain treatment gains post-discharge.

The authors' utilized machine learning techniques to predict if an individual who was part of the clinical trial would maintain their gains 12-month post-discharge from the trial. The researchers utilized bootstrapping to impute missing data in the dataset utilize the expectation-maximization method. The dataset was then split at 60:40, with 60% being training and 40% being a validation of the algorithm. The primary algorithm utilized was gradient boosting, where the outcome variable was a binary PHQ-9 score over or under 10. This indicated if the individual was still endorsing depressive symptomology. The researchers also used logistic regression to predict the same outcome variable and then assessed the machine learning algorithm's predictive power.

Machine learning has also been utilized to understand and predict who is at risk for completing suicide and what factors are associated with this risk. Current suicide risk prediction today is extremely antiqued. Typically, an individual completes a structured interview with a

trained clinician. These interviews are lengthy, and due to the training required to administer the interview, not easily accessible to the public. Additionally, a 30-year meta-analysis has shown that these aforementioned interviews have low power in predicting a suicide attempt. The field of psychology has historically utilized less than advanced statistical techniques. As this is essentially a classification problem, machine learning was an appropriate methodology to predict attempts. Utilizing medical records from 5,543 adult patients at Vanderbilt University Medical Center, the authors coded all E95x, International Classification of Diseases 10th Edition (ICD-10) codes.

The author's utilized machine learning techniques in order to predict suicide attempts temporally. They utilized individuals with a previous attempt and a control group of adults with no prior documented history of suicide attempts. An ensemble-based learning method, random forests, was utilized to predict the attempts. Decision trees are generated via recursive sampling techniques of the predictor data. In this study, 500 trees were generated, and the risk estimated was determined based on the proportion of trees that predicted the correct outcome. The authors utilized standard demographic information from the medical record in their data. The authors published 92% accuracy of the prediction, with a low discrepancy between precision and recall.

Both of these examples highlight an advanced state of affairs in the behavioral health setting. Machine learning is beginning to be introduced into the field but shows promise in achieving increased predictive power of current applications. Applying these techniques to predict the readmission state in individuals with OUD is believed to be a current gap in the literature.

Ensemble-based machine learning techniques

Random Forests

Random forest is an ensemble-based algorithm that generates many different decisions in each iteration of the algorithm (Breiman, 2001). At its core, random forest models operate off of the following principles:

> "A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models. "

The algorithm generates decision trees with subtleties in the parameters; then, it takes the model's run with the most accuracy on each iteration. Each tree that is generated then produces a prediction of the binary outcome variable. Each tree that accurately identifies the outcome variable state gets a vote. The trees with the most votes become the class prediction. By leveraging the low correlation between the models, the trees "protect" each from individuals errors. Typically, producing a model with a good ability to predict the outcome variable.

Gradient Boosting and AdaBoost

Gradient boosting AdaBoosting is another type of machine learning algorithm that allows the ability to predict an outcome variable. Gradient boosting comprises of three separate and distinct elements, a loss function that is allowed to be optimized, a weak learner that makes predictions, and an additive model to add to the weak learners to minimize the loss function (Natekin & Knoll, 2013).

The loss function allows us to quickly evaluate the model while ensuring each model does better than previous iterations of the model. In essence, a loss function quantifies error between the actual results, the predicted results and is a distillation of many variables into a

specific number. This specific loss function number allows multiple iterations of the algorithm to ooccur and the continued evaluation of which iteration "better" than the others (James, 2003).

A weak learner is something that is computationally simple and provides accuracy at best, relatively poorly. These weak learners could be simple classifiers of predictors or a regression analysis; however, they create a robust classifier or an ensemble-based classification when these are pooled together in multiple iterations. Gradient boosting relies on the multiple iterations of weak learners while minimizing the loss function to evaluate which weak learners provide the best results. The additive model allows each weak learner's learnings to be applied to each new iterative run, landing on the best solution(Joshi et al., 2002).

Adaboost or adaptive boosting is a variation on gradient boosting; the algorithm builds off the weak learners optimizing the loss function. However, in adaptive boosting, the sample distribution changes in each iteration. Changing the sample distributions, the weights on mispredicted weak learners increase, thus allowing the weights of correctly predicted weak learners to increase. This allows the algorithm to better focus on the more difficult iterations and, in many instances, leads to better predictive power (Friedman, 2001).

J48 Trees

J48 trees, a classification algorithm, allows the algorithm to identify numerous factors in a decision tree fashion to predict the outcome variable (Salzberg, 1994). The algorithm is a modification of the C4.5 algorithm developed by Ross Quinlan and an extension of the iterative dichtomiser 3, developed by the University of Waikato. The algorithm builds multiple decision trees utilizing the concept of information entropy. Since this is an ensemble-based algorithm, multiple trees are built and evaluated to take the best concepts of each tree and formulate them

down to one final tree. J48 trees are available in multiple different code types, including R-statistics, Python, and open-source Java.

Further Statistical Evaluation

Logistic Regression

Logistic regression is a common type of multivariate statistic which allows for the prediction of a binary outcome variable. Logistic regressions are wildly used in the psychiatric setting to evaluate treatment outcomes and the evaluation of psychometrics. Logistic regressions are derived from using a logistic function where the binary outcome variable is fit to regression and then predicted as a categorical binary variable. The general logistic regression equation can be derived as the following, where P is the probability of the state occurring, and $a + bX$ is the regression equation:

Equation 1

Logistic regression

$$\ln\left(\frac{P}{1-P}\right) = a + bX$$

Deep learning

Deep learning is another machine learning technique that utilizes neural networks to predict outputs from a dataset. Deep learning mimics the thinking of an animal's brain by simulating different layers that aid in the algorithm's decision-making process. These layers are the input layer, the hidden layer, and the output layer. The input layer is the data fed into the algorithm and the data that will be utilized in the mathematical computations of the hidden layer. The hidden layer in a deep learning algorithm is where the algorithms mathematical connections are made. There could be a vast number of hidden layers depending on the complexity of the

problem. The output layer is where the predicted value from the model is exposed, or the prediction is made in the case of opioid prediction readmission; this would be the prediction if the patient were readmitted into treatment or not readmitted into treatment (Indolia et al., 2018)

Figure 1

Conceptual Diagram Convoluted Neural Network

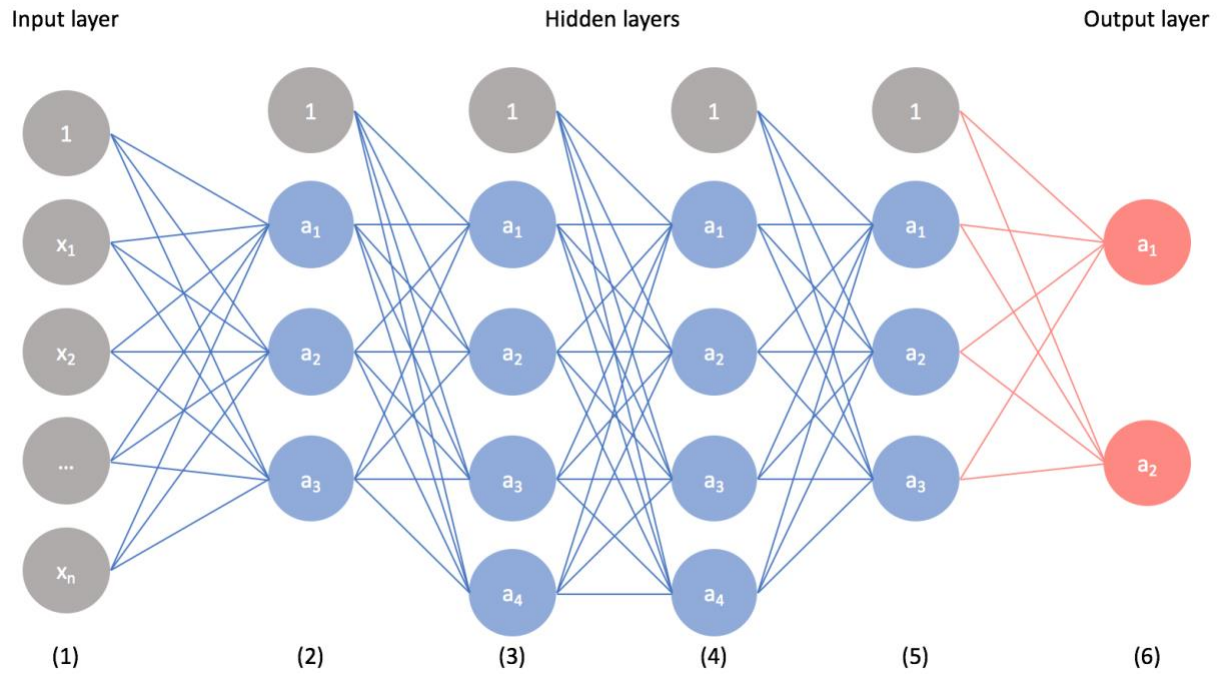


Image Credit: Jordan (2017)

Deep learning also relies on the use of a loss function, but in deep learning algorithms, it is called gradient descent. The gradient descent is a function where we could minimize the total number of errors in each associated model. The reduction in the area is due to modifying the weights in small increments on each of the input variables, which, once iterated over time, allows the cost function to be reduced. Because these weights are iterated many times, deep learning is a

computationally costly algorithm. Graphically depicted, J(w) is the function of error in the model, and w being the global cost minimum. In essence, the model attempts to reduce the total amount of error by adjusting weights to reach the $J_{min}(w)$, the lowest amount of global error in the outputs (Ruder, 2016).

Figure 2

Decision Gradient

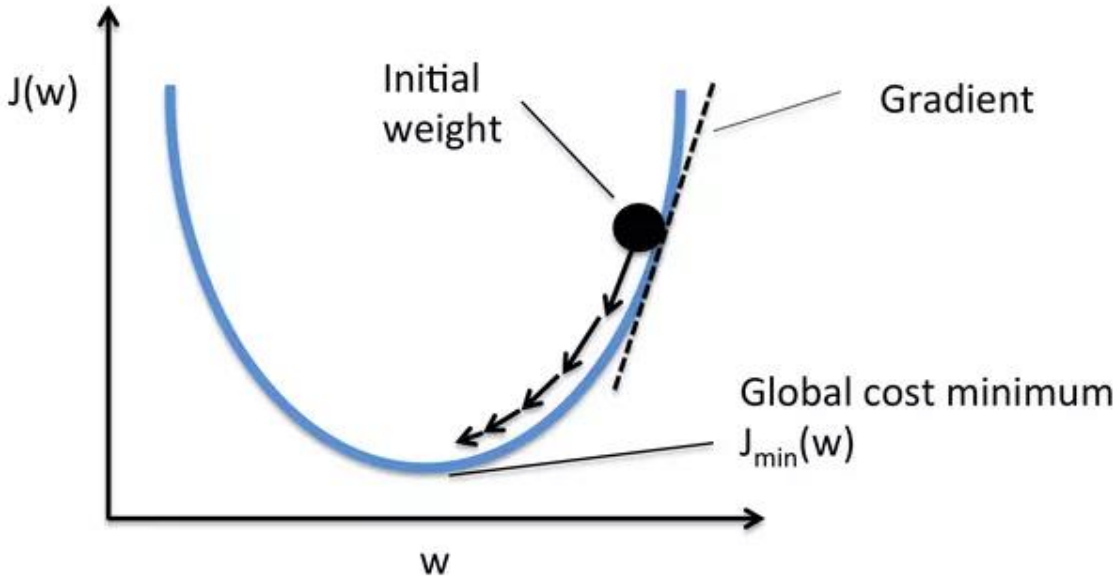


Image credit: Raschka (2020)

Unfortunately, this results in a "black box" where simple sets of rules are not easily obtained from the model due to complex interactions and relationships. Deep learning, however, has the capability of producing highly accurate models and has been utilized in pathology applications to identify lymph node metastases more accurately than a panel of pathologists (Ehteshami Bejnordi et al., 2017).

Dataset

The dataset utilized in this analysis was a de-identified dataset provided by Rogers Behavioral Health. Permission to utilize data was obtained by Rogers Behavioral Health, and ethical considerations were reviewed by the University of Wisconsin Milwaukee Institutional Review Board. The dataset was transmitted for analysis utilizing secure means and was received fully de-identified according to the Health Insurance Portability and Accountability Act (HIPAA) safe harbor method. The data was a merged dataset containing information from the electronic health record and Rogers Outcome Assessment System (ROAS), which supplied the patient-reported outcome measure variables. The dataset file was provided as a comma-separated values file and securely transferred to this author, providing encryption while in transit.

The dataset included patients from an inpatient medically managed detoxification for opioid use disorder unit with patients discharged from 2017 through 2020, at three separate facilities. Demographics that were analyzed included; age, discharge disposition, psychiatric diagnosis, employment status, ethnicity, highest education, length of stay, marital status, if commercial or Medicare or Medicaid insurance was utilized, the use of FDA approved recovery medications, sex, and as well as the QIDS, and QLESQ. Finally, patient flow characteristics were recorded with each encounter, including identifying if the patient stepped down to a less intensive level of care such as partial hospitalization, residential treatment, or intensive outpatient.

The dataset contains patient-reported outcome measures administered at admission to the program, every four days after admission, and finally, within 24 hours of discharge. Additionally, the dataset contains a flag for every encounter the patient was readmitted to the inpatient level of care. This readmission flag was primarily utilized as the outcome variable in all

subsequent analyses. Both the QIDS and QLESQ were self-reported, with the patient filling out each measure electronically utilizing a tablet and the ROAS system.

Methods

The analysis, in general, followed a similar methodology for the evaluation of each machine learning algorithm. All analysis and cleaning of the data were completed utilizing R-statistics version 4.0.3 "Bunny-Wunnies Freak Out" (R Core Team, 2020), as was the integrated development environmental R-Studio version 1.3.1093 (RStudio Team, 2021).

The dataset of information from the electronic health record in the patient-reported outcome system was cleaned to remove outliers, inconsistencies in data formatting and identifying and coercing the data into appropriate data classes in preparation for machine learning and analysis. Individuals who were flagged as readmitted were then classified in a binary variable stating if the patient was readmitted or has not readmitted. This binary class was utilized as the target variable.

The dataset utilized consisted of 2,103 patients admitted to Rogers Behavioral health's adult inpatient hospitalization unit in either Oconomowoc, West Allis, or Brown Deer, Wisconsin, between March of 2017 and December of 2020. Patients were included in the dataset if they had a primary diagnosis for OUD. In the sample, 160 patients readmitted back into one of the inpatient units at Rogers Behavioral Health, the individuals were labeled as "Readmitted," and all individuals who did not readmit to an inpatient unit at Rogers Behavioral Health were labeled as "No-Readmit." Demographic characteristics of the two groups are contained in Table 1:

Table 1

Sample Demographics

Variables	Readmitted (N=160)	No-Readmit (N=1943)
Age		
Mean (SD)	33.93 (10.80)	34.57 (10.07)
Min-Max	19-69	18-79
Sex (%)		
Female	52 (32.5)	777 (40)
Male	108 (67.5)	1166 (60)
Marital status (%)		
Single	107 (66.9)	1310 (67.4)
Separated	15 (9.4)	153 (7.9)
Married	14 (8.8)	215 (11.1)
Divorced	11 (6.9)	140 (7.2)
Widowed	1 (.6)	9 (.5)
Unknown	12 (7.5)	116 (6.0)
Race		
American Indian or Alaska Native	3 (1.9)	27 (1.4)
Asian	0 (0)	9 (.5)
Black or African American	12 (7.5)	151 (7.8)
Hispanic or Latino	0 (0)	0 (0)
Native Hawaiian or Pacific Islander	0 (0)	7 (.4)
White	134 (83.8)	1602 (82.4)
Unknown	11 (6.9)	147 (7)
Number of Diagnosis		
Mean (SD)	3.72 (1.57)	3.49 (1.62)
Min-Max	1-8	1-8
Discharge Disposition		
Home or Self Care	120 (75)	1539 (79.2)
Left Against Medical Advice	11 (6.9)	132 (6.8)
Other Healthcare Facility	10 (6.2)	47 (2.5)
Lower Level of Care (Residential, Partial Hospital, Intensive Outpatient)	19 (11.8)	225 (11.5)

QIDS Total Score		
Mean (SD)	14.70 (6.01)	14.01 (5.88)
QIDS Percent change admission to discharge		
Mean (SD)	-.31 (.59)	-6.04 (6.02)
Recovery Medications (%)		
Prescribed	96 (59.4)	1254 (64.5)
Not Prescribed	65 (40.6)	689 (35.5)
Length of Stay Days		
Mean (SD)	4.79 (2.35)	4.64(2.42)
Min-Max	1-20	1-47

As readmission to a hospital setting can be constituted as a relatively rare event, the dataset utilized was imbalanced in terms of individuals who readmitted versus those who did not readmit. To increase the number of cases and balance the dataset, a combination of both undersampling and oversampling was utilized. The dataset was oversampled with the readmitted cases, and those who did not readmit undersampled without replacement. This created a dataset with 1,010 individuals who did not readmit back to medically managed inpatient and 990 individuals who were readmitted back to treatment—all algorithms trained off the identical training dataset and were validated utilizing an identical validation dataset.

For all of the algorithms that were utilized in the analysis, four evaluation statistics were generated. These evaluation statistics include sensitivity, precision, accuracy, and the Matthews coefficient. The evaluation statistics were derived utilizing a confusion matrix for each algorithms output. A confusion matrix is a tool that allows for the representation of a model's accuracy. The confusion matrix aids in the identification of the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Table 2 depicts a typical confusion matrix is constructed:

Table 2

Confusion Matrix Example

Predicted/ Actual	Positive	Negative
Positive	TP	FP
Negative	FN	TN

Sensitivity is the statistic that measures the number of true positives within the algorithm. The higher the sensitivity value, the more accurately the algorithm was able to identify the number of true positives. Sensitivity is derived below. In the following equation:

Equation 2

$$Sensitivity = \frac{TP}{(TP + FN)}$$

Precision is the positive predictive value or the amount of correct positive predictions that were made. The higher the value on the positive predicted value, the more correct positive predictions the algorithm could make. The calculation for precision is derived below:

Equation 3

$$Precision = \frac{TP}{(TP + FP)}$$

Accuracy is how many data points the model correctly predicted. The higher the value for accuracy, the more correct true positives and true negatives that the algorithm was able to identify. Accuracy is derived below in the following equation:

Equation 4

$$Accuracy = \frac{(TP + TN)}{(P + N)}$$

The Matthews Correlation Coefficient is derived the same way as Pearson's phi and is a coefficient that takes into all aspects of the confusion matrix. Because the evaluation statistics take into account both true and false positives and negatives, it is regarded as one of the most balanced evaluation statistics. The higher the value of the Matthews coefficient, the better the algorithms' ability to predict the binary classes. The Matthews coefficient is derived below:

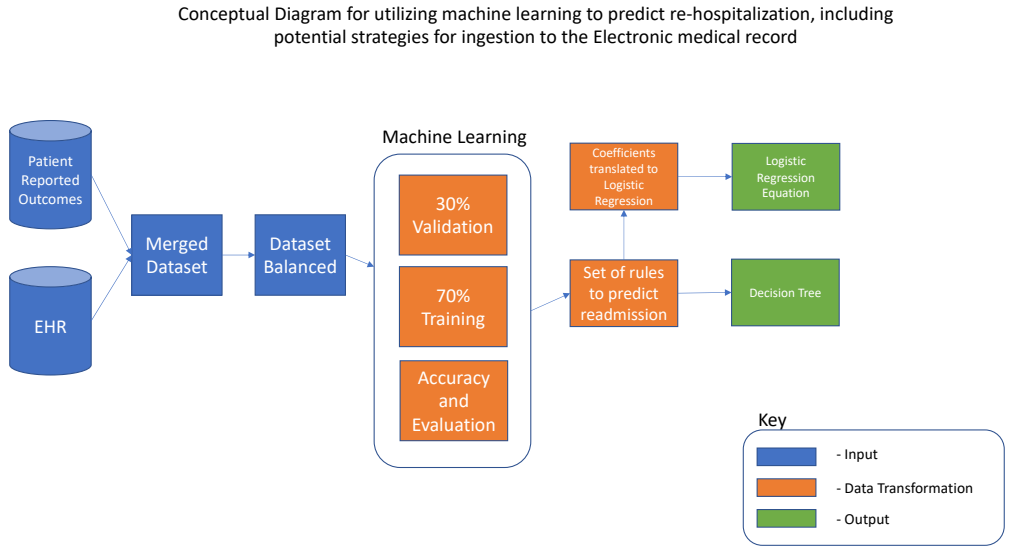
Equation 5

$$Matthews\ Correlation\ Coefficient = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

In all the algorithms, feature importance was also obtained. Feature importance allows an individual to understand which variables have a higher importance in making the overall prediction. A graphical representation of the entire method is contained in Figure 3:

Figure 3

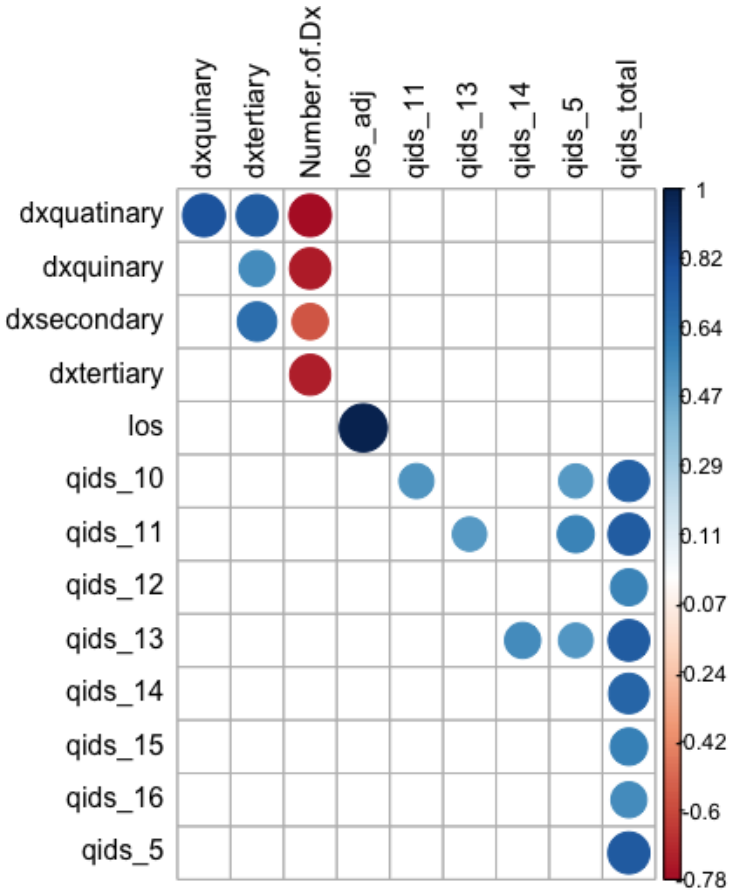
Conceptual diagram for utilizing machine learning to predict outcome states



Five distinct types of algorithms were evaluated in terms of sensitivity, precision, accuracy, and Matthew's Coefficient to evaluate each algorithm's overall accuracy. Before these evaluation statistics were derived, the dataset was evaluated for collinear variables through the use of correlation analysis. Individuals' diagnoses were highly correlated with the total number of diagnoses. As a result, the individual diagnosis was removed from the dataset in favor of the total number of diagnoses. Additionally, individual questions on the QIDS were correlated to the total score of the QIDS; to keep the differentiation of the responses to the QIDS items, QIDS total score was removed. Items 10 and 11, items 13 and 11, and items 13 and 14 were also correlated. However, this is likely due to the design of the QIDS as multiple questions assess the same symptoms. Due to this, it was determined that these questions remain within the dataset despite their collinearity. Figure 4 summarizes the correlation between the variables:

Figure 4

Correlated variables



All algorithms trained off the same dataset and were validated against the same validation dataset. The training dataset contained 758 patients who did not readmit and 742 patients who were readmitted to medically managed inpatient detoxification. The validation dataset consisted of 252 patients who did not readmit and 248 patients who were readmitted to medically managed inpatient detoxification.

Results

Random forest

A random forest model was produced utilizing the r package "Random forest" (Wiener, 2002). The algorithm was tuned with the following parameters, the initial set of trees to be produced was 2,000 trees, omit any instances that did not have a classification label, and proximity set as true. The random forest model produced a sensitivity of 0, a precision of 0, accuracy of 88%, and a Matthews coefficient of .-.02 The confusion matrix in Table 3 illustrates the output:

Table 3

Confusion Matrix: Random Forest

	No readmit	Readmit
No readmit	259	1
Readmit	33	0

The model did well at predicting individuals who were likely to not readmit into treatment. However, it did not accurately predict any individuals who were readmitted into treatment. The top five features obtained through the decision trees to predicting readmission included age, discharge disposition, payer type, and the total reduction of the quick inventory of depressive symptomology.

Gradient Boosting

The gradient boosted model utilized the method xgbTree, through the r package "XgBoost" (Tianqi Chen et al., 2020); the model parameters omitted any instances where there was no label to predict class. In this model, cross-fold validation was employed, with ten

crossfolds being applied. After training, tuning parameters were applied, including the number of rounds equaling 150, the max depth 150, an eta of 3, and gamma of .4. The gradient boosted model produced a sensitivity of 0%, a specificity of 100% , accuracy of 100%, and a Matthews coefficient of 0. The confusion matrix in Table 4 illustrates the output:

Table 4

Confusion Matrix: Gradient Boosting

	No readmit	Readmit
No readmit	500	0
Readmit	0	0

The model performed poorly, predicting that all individuals were predicting not to readmit, resulting in a Matthews coefficient of 0. Due this this, there was no stratification between the binary classes in this model, and the usability of the model impacted, as all individuals were labeled as not predicted to not readmit. The top features by importance that the model identified were difference in the QIDS score from admission to discharge, age, length of stay, and the total number of diagnoses.

J48 Trees

The J48 tree utilized the package "Rweka" for all model development and validation (Hornik K, 2009). The model parameters omitted any instances where there was not a label to predict class. Default settings were utilized in all other parameters. The J48 model produced a

sensitivity of 0%, a precision of 0%, accuracy of 88%, a Matthews coefficient of 0, and an accuracy of 91%. The confusion matrix in Table 5 illustrates the output.

Table 5

Confusion Matrix: J48

	No readmit	Readmit
No readmit	1186	0
Readmit	107	0

The model did well at predicting individuals who were likely to not readmit into treatment. However, it did not accurately predict any individuals who were readmitted into treatment. This resulted in a Mathews coefficient of 0. The models usability may be impacted given that the majority of individuals were predicted not to readmit, and those predicted readmit were incorrectly classified. The top features by importance that the model identified were age, discharge disposition, the total number of diagnoses, the use of FDA-approved recovery medication, and patient's length of stay on their original medically managed detoxification encounter.

Logistic Regression

The logistic regression created utilized the base package of the r-statistics. The model parameters omitted any instances where there was not a label to predict class. A train control was created utilizing five cross-fold validation. Additionally, the method utilized was a general linear model, with the family set as binomial. Default settings were utilized in all other parameters built

within the model. The logistic regression model produced a sensitivity of 59%, a precision of 59%, an accuracy of 61%, and a Matthews coefficient of 23%. The confusion matrix in Table 6 illustrates the output.

Table 6

Confusion Matrix: Logistic Regression

	No readmit	Readmit
No readmit	175	100
Readmit	100	147

The model did moderately well at predicting both those who did not readmit and those who did readmit. The models identified more important features from the QIDS, compared to the other models, with the top features being the QIDS item 5, QIDS item 15, QIDS item 11, the total number of diagnoses, and the age of the patient.

Adaboosting

The AdaBoost algorithm utilized the r package "adabag" (Alfaro, 2003). The model parameters omitted any instances where there was not a label to predict class, bootstrapped the training data utilizing the weights of each observation, and produced a total number of iterations as 50. Default settings were utilized in all other parameters. The AdaBoost model produced a sensitivity of 100%, a precision of 94%, an accuracy of 97%, and a Matthews coefficient of .94. The confusion matrix in Table 7 illustrates the output.

Table 7

Confusion Matrix: Adaboost

	No readmit	Readmit
No readmit	238	0
Readmit	14	248

The model did well at predicting individuals who were likely to not readmit into treatment and well at predicting who would readmit back into treatment. The top features by importance that the model identified were age, discharge disposition, the total number of diagnoses, the use of FDA-approved recovery medication, and patient's length of stay on their original medically managed detoxification encounter.

Deep learning

The deep learning model utilized the package H2o in order to build and tune the algorithm (Tom Kraljevic & Malohlava, 2020). The model required a considerable number of hyperparameter tuning in order to produce the best results. Parameters tuned included the number of hidden layers, number of epochs, momentum, and downsampling. The final hyperparameter tuned model included, 100000 epochs, 128 hidden layers, a momentum start of .2, with a momentum ramp of 1e7. . As a result of the hyperparameter tuning, the model produced a sensitivity of 98%, a precision of 100%, an accuracy of 99%, and a Matthews coefficient of .98. The confusion matrix in Table 8 illustrates the output.

Table 8

Confusion Matrix: Deep Learning

	No readmit	Readmit
No readmit	248	4
Readmit	0	252

The model performed exceptionally well at not only predicting who would readmit but who would not readmit utilizing the validation training set. The top features by importance were similar to the other models: QIDS item 16, the patient's age, the total number of diagnoses, the length of stay of the patient's original encounter, and the overall reduction of QIDS score throughout the stay.

Summary of model prediction

The models performed with a wide range of variation sensitivity, accuracy, precision, and Matthew's coefficient. However, the model with the most optimal performance in terms of all the evaluation statistics is the deep learning model. Table 9 summarizes the evaluation statistics between all of the models.

Table 9

Summary of evaluation statistics

	Sensitivity	Precision	Accuracy	Matthews
Random Forest	0	0	88%	-0.02
Gradient Boosting	0	0	100%	0
J48	0	0	91%	0
Logistic Regression	59%	59%	61%	0.23
Adaboosting	100%	94%	97%	.94
Deep Learning	98%	100%	98%	.99

The models produced similarly within the feature importance, will all the models selecting age as an essential factor of readmission prediction. Additionally, all but one model selected the total number of diagnoses as an important feature. The model's utilized items of the QIDS or the total reduction of the QIDs as an important feature for prediction. The models ranked different features as more important based on the model which was applied, with the logistic regression weighting the items from the QIDs more than other models. This may be due to the logistic regression model weighting all of the variables at equal importance, rather than an iterative, ensemble-based decision. With the machine learning algorithms, the features were more evenly spread between demographic variables and items on the QIDs; likely due to voting of features on individual runs of the machine learning algorithms. Additionally, through hyperparameter tuning, the deep learning model incorporated 128 hidden layers, allowing the model to identify more complex interactions, not possible through the other algorithms evaluated. Table 10 summarizes each model's top five features and illustrates the overlap of the features between each algorithm.

Table 10

Top Five Model Features

Feature	Random Forest	Gradient Boosting	J48	Logistic Regression	Adaboost	Deep Learning
Age	X	X	X	X	X	X
Discharge Disposition	X		X		X	

Payer Type	X					
Length of Stay		X	X		X	X
Total Number of Dx		X	X	X	X	X
FDA Recovery Meds			X		X	
QIDS Item 5				X		
QIDS Item 11				X		
QIDS Item 15				X		
QIDS Item 16						X
Reduction in QIDS	X	X				X

Discussion

Advanced data analytics, particularly machine learning techniques, appears to help gain new insights into sizeable behavioral health datasets, creating models with high accuracy at identifying who will readmit and those who will not readmit. Of all of the models generated, deep learning provided the overall best prediction as evaluated by the Matthews coefficient. The ability to produce a model with such high accuracy is promising to implement these decisions into behavioral healthcare. Additionally, the deep learning model had features consistent with the other models evaluated, lending credence that the factors displayed in Table 3 are at the very least correlated with identifying who is likely to readmit to an inpatient level of care.

A large amount of the models identified age, length of stay, and the total number of diagnosis as essential features to predicting outcome class and is consistent with the current state of the literature. The lower the age and length of stay was associated with a higher risk of readmission.Furthmore, the higher number of total diagnosis were associated with an increased

risk of readmission. The presence of multiple comorbidities is also congruent with the literature. Typical comorbidities including, depression, sleep disorders, personality disorders, and anxiety disorders (Grella et al., 2009), are all associated with OUD and all identified in the dataset. In medically managed detoxification, the primary goal of treatment is withdrawal management, ensuring medical stability, and discharge preparation. The ability to introduce standard behavioral health management techniques such as cognitive-behavioral therapy (CBT) and behavioral activation (BA) may be advantageous to reducing symptoms associated with common comorbid disorders.

Furthermore, the relationship between the length of stay and improved outcomes is also documented by (Oh et al., 2020), illustrating different treatment trajectories based on diagnosis on inpatient psychiatric treatment. Identifying a longer length of stay was consistently identified between all of the machine learning models evaluated. Length of stay in a psychiatric setting has varied across the years and within a geographic region, with the average length of stay declining close to the three days from the 1990s to the 2010s (Lee et al., 2012). This variation of length of stay may likely be due to a lack of robust outcome information, where the criteria to discharge are guided primarily through clinical judgment rather than care guidelines typical in the medical-surgical sector. This lack of length of stay optimization may be a critical factor in the readmission of opioid use disorder patients and psychiatric inpatients in general. The utilization of deep learning techniques may be a critical method to establishing the optimal length of stay.

Encouragingly, many of the aforementioned features are modifiable through treatment and giving clinicians the ability to actively reduce the probability that a patient would readmit to an intensive level of care. Since many of these factors are collected through the standard treatment documentation of opioid use disorder, it would be practical to implement these as a set of rules

within the EHR. The rules can either be quantified and weighted as an overall score, illustrating the probability of readmission, or utilized as flags for treatment interventions. By flagging which conditions a patient met, a clinician would optimize one's length of stay, treating comorbidities through targeted treatment interventions and proper discharge planning.

Additionally, this can create an even more robust treatment plan, individualized to not only the individual but to long-term outcomes. Treatment designed like this introduces personalized medicine into behavioral health in a way that is not prevalent in the space today and may help counter some of the ethical challenges of personalized medicine in psychiatry (Evers, 2009). It can also produce a framework for treatment that would be generalizable to other behavioral health disorders, identifying individuals' characteristics that place the patient at a higher risk of an adverse outcome.

However, there is a trade-off between algorithms that produce defined decision trees with less accuracy and deep learning algorithms with increasing accuracy and more obfuscated decision pathways. These factors may lead to less utilization of the features outputted by the algorithm due to clinicians having hesitancy to use when the algorithm is primarily a "black box" (Stead, 2018). With the potential lack of buy-in to utilize the information gained by the deep learning model, the work becomes more theoretical, rather than implementing functional changes to reduce readmission. The mistrust of deep learning techniques may prove to be a long term barrier to adopting many insights gained by applying these techniques. Further education to clinicians regarding machine learning techniques and the strategies on how to evaluate if an algorithm produces a reliable and robust model may increase adoption.

In addition to the "black box" produced by the deep learning algorithm, the other algorithms produced very robust decision trees. For example, the AdaBoost algorithm produced a decision tree with over 400 different decision points. Due to the sheer amount of decision points contained in the model, there may be barriers to implementing the findings at an individual patient level. Implementing this processing to the EHR may not only be difficult but not feasible given the current architecture of EHR's today. Additionally, data processing modules may need to be added to the EHR in order to achieve this level of processing.

There were also some underlying limitations within the dataset itself. All of the readmissions analyzed were patients who were admitted and readmitted to a single behavioral health system. This does not provide insight into the individuals who may have sought treatment at a different health system, or a different level of care, potentially missing important features when looking at readmissions from a more global perspective. This may bias the algorithm due to factors such as overall satisfaction with the treatment experience and satisfaction with providers not accounted for in the data. In essence, the data does not provide the opportunity to parse if a patient did not readmit due to being successful in maintaining sobriety or simply if they were unhappy with their treatment experience and sough another provider.

One way to potentially counter this limitation is data sharing with entities such as payers, the fidelity of the outcome variable can be increased due to the payor collecting claims data not specific to one entity. The ability to use claims data as the outcome variable could further strengthen the predictive factors and potentially lead to other features essential to predicting readmission, such as emergency room utilization.

From a data analysis perspective, the dataset was unbalanced, and the use of over and under sampling was applied, which results in duplication of readmission cases. The algorithms learned features over duplicated cases; this may have biased the algorithm to features that were apparent only in the duplicated cases. However, the use of undersampling and oversampling was necessary to produce a viable model throughout the analysis. As the dataset accumulates new patients over time, these issues may no longer be relevant; they may also be mitigated through further collaborations with payers.

Regardless of these limitations, the use of machine learning techniques in behavioral health appears to be a valuable next step in reducing the total number of readmissions and potentially the costs associated with subsequent readmissions. Additionally, identifying modifiable features may be one of the first steps to personalized medicine in behavioral health treatment and allows for optimized treatment outcomes. As datasets in behavioral healthcare continue to grow with high-fidelity structured data, more insights may be possible to predict outcome states. Furthermore, with more precise education to clinical providers, the adoption of machine learning algorithms will further increase, providing more robust and patient-driven outcomes.

References

About the ASAM Criteria. https://www.asam.org/asam-criteria/about

Alfaro, E., Gamez, M. Garcia, N. (2003). adabag: An R Package for Classification with Boosting and Bagging. *Journal of Statistical Software, 54*(2), 1-35. http://www.jstatsoft.org/v54/i02/

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.).

Banks, M. A. (2020, 2020/01/01). Sizing up big data. *Nature Medicine, 26*(1), 5-6. https://doi.org/10.1038/s41591-019-0703-0

Breiman, L. (2001, 2001/10/01). Random Forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Choy, G., Khalilzadeh, O., Michalski, M., Do, S., Samir, A. E., Pianykh, O. S., Geis, J. R., Pandharipande, P. V., Brink, J. A., & Dreyer, K. J. (2018, 2018/08/01). Current Applications and Future Impact of Machine Learning in Radiology. *Radiology, 288*(2), 318-328. https://doi.org/10.1148/radiol.2018171820

Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J. A. W. M., & Consortium, a. t. C. (2017). Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA, 318*(22), 2199-2210. https://doi.org/10.1001/jama.2017.14585

Ellis, R. J., Wang, Z., Genes, N., & Ma'ayan, A. (2019, 2019/01/29). Predicting opioid dependence from electronic health records with machine learning. *BioData Mining, 12*(1), 3. https://doi.org/10.1186/s13040-019-0193-0

Endicott, J., Nee, J., Harrison, W., & Blumenthal, R. (1993). *Quality of life enjoyment and satisfaction questionnaire: A new measure*

Evers, K. (2009). Personalized medicine in psychiatry: ethical challenges and opportunities. *Dialogues in clinical neuroscience, 11*(4), 427-434. https://doi.org/10.31887/DCNS.2009.11.4/kevers

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics, 29*(5), 1189-1232. http://www.jstor.org/stable/2699986

Grella, C. E., Karno, M. P., Warda, U. S., Niv, N., & Moore, A. A. (2009, Jun-Jul). Gender and comorbidity among individuals with opioid use disorders in the NESARC study. *Addictive behaviors, 34*(6-7), 498-504. https://doi.org/10.1016/j.addbeh.2009.01.002

Hatfield, D., McCullough, L., Frantz, S. H., & Krieger, K. (2010, Jan-Feb). Do we know when our clients get worse? an investigation of therapists' ability to detect negative client change. *Clin Psychol Psychother, 17*(1), 25-32. https://doi.org/10.1002/cpp.656

Hatton, C. M., Paton, L. W., McMillan, D., Cussens, J., Gilbody, S., & Tiffin, P. A. (2019). Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare. *Journal of affective disorders, 246*, 857-860.

Hornik K, B. C., Zeileis A. (2009). Open-Source Machine Learning: R MeetsWeka. *Computational Statistics, 24*(2), 225-232. https://doi.org/10.1007/s00180-008-0119-7

Indolia, S., Goswami, A. K., Mishra, S. P., & Asopa, P. (2018, 2018/01/01/). Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. *Procedia Computer Science, 132*, 679-688. https://doi.org/https://doi.org/10.1016/j.procs.2018.05.069

James, G. M. (2003, 2003/05/01). Variance and Bias for General Loss Functions. *Machine Learning, 51*(2), 115-135. https://doi.org/10.1023/A:1022899518027

Jordan, J. (2017). *Convolutional neural networks.* https://www.jeremyjordan.me/convolutional-neural-networks/

Joshi, M. V., Agarwal, R. C., & Kumar, V. (2002). *Predicting rare classes: can boosting make any weak learner strong?* https://doi.org/10.1145/775047.775092

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection* Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2, Montreal, Quebec, Canada.

Kosten, T. R., & Baxter, L. E. (2019, Feb). Review article: Effective management of opioid withdrawal symptoms: A gateway to opioid dependence treatment. *Am J Addict, 28*(2), 55-62. https://doi.org/10.1111/ajad.12862

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine, 16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lee, S., Rothbard, A. B., & Noll, E. L. (2012, Sep 1). Length of inpatient stay of persons with serious mental illness: effects of hospital and regional characteristics. *Psychiatr Serv, 63*(9), 889-895. https://doi.org/10.1176/appi.ps.201100412

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics, 7*, 21-21. https://doi.org/10.3389/fnbot.2013.00021

NIDA. (2020). *Effective Treatments for Opioid Addiction*. Retrieved 7/12/2020 from

Oh, H., Lee, J., Kim, S., Rufino, K. A., Fonagy, P., Oldham, J. M., Schanzer, B., & Patriquin, M. A. (2020). Time in treatment: Examining mental illness trajectories across inpatient psychiatric treatment. *medRxiv*, 2020.2006.2026.20140293. https://doi.org/10.1101/2020.06.26.20140293

R Core Team. (2020). *R: A Language and Environment for Statistical Computing.* In R Foundation for Statistical Computing. http://www.R-project.org/

Raschka, S. (2020). *Gradient Descent and Stochastic Gradient Descent*. http://rasbt.github.io/mlxtend/user_guide/general_concepts/gradient-optimization/

Rosenblum, A., Marsch, L. A., Joseph, H., & Portenoy, R. K. (2008). Opioids and the treatment of chronic pain: controversies, current status, and future directions. *Experimental and clinical psychopharmacology, 16*(5), 405-416. https://doi.org/10.1037/a0013628 (Experimental and clinical psychopharmacology)

Roy, A., Nikolitch, K., McGinn, R., Jinah, S., Klement, W., & Kaminsky, Z. A. (2020, 2020/05/26). A machine learning approach predicts future risk to suicidal ideation from social media data. *npj Digital Medicine, 3*(1), 78. https://doi.org/10.1038/s41746-020-0287-6

RStudio Team. (2021). *RStudio: Integrated Development for R.* In RStudio, PBC. http://www.rstudio.com/.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003, 2003/09/01/).

The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573-583. https://doi.org/https://doi.org/10.1016/S0006-3223(02)01866-8

Salzberg, S. L. (1994, 1994/09/01). C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Machine Learning, 16*(3), 235-240. https://doi.org/10.1007/BF00993309

Stead, W. W. (2018). Clinical Implications and Challenges of Artificial Intelligence and Deep Learning. *JAMA, 320*(11), 1107-1108. https://doi.org/10.1001/jama.2018.11029

Substance Abuse and Mental Health Services Administration (2006). *Center for Substance Abuse Treatment. Detoxification and Substance Abuse Treatment.* . https://www.ncbi.nlm.nih.gov/books/NBK64109/

Tianqi Chen, T. H., Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu, Cho, K. C., Rory Mitchell, Ignacio Cano, Tianyi Zhou, Mu Li, Junyuan, & Xie, M. L., Yifeng Geng and Yutian Li. (2020). *xgboost: Extreme Gradient Boosting.* In https://CRAN.R-project.org/package=xgboost

Tom Kraljevic, T. N., Patrick Aboyoun, Michal Kurka and Michal, & Malohlava. (2020). *h2o: R Interface for the 'H2O' Scalable Machine LearningPlatform.* In https://CRAN.R-project.org/package=h2o

Volkow, N. D. (2018, Jan 27). Medications for opioid use disorder: bridging the gap in care. *Lancet, 391*(10118), 285-287. https://doi.org/10.1016/s0140-6736(17)32893-3

Wakeman, S. E., Green, T. C., & Rich, J. (2020, 2020/06/01). An overdose surge will compound the COVID-19 pandemic if urgent action is not taken. *Nature Medicine, 26*(6), 819-820. https://doi.org/10.1038/s41591-020-0898-0 (Nature Medicine)

Wiener, A. L. a. M. (2002). Classification and Regression by randomForest. *R News, 2*(3), 18-22. https://CRAN.R-project.org/doc/Rnews/

Zimmerman, M., & McGlinchey, J. B. (2008, Dec). Why don't psychiatrists use scales to measure outcome when treating depressed patients? *J Clin Psychiatry, 69*(12), 1916-1919. https://doi.org/10.4088/jcp.v69n1209

APPENDICES

Quick Inventory of Depressive Symptomatology 16 item short form

The Quick Inventory of Depressive Symptomatology (16-Item) (Self-Report) (QIDS-SR 16)

Name or ID: _____ Date: _____

CHECK THE ONE RESPONSE TO EACH ITEM THAT BEST DESCRIBES YOU FOR THE PAST SEVEN DAYS.

During the past seven days...

1. Falling Asleep:

☐ 0 I never take longer than 30 minutes to fall asleep.

☐ 1 I take at least 30 minutes to fall asleep, less than half the time.

☐ 2 I take at least 30 minutes to fall asleep, more than half the time.

☐ 3 I take more than 60 minutes to fall asleep, more than half the time.

2. Sleep During the Night

☐ 0 I do not wake up at night.

☐ 1 I have a restless, light sleep with a few brief awakenings each night.

☐ 2 I wake up at least once a night, but I go back to sleep easily.

☐ 3 I awaken more than once a night and stay awake for 20 minutes or more, more than half the time.

3. Waking Up Too Early:

☐ 0 Most of the time, I awaken no more than 30 minutes before I need to get up.

☐ 1 More than half the time, I awaken more than 30 minutes before I need to get up.

☐ 2 I almost always awaken at least one hour or so before I need to, but I go back to sleep eventually.

☐ 3 I awaken at least one hour before I need to, and can't go back to sleep.

4. Sleeping Too Much:

☐ 0 I sleep no longer than 7-8 hours/night, without napping during the day.

☐ 1 I sleep no longer than 10 hours in a 24-hour period including naps.

☐ 2 I sleep no longer than 12 hours in a 24-hour period including naps.

☐ 3 I sleep longer than 12 hours in a 24-hour period including naps.

During the past seven days...

5. Feeling Sad:

☐ 0 I do not feel sad.

☐ 1 I feel sad less than half the time.

☐ 2 I feel sad more than half the time.

☐ 3 I feel sad nearly all of the time.

Please complete either 6 or 7 (not both)

6. Decreased Appetite:

☐ 0 There is no change in my usual appetite.

☐ 1 I eat somewhat less often or lesser amounts of food than usual.

☐ 2 I eat much less than usual and only with personal effort.

☐ 3 I rarely eat within a 24-hour period, and only with extreme personal effort or when others persuade me to eat.

- OR -

7. Increased Appetite:

☐ 0 There is no change from my usual appetite.

☐ 1 I feel a need to eat more frequently than usual.

☐ 2 I regularly eat more often and/or greater amounts of food than usual.

☐ 3 I feel driven to overeat both at mealtime and between meals.

Please complete either 8 or 9 (not both)

8. Decreased Weight (Within the Last Two Weeks):

☐ 0 I have not had a change in my weight.

☐ 1 I feel as if I have had a slight weight loss.

☐ 2 I have lost 2 pounds or more.

☐ 3 I have lost 5 pounds or more.

- OR -

9. Increased Weight (Within the Last Two Weeks):

☐ 0 I have not had a change in my weight.

☐ 1 I feel as if I have had a slight weight gain.

☐ 2 I have gained 2 pounds or more.

☐ 3 I have gained 5 pounds or more.

Quality of Life Enjoyment and Satisfaction Short Form (Q-LES-Q- SF)

Name:_____ Date:_____

Quality of Life Enjoyment and Satisfaction Questionnaire – Short Form (Q-LES-Q-SF)

Taking everything into consideration, during the past week how satisfied have you been with your……

	Very Poor	Poor	Fair	Good	Very Good
…..physical health?	1	2	3	4	5
…..mood?	1	2	3	4	5
…..work?	1	2	3	4	5
…..household activities?	1	2	3	4	5
…..social relationships?	1	2	3	4	5
…..family relationships?	1	2	3	4	5
…..leisure time activities?	1	2	3	4	5
…..ability to function in daily life?	1	2	3	4	5
…..sexual drive, interest and/or performance?*	1	2	3	4	5
…..economic status?	1	2	3	4	5
…..living/housing situation?*	1	2	3	4	5
…..ability to get around physically without feeling dizzy or unsteady or falling?*	1	2	3	4	5
…..your vision in terms of ability to do work or hobbies?*	1	2	3	4	5
…..overall sense of well being?	1	2	3	4	5
…..medication? (If not taking any, check here _____ and leave item blank.)	1	2	3	4	5
…..How would you rate your overall life satisfaction and contentment during the past week?	1	2	3	4	5

*If satisfaction is very poor, poor or fair on these items, please UNDERLINE the factor(s) associated with a lack of satisfaction.

R-Statistics Code

```
### Author: Brian Kay
### Predicting readmission in OUD- Exploratory Data Analysis, and predictive models
### Updated: 12/31/2020
### PACKAGE LOAD

library(openxlsx)
library(Boruta)
library(car)
library(corrplot)
library(data.table)
library(plyr) ## Load this before dplyr
library(dplyr)
library(olsrr)
library(psych)
library(questionr)
library(randomForest)
require(caTools)
library(tidyverse)
library(caret)
library(xgboost)
library(imputeTS)
library(stringr)
library(tree)
library(FFTrees)
library(RWeka)
library(party)
library(FSelector)
library(caret)
library(pROC)
library(adabag)
library(mlbench)
library(forcats)
library(Amelia)
library(tidyr)
library(h2o)

# Run Descriptive Statistics:

# Subset to only baseline & by readmit vs non readmit

OUD_Baseline<-subset(OUD_data, phase=="Baseline")
OUD_Baseline$phase<-NULL
OUD_Baseline_ML<-OUD_Baseline[!is.na(OUD_Baseline$Readmit_Orig),]
descriptivestatsreadmit<-subset(OUD_Baseline_ML, Readmit_Orig=="Readmit")
```

```
descriptivestatsnoreadmit<-subset(OUD_Baseline_ML, Readmit_Orig=="No Readmit")

# Summary of demographic information in each group
describeBy(OUD_Baseline_ML,OUD_Baseline_ML$Readmit_Orig=="Readmit")
describeBy(OUD_Baseline_ML,OUD_Baseline_ML$Readmit_Orig=="No Readmit")
#Frequencies for readmit
questionr::freq(descriptivestatsreadmit$sex, cum = TRUE, sort = "dec", total = TRUE) #
Discharge sex

questionr::freq(descriptivestatsreadmit$dcdisposition, cum = TRUE, sort = "dec", total = TRUE)
# Discharge disposition

questionr::freq(descriptivestatsreadmit$FDA_Recovery_Med, cum = TRUE, sort = "dec", total =
TRUE) # Recovery Medications

questionr::freq(descriptivestatsreadmit$race, cum = TRUE, sort = "dec", total = TRUE) # Race

questionr::freq(descriptivestatsreadmit$marital, cum = TRUE, sort = "dec", total = TRUE) #
Marital


# Frequencies for no readmit

questionr::freq(descriptivestatsnoreadmit$dcdisposition, cum = TRUE, sort = "dec", total =
TRUE) # Discharge disposition

questionr::freq(descriptivestatsnoreadmit$FDA_Recovery_Med, cum = TRUE, sort = "dec",
total = TRUE) # Recovery Medications

questionr::freq(descriptivestatsnoreadmit$marital, cum = TRUE, sort = "dec", total = TRUE)
#Marital

questionr::freq(descriptivestatsnoreadmit$sex, cum = TRUE, sort = "dec", total = TRUE) #Sex


questionr::freq(descriptivestatsnoreadmit$race, cum = TRUE, sort = "dec", total = TRUE) #
Race


# Understand which variables are highly correlated
corr_simple <- function(data=df,sig=0.5){
  #convert data to numeric in order to run correlations
  #convert to factor first to keep the integrity of the data - each value will become a number
rather than turn into NA
  df_cor <- data %>% mutate_if(is.character, as.factor)
  df_cor <- df_cor %>% mutate_if(is.factor, as.numeric)
```

```
  #run a correlation and drop the insignificant ones
  corr <- cor(df_cor)
  #prepare to drop duplicates and correlations of 1
  corr[lower.tri(corr,diag=TRUE)] <- NA
  #drop perfect correlations
  corr[corr == 1] <- NA
  #turn into a 3-column table
  corr <- as.data.frame(as.table(corr))
  #remove the NA values from above
  corr <- na.omit(corr)
  #select significant values
  corr <- subset(corr, abs(Freq) > sig)
  #sort by highest correlation
  corr <- corr[order(-abs(corr$Freq)),]
  #print table
  print(corr)
  #turn corr back into matrix in order to plot with corrplot
  mtx_corr <- reshape2::acast(corr, Var1~Var2, value.var="Freq")

  #plot correlations visually
  corrplot(mtx_corr, is.corr=FALSE, tl.col="black", na.label=" ")
}
corr_simple(OUD_Impute)


#Duplicative Variables and Colinear Variables
OUD_Impute$los<-NULL
OUD_Impute$ethnic<-NULL
OUD_Impute$employment<-NULL
OUD_Impute$highest_education<-NULL
OUD_Impute$dxquatinary<-NULL
OUD_Impute$dxquinary<-NULL
OUD_Impute$dxtertiary<-NULL
OUD_Impute$Continuation<-NULL
OUD_Impute$dxprimary<-NULL
OUD_Impute$dxsecondary<-NULL
OUD_Impute$qids_percent_change<-NULL
OUD_Impute$qids_total<-NULL
missmap(OUD_Impute)

#Oversample and undersample the Dataset to balance
library(ROSE)
data_balanced_both<- ovun.sample(Readmit_Orig ~., data=OUD_Impute ,method= "both",
N=2000, seed = 123)$data
table(data_balanced_both$Readmit_Orig)
OUD_Baseline_ML<-data_balanced_both
```

```
#Reduce dataset to feature selected variables

# Apply Random forest model
set.seed(101)
sample = sample.split(OUD_Baseline_ML$Readmit_Orig,SplitRatio=.75)
train= subset(OUD_Baseline_ML,sample ==TRUE)
test= subset(OUD_Baseline_ML,sample ==FALSE)

train<-train %>% mutate_if(is.character, as.factor)

rf<-randomForest(
  Readmit_Orig ~.,
  data=train,
  ntree= 2000,
  importance=TRUE,
  proximity=TRUE,
  na.action = na.omit
)

getTree(rf,1,labelVar = TRUE)
pred= predict(rf, newdata=test[-36])
cm= table(test[,36],pred)
cm


#Apply gradient boosting model

set.seed(123)
OUD_gradient <- train(
  Readmit_Orig ~., data=train, method= "xgbTree",
  trControl=trainControl("cv",number=10),
  na.action= na.omit
  )

#Best tuning parameter
OUD_gradient$bestTune

#Make predictions on the test data
pred<-predict(OUD_gradient,test)
pred.resp <- ifelse(pred >0.86, 1, 0)
predicted.classes<-OUD_gradient %>% predict(test)
head(predicted.classes)
confusionMatrix(pred.resp, Readmit_Orig, positive="Readmit")
#Compute model prediction accuracy rate
mean(predicted.classes == test$Readmit_Orig)
varImp(OUD_gradient)
```

```
plot(varImp(OUD_gradient), top=10)
y_pred=predict(OUD_gradient,test)
err <- mean(as.numeric(pred > 0.5))
print(paste("test-error=", err))


#Apply J48
j48 <- J48(Readmit_Orig~., data = train)

summary(j48)

plot(information.gain(Readmit_Orig~., data = OUD_Baseline_ML), top=10)
y_pred=predict(j48,test)
err <- mean(as.numeric(pred > 0.5))
print(paste("test-error=", err))
#Information gain of J48
information.gain(Readmit_Orig~., data = OUD_Baseline_ML)
if(require("party", quietly = TRUE)) plot(j48)

#Apply Logistic Regression

trCntl <- trainControl(method = "CV",number = 5)
glmModel <- train(Readmit_Orig ~ .,data = train,trControl = trCntl,method="glm",family =
"binomial")
# print the model info
summary(glmModel)
glmModel
confusionMatrix(glmModel)
# generate predictions on hold back data
trainPredicted <- predict(glmModel,test)
# generate confusion matrix for hold back data
confusionMatrix(trainPredicted,reference=test$Readmit_Orig)


##adaboost

adaboost<-boosting(Readmit_Orig~ ., data=train ,boos=TRUE, mfinal=50)
predadaboost<-predict(adaboost,newdata=test)
print(predadaboost$error)
get_tree(adaboost,1)
print(predadaboost$confusion)
print(predadaboost$trees)

#Boosted model with Crossfold validation
adaboostcv<-boosting.cv(Readmit_Orig~ ., data=train ,boos=TRUE, mfinal=50, v=10)
print(adaboostcv[-1])
```

```
## Deep Learning Model,derived from h2o tutorial

library(h2o)
h2o.init(nthreads = -1)

#Convert To H2o dataframe

train<-as.h2o(train)
test<-as.h2o(test)

#Set target variable and predictors
response<-"Readmit_Orig"
predictors <- setdiff(names(train), response)
predictors

#Model 1
m1 <- h2o.deeplearning(
  model_id="dl_model_first",
  training_frame=train,
  validation_frame=test,   ## validation dataset: used for scoring and early stopping
  x=predictors,
  y=response,
  activation="Rectifier",
  hidden=c(200,200),
  epochs=1,
  variable_importances=T
)
summary(m1)

head(as.data.frame(h2o.varimp(m1)))

#Model 2
m2 <- h2o.deeplearning(
  model_id="dl_model_faster",
  training_frame=train,
  validation_frame=test,
  x=predictors,
  y=response,
  hidden=c(256,256,256),
  epochs=1000000,
  score_validation_samples=10000,
  stopping_rounds=2,
  stopping_metric="misclassification",
  stopping_tolerance=0.01
)
```

```
summary(m2)
plot(m2)

#final Model Tuned
m3 <- h2o.deeplearning(
  model_id="dl_model_tuned",
  training_frame=train,
  validation_frame=test,
  x=predictors,
  y=response,
  overwrite_with_best_model=F,
  hidden=c(128,128,128,128),
  epochs=1000000,
  score_validation_samples=10000,
  score_duty_cycle=0.025,
  adaptive_rate=F,
  rate=0.01,
  rate_annealing=2e-6,
  momentum_start=0.2,
  momentum_stable=0.4,
  momentum_ramp=1e7,
  l1=1e-5,
  l2=1e-5,
  max_w2=10
)
summary(m3)

h2o.performance(m3, train=T)
h2o.performance(m3, test=T)
h2o.performance(m3, newdata=train)
h2o.performance(m3, newdata=test)

#Tuning Hyperparameters
hyper_params <- list(
  hidden=list(c(256,256,256),c(64,64)),
  input_dropout_ratio=c(0,0.05),
  rate=c(0.01,0.02),
  rate_annealing=c(1e-8,1e-7,1e-6)
)

hyper_params
grid <- h2o.grid(
  algorithm="deeplearning",
  grid_id="dl_grid",
  training_frame=train,
  validation_frame=test,
```

```
  x=predictors,
  y=response,
  epochs=10,
  stopping_metric="misclassification",
  stopping_tolerance=1e-2,        ## stop when misclassification does not improve by >=1% for 2
scoring events
  stopping_rounds=2,
  score_validation_samples=10000,
  score_duty_cycle=0.025
  adaptive_rate=F,
  momentum_start=0.5,
  momentum_stable=0.9,
  momentum_ramp=1e7,
  l1=1e-5,
  l2=1e-5,
  activation=c("Rectifier"),
  max_w2=10,
  hyper_params=hyper_params
)
grid

dlmodel <- h2o.deeplearning(
  x=predictors,
  y="bin_response",
  training_frame=train,
  hidden=c(10,10),
  epochs=0.1
)
summary(dlmodel)

grid <- h2o.getGrid("dl_grid",sort_by="err",decreasing=FALSE)
grid

## To see what other "sort_by" criteria are allowed
#grid <- h2o.getGrid("dl_grid",sort_by="wrong_thing",decreasing=FALSE)

## Sort by logloss
h2o.getGrid("dl_grid",sort_by="logloss",decreasing=FALSE)

## Find the best model and its full set of parameters
grid@summary_table[1,]
best_model <- h2o.getModel(grid@model_ids[[1]])
best_model

print(best_model@allparameters)
print(h2o.performance(best_model, valid=T))
```

```
print(h2o.logloss(best_model, valid=T))


hyper_params <- list(

activation=c("Rectifier","Tanh","Maxout","RectifierWithDropout","TanhWithDropout","Maxou
tWithDropout"),
  hidden=list(c(20,20),c(50,50),c(30,30,30),c(25,25,25,25)),
  input_dropout_ratio=c(0,0.05),
  l1=seq(0,1e-4,1e-6),
  l2=seq(0,1e-4,1e-6)
)
hyper_params

## Stop once the top 5 models are within 1% of each other (i.e., the windowed average varies
less than 1%)
search_criteria = list(strategy = "RandomDiscrete", max_runtime_secs = 360, max_models =
100, seed=1234567, stopping_rounds=5, stopping_tolerance=1e-2)
dl_random_grid <- h2o.grid(
  algorithm="deeplearning",
  grid_id = "dl_grid_random",
  training_frame=train,
  validation_frame=test,
  x=predictors,
  y=response,
  epochs=1,
  stopping_metric="logloss",
  stopping_tolerance=1e-2
  stopping_rounds=2,
  score_validation_samples=10000,
  score_duty_cycle=0.025,
  max_w2=10,
  hyper_params = hyper_params,
  search_criteria = search_criteria
)
grid <- h2o.getGrid("dl_grid_random",sort_by="logloss",decreasing=FALSE)
grid

grid@summary_table[1,]
best_model <- h2o.getModel(grid@model_ids[[1]])
best_model

grid <- h2o.getGrid("dl_grid",sort_by="err",decreasing=FALSE)
best_model <- h2o.getModel(grid@model_ids[[1]])
h2o.confusionMatrix(best_model,valid=T)
best_params <- best_model@allparameters
```

```
best_params$activation
best_params$hidden
best_params$input_dropout_ratio
best_params$l1
best_params$l2


max_epochs <- 10000 ## Add two more epochs
m_cont <- h2o.deeplearning(
  model_id="dl_model_tuned_continued",
  checkpoint="dl_model_tuned",
  training_frame=train,
  validation_frame=test,
  x=predictors,
  y=response,
  hidden=c(128,128),
  epochs=max_epochs,
  stopping_metric="logloss",
  stopping_tolerance=1e-2,        scoring events
  stopping_rounds=2,
  score_validation_samples=10000,
  score_duty_cycle=0.025,
  adaptive_rate=F,
  rate=0.01,
  rate_annealing=2e-6,
  momentum_start=0.2,
  momentum_stable=0.4,
  momentum_ramp=1e7,
  l1=1e-5,
  l2=1e-5,
  max_w2=10
)
summary(m_cont)
plot(m_cont)

dlmodel <- h2o.deeplearning(
  x=predictors,
  y=response,
  training_frame=train,
  hidden=c(10,10),
  epochs=1,
  nfolds=5,
  fold_assignment="Modulo" # can be "AUTO", "Modulo", "Random" or "Stratified"
)
dlmodel
```

```
#Binomial Model
dlmodel <- h2o.deeplearning(
  x=predictors,
  y="Readmit_Orig",
  training_frame=train,
  validation_frame = test,
  hidden=c(128,128,128,128),
  epochs=1000,
  reproducible = T
)
summary(dlmodel)
h2o.varimp(dlmodel)
h2o.varimp_plot(dlmodel,num_of_features = 5)
plot(h2o.performance(dlmodel))
```

CURRICULUM VITAE

Brian Kay | Curriculum Vitae

Education

Doctor of Philosophy, Biomedical Informatics- University of Wisconsin- Milwaukee — 2017- 2021

Master of Science, Healthcare Informatics — University of Wisconsin- Milwaukee — 2012 — 2013
Thesis: *Alcohol Biomarkers as Predictive factors of Rearrest in High Risk Repeat Offense Drunk Drivers*
Defended: August 2013

Bachelor of Arts, Psychology — Marquette University, Milwaukee — 2006 — 2009

Employment

Vice President, Continuous Improvement— Rogers Memorial Hospital, Oconomowoc Wisconsin — 2017— Present

Directs teams utilizing lean methodologies to improve clinical quality, operational efficiencies, and data science directives through a large national behavioral health organization. Oversees three million dollar annual budget and 20 direct reports. Manages data science initiatives with outside collaborators for the enterprise.

Director of Clinical Effectiveness— Rogers Memorial Hospital, Oconomowoc Wisconsin — 2015— 2017

Leads an interdisciplinary data mining and analysis team; fusing clinical, quality, an operations data. Collaborates with clinical and operational leadership to implement findings into processes and procedures throughout a large national behavioral health system. Leads the research and development of greenfield mobile health application.

Peer-Reviewed Publications

Meier, Marieke & Kossakowski, Jolanda & Jones, Payton & **Kay, Brian** & Riemann, Bradley & McNally, Richard. (2019). Obsessive-compulsive symptoms in eating disorders: A network investigation. International Journal of Eating Disorders. 10.1002/eat.23196.

Gregory, Sean & **Kay, Brian** & Riemann, Bradley & Goodman, Wayne & Storch, Eric. (2019). Cost-effectiveness of treatment alternatives for treatment-refractory pediatric obsessive-compulsive disorder. Journal of Anxiety Disorders. 102151. 10.1016/j.janxdis.2019.102151.

Højgaard, Davíð & Schneider, Sophie & La Buissonniere Ariza, Valerie & **Kay, Brian** & Riemann, Bradley & Jacobi, David & Eken, Stephanie & Lake, Peter & Goodman, Wayne & McIngvale, Elizabeth & Storch, Eric. (2019). Predictors of treatment outcome for youth receiving intensive residential treatment for obsessive-compulsive disorder (OCD). Cognitive Behaviour Therapy. 1-13. 10.1080/16506073.2019.1614977.

D. Molinari, Arielle & L. Andrews, Jessica & Zaboski, Brian & **Kay, Brian** & Hamblin, Rebecca & Gilbert, Alexandra & Ramos, Amaya & C. Riemann, Bradley & Eken, Stephanie & M. Nadeau, Joshua & Storch, Eric. (2019). Quality of life and anxiety in children and adolescents in residential treatment facilities. Residential Treatment for Children & Youth. 1-15.

Gregory, Sean & **Kay, Brian** & Smith, Joseph & Hall, Kristin & De Nadai, Alessandro & Quast, Troy & C. Riemann, Bradley & Storch, Eric. (2018). Treatment-Refractory Obsessive-Compulsive Disorder in Adults: A Cost-Effectiveness Analysis of Treatment Strategies. The Journal of Clinical Psychiatry. 79.

La Buissonniere Ariza, Valerie & Schneider, Sophie & Højgaard, Davíð & C. **Kay, Brian** & C. Riemann, Bradley & C. Eken, Stephanie & Lake, Peter & M. Nadeau, Joshua & Storch, Eric. (2017 Family Accommodation of Anxiety Symptoms in Youth Undergoing Intensive Multimodal Treatment for Anxiety Disorders and Obsessive-Compulsive Disorder: Nature, Clinical Correlates, and Treatment Response. Comprehensive Psychiatry. 80.

Storch, Eric & **Kay, Brian.** (2017). Commentary on Spradlin et al.: Is marijuana use common in OCD?. Addictive Behaviors. 10.1016/j.addbeh.2017.07.028.

Iniesta-Sepúlveda, Marina & Nadeau, Joshua & Ramos, Amaya & **Kay, Brian** & Riemann, Bradle & Storch, Eric. (2017). An Initial Case Series of Intensive Cognitive-Behavioral Therapy for Obsessive-Compulsive Disorder in Adolescents with Autism Spectrum Disorder. Child psychiatry and human development. 49.

Storch, E., **Kay, B**., Wu, S., Nadeau, J., & Riemann, B..“Suicidal and death ideation among adults with obsessive-compulsive disorder presenting for intensive intervention" Annals Of Clinical Psychiatry, Volume 29 Issue 1.

Storch,E.,Naadeau, J., De Nadai, A., & **Kay,B**.. "Symptom Correspondence Between Clinicians and Patients on the Yale-Brown Obsessive Compulsive Scale" Comprehensive Psychiatry, Volume 73.

Kay, B., Eken, S., Jacobi, J., Riemann, B., & Storch, E..Gene "Outcome of Multidisciplinary, CBT-focused Treatment for Pediatric OCD" ral Hospital Psychiatry, Volume 42.

Nadeau, J., De Nadai, A., Viar-Paxton, M., Olatunji, B., Jacobi, D., Eken, S., **Kay, B**., Riemann. B., & Storch, E.. "Further Psychometric Evaluation of the Child Disgust Scale." Child Psychiatry and Human Development, *In Press*.

Weltzin,T., **Kay**, **B.**, Cornella-Carlson, T., Timmel, P., Klosterman, E., Kinnear, K., Welk-Richards R., Lee, H., & Bean, P.. *"Long-Term Effects of a Multidisciplinary Residential Treatment Model or*

Improvements of Symptoms and Weight in Adolescents with Eating Disorders." Journal of Groups ir Addiction & Recovery, Volume 9 Issue 1.

Bean, P., **Kay**, **B.**, Bean, J., Roska, C., Pearson, J., Garuz, C., & Hallinan, P.. *"Recidivism Risk of Repeat Intoxicated Drivers Monitored with Alcohol Biomarkers".* Alcoholism Treatment Quarterly, Volume 32 Issue 4, 433

Bean,P. ,Louks, H.,**Kay, B.**, Cornella-Carlson, T. ,& Weltzin, T.. "*Clinical Observations of the Imp of Maudsley Therapy in Improving Eating Disorder Symptoms, Weight, and Depression in Adolesce Receiving Treatment for Anorexia Nervosa.*" Journal of Groups in Addiction & Recovery, Volume 5 Issue 1, 70.

Bean,P. , Roska, C.,Harasymiw, J., Pearson, J. ,**Kay**, **B.,** & Louks, H.. "*Alcohol Biomarkers as Tool Guide and Support Decisions About Intoxicated Driver Risk.*" Traffic Injury Prevention, Volume 10 Issue 6, 519.

Invited Presentations

Metric Driven Strategy for EHR Continuous Improvement , speaker at Healthcare Information Management System Society, Orlando, FL February 12th, 2019.

Behavioral Health and Cerner: Improved Revenue and Clinical Outcomes, speaker at Cerner Health Conference, Kansas City, MS, October 10th, 2017.

Leveraging Data Analytics to Improve Quality and Outcomes, speaker at Cerner Health Conference, Kansas City, MS, November 16th, 2016.

Patents

Kay, B., Rousey, J., Sagrillo, K., *Graphical User interface for the display of census information*— Patent in Review US, EU Design Number:002740035-0001, Granted; July 2015

Kay, B., Rousey, J., Sagrillo, K., *Graphical User interface for the display of hours worked*— Patent in Review US, EU Design Number: 002725184-0001, Granted; July 2015

Kay, B., Rousey, J., Badillo, J., Jens, K., Sagrillo, K., *Graphical User interface for the display of date last worked*— Patent in Review US, EU Design Number:002740373-0001, Granted; July 2015