Paper 5171-2020

Machine Learning Approach to Combat the Opioid Epidemic

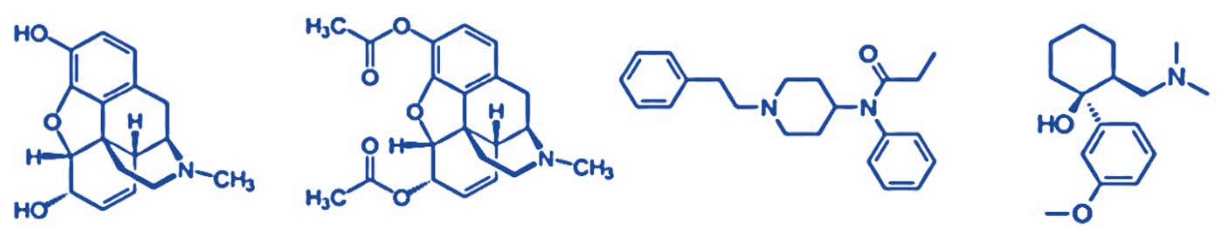
Author: Navjot Kaur, Oklahoma State University
Co-Author: Goutam Chakraborty, Oklahoma State University
Miriam Mcgaugh, Oklahoma State University

ABSTRACT

The ever-increasing practice of using opioids or street drugs in the United States has caused the rates of mortality from drug abuse to hit the roof. Although prescribed opioids are mainly used for pain relief, there is a prevalence of illegal use of opioids and the likelihood of becoming dependent on an opioid long-term spike after just five days of use. Unfortunately, the epidemic has affected almost every age group in every U.S. population, but the opioid addiction has been disproportionately affecting older adults living in rural America. The impact is so serious that every day more than 130 Americans die from an opioid overdose. In this study, we are going to examine whether there is any correlation between the prescription of opioids and deaths from overdose. Further, our objective is to find correlations among risk factors and identify statistically significant factors in individuals or groups who are susceptible to opioid abuse. Finally, we aim to develop a risk model to predict a patient's risk of opioid abuse or death from future opioid use.

INTRODUCTION

Opioids are the class of drugs that are used for moderate to severe pain treatment. Opioids occur naturally or can be produced in labs and are usually classified into three categories: Natural, Semi-Synthetic and Synthetic. Natural opiates such as opium, morphine, and codeine, etc. are alkaloids that are a member of chemical compounds containing nitrogen and usually derived from opium poppy plants. Semi-Synthetic opioids are the man-made opioids that are chemically related to opiates and are created in labs from natural opiates. Examples of semi-synthetic opioids include hydrocodone, oxycodone, oxymorphone, and hydromorphone.



Morphine (Occurs naturally) | **Heroin** (Semi-Synthetic) | **Fentanyl** (Synthetic) | **Tramadol** (Synthetic)

Figure 1: Chemical Structures of different kinds of Opioids

Synthetic opioids are a class of drugs that are not chemically related to opiates but emulate their properties. Methadone, fentanyl, dextropropoxyphene, tramadol, for example, are some of the synthetic opioids.

Opioids work by interacting with opioid receptors to block pain and stimulate endorphins that produce feelings of well-being and pleasure. Over time, opioids cause the body to slow its natural endorphin development, which causes drug dependence. Furthermore, as the body builds tolerance, increasing doses of opioids are required to create the same impact and thus, leading to addiction.

The roots of the epidemic of opioids incubated back in the 1990s, when there was a strong focus on improving pain management and the aggressive advertisement of OxyContin (a drug that belongs to opioid anesthetics) from 1996 to 2001 influenced physicians to prescribe opioids for complicated pain treatments. Pharma companies not only misrepresented the risk of opioid addiction but also promoted opioids as one the best solution for chronic pain treatments. Consequently, it led to an increased usage of opioids in the pain market and the liberalization of the use of opioids has increased the abuse of all opioids. As a result, drug overdose mortality rates owing to opioid abuse have dramatically increased by 2002 and it has become the primary cause of accidental deaths in the United States. Additionally, the crisis has devastating effects on physical, social and economic dimensions.

PROBLEM STATEMENT

The opioid epidemic is multidimensional and there are many reasons behind it. Health care providers, hospitals, pharmaceutical companies, and the people themselves who are using pain killers without prescription or who are misusing prescription, all are responsible for the opioid crisis. According to the U.S. Department of Health and Human Services, the opioid crisis continues to devastate communities across the country and below are some of the alarming statistics:

i. More than 130 people died every day from an opioid-related drug overdose.
ii. 10.3 million people misuse prescription opioids in 2018.
iii. 47,600 died from overdosing opioids in 2018.
iv. 2 million had an opioid overdose in 2018.
v. 2 million people misused prescription opioids for the first time.
vi. An estimated 40% of opioid overdose deaths involved a prescription opioid.
vii. The economic burden of prescription opioids contributes to $78.5 billion a year.

The disturbing consequences of opioids alarmed the government and the opioid crisis was declared a public health emergency in 2017. However, in spite of all the attention placed on the opioid epidemic, results have not been delivered at the required speed and the nation is still under critical situation. The reason is that the problem neither has an easy fix nor it can be handled by a single-layered strategy due to its multidimensional structure. Moreover, economic, social, and geographic disparities among the abusers make it difficult to tackle the problem through traditional methods. Hence, it has become imperative to remain vigilant to discover ways to combat the problem through the use of data and analysis using machine learning techniques.

DATA COLLECTION AND PREPARATION

Data Sources

The data sets for this analysis were acquired from CDC, CMS HCUP, HHS and Data World. The details of datasets along with their sources are described as following:

Drug Overdose or Mortality related data: The data file includes information about the provisional drug overdose death as per the National Vital Statistics System (NVSS).The datafile 'VSRR_Provisional_Drug_Overdose_counts.csv' was used mainly for descriptive analysis and was attained from https://data.cdc.gov/NCHS/VSRR-Provisional-Drug-Overdose-Death-Counts/xkb8-kh2a

Multiple causes of deaths: The Multiple Cause of Death (1999-2017) data available on CDC WONDER (https://wonder.cdc.gov/mcd.html) are county-level national mortality and population data. Data are based on death certificates for U.S. residents by age groups (single year age cohorts, 5-year age groups, 10-year age groups, or infant age groups), race (4 groups), ethnicity, sex, state, county, underlying cause of death and multiple cause of death, urbanization, year and month of death, weekday of death, place of death, and autopsy status.

Medicare Part D Opioid Prescriber Summary File 2017: The Centers for Medicare & Medicaid Services (CMS) has prepared a public data set, the Medicare Part D Opioid Prescriber Summary File, which presents information on the individual opioid prescribing rates of health providers that participate in Medicare Part D program. This file is a prescriber-level data set that provides data on the number and percentage of prescription claims (includes new prescriptions and refills) for opioid drugs and contains information on each provider's name, specialty, state, and ZIP code. This summary file was derived from CMS at https://data.cms.gov/Medicare-Part-D/Medicare-Part-D-Opioid-Prescriber-Summary-File-201/sakz-a2rp

Inpatient Sample: This dataset was extracted from the National (Nationwide) Inpatient Sample (NIS) database that contained all-payer inpatient care. The data sample for the year 2017 was obtained for predictive analysis in three different files:2017 NIS Inpatient Core Files,2017 NIS Disease Severity Measures Files and 2017 NIS Hospital Files

The NIS files contain clinical and resource-use information that is included in a typical discharge abstract, with safeguards to protect the privacy of individual patients, physicians, and hospitals (as required by data sources). It contains clinical and nonclinical data elements for each hospital stay, including:

i. International Classification of Diseases, Tenth Revision, Clinical Modification/Procedure Coding System (ICD-10-CM/PCS) diagnosis, procedures, and external cause of morbidity codes beginning October 1, 2015
ii. Patient demographic characteristics (e.g., sex, age, race, median household income for ZIP Code)
iii. Hospital characteristics (e.g., ownership)
iv. Expected payment source
v. Total charges
vi. Discharge status
vii. Length of stay
viii. Severity and comorbidity measures

For detailed elements of the files, see references [3] for NIS File Structures.

OSU CHSI datasets: The Center for Health Systems Innovation (CHSI) at Spears School of Business at Oklahoma State University provided us the opioid patient data sample. The dataset was divided into five categories:

i. Diagnosis: It contained information related to diagnosis such as Diagnosis Type, Diagnosis Code, Diagnosis Description, Diagnosis Priority, Patient Condition

ii. Procedure: This data table contained procedure code (ICD-9-CM or ICD-10-PCS), procedure name (MetabolicPanelBasic, AlcoholScreenUrine, etc), procedure groups details reported for opioid patients
iii. Patient demographic data: census region,bed_size,medical speciality, care setting, marital status,acute_status,gender,age,patient type description, etc.
iv. Clinical events: The dataset records clinical events (blood pressure systolic, heart rate, respiration rate, etc), description of events (pain assessment, chief complaint, etc.) and clinical event results and clinical event normalcy level (normal, critical, panic high, etc.)
v. Medications: It contains details about medication details like dosage_unit, name of medicine, generic name, drug duration, volume, etc.

Data Cleaning

Most of the files contained data of all drugs. As this study was related to Opioids, therefore, Opioid-related data were extracted from all files. Prescription rates and opioid overdose death rate files were merged on the basis of counties.

The datasets provided by OSU CHSI were merged using patient encounter_id and irrelevant columns like column repeating in multiple files, lab results like heartbeat rate, _systolic were discarded. Normalcy level (whether the test results from the lab test are normal, critical-high, critical-low) of lab_results were taken into account to define the condition of the patient.

ICD-10-CM diagnoses codes were also used to identify diseases like chronic pain, alcohol use disorder, etc. and to identify opioid users. From primary to secondary there were 40 columns (I10_DX1-I10_DX40) for diagnoses. However, from I10-DX11-I10DX40, very few rows contained data. Therefore, to identify if the patient is diagnosed with any type of disease, I10-DX10 columns were used. In other words, 1 principal and 9 secondary diagnoses of the patients were used to identify the particular disease. Similarly, there were 40 columns for procedure codes. The 10 (1 principal and 9 secondary) procedure codes were selected. NIS-Core-Patient, NIS-Hospital and NIS-Severity files were merged using the given primary key (NIS_KEY).

A cohort of opioid patients was extracted from the merged inpatient sample. The severity file contained all patient refined drug and below two columns that were used to extract the nominal target variable:

Column Name	Column Details	Coding Notes
APRDRG_Risk_Mortality	Risk of Mortality Subclass	0) No class specified (1) Minor likelihood of dying, (2) Moderate likelihood of dying, (3) Major likelihood of dying, (4) Extreme likelihood of dying
APRDRG_Severity	Severity of Illness Subclass	(0) No class specified, (1) Minor loss of function (no comorbidity or complications), (2) Moderate loss of function, (3) Major loss of function, (4) Extreme loss of function

Table 1. Column details

A nominal target variable 'risk_mortality' was calculated as

i. High Risk: Major or extreme likelihood of dying and major or extreme loss of function
ii. Medium: Moderate likelihood of dying and moderate loss of function
iii. Low Risk: No or Minor likelihood of dying and minor or no loss of function

METHODOLOGY

Broadly, the methodology is defined in three phases:

1. Descriptive Analysis: It was performed to determine the key metrics that would be the decisive points of the study. This included exploratory analysis by examining the trends of recorded deaths due to various types of opioids and by analyzing the data of various age-groups who were impacted by this epidemic. The state-wise mortality data owing to drug abuse was analyzed too.
2. Correlation Analysis: Various studies claimed that physicians, insurance companies, pharma companies and even the individuals who are misusing the prescription, are all responsible for fueling up the crisis. The new study in JAMA Network Open Drug states that drug companies have been spending billions of dollars to market their products to doctors, and other prescribers with speaking fees, free dinners, paid trips, and more. This paper focused to find the correlation between:
    i. Prescription rate vs drug overdose deaths
    ii. Approved Medicare opioid claims vs the number of prescriptions
3. Predictive Analysis: Along with physicians who are over-prescribing opioids and the insurance companies that are promoting prescription opioids, there are various other factors at an individual level that may have been contributing to the rise of the epidemic. For instance, lack of awareness on the effective use of these drugs, high consumer demand for drugs, broken health monitoring program and unavailability of predictive tools to predict the abuser of the prescribed drug. Therefore, a predictive analysis was performed to construct two different models.
    i. The first model was constructed to identify significant factors leading to opioid dependence and to predict opioid abusers. For this model, a binary target variable "opioid-oud" and various clinical and demographical independent variables of all patients (opioids and non-opioids) were used to build the model.
    ii. If the patient has been identified as an opioid-user or opioid-dependent, a risk stratification model was constructed to classify opioid patients into three classes: high-risk, medium-risk and low-risk. For this model, clinical, medical, demographical variables were accessed to classify the patients as per their risk level.

All the analysis was performed using Base SAS®, SAS® Studio, and SAS® Enterprise Miner™.

ANALYSIS

Descriptive Analysis

To begin with, we analyzed the drug overdose death count dataset to determine the cause of death and to determine various kinds of opioids that are resulting in deaths across the nation.

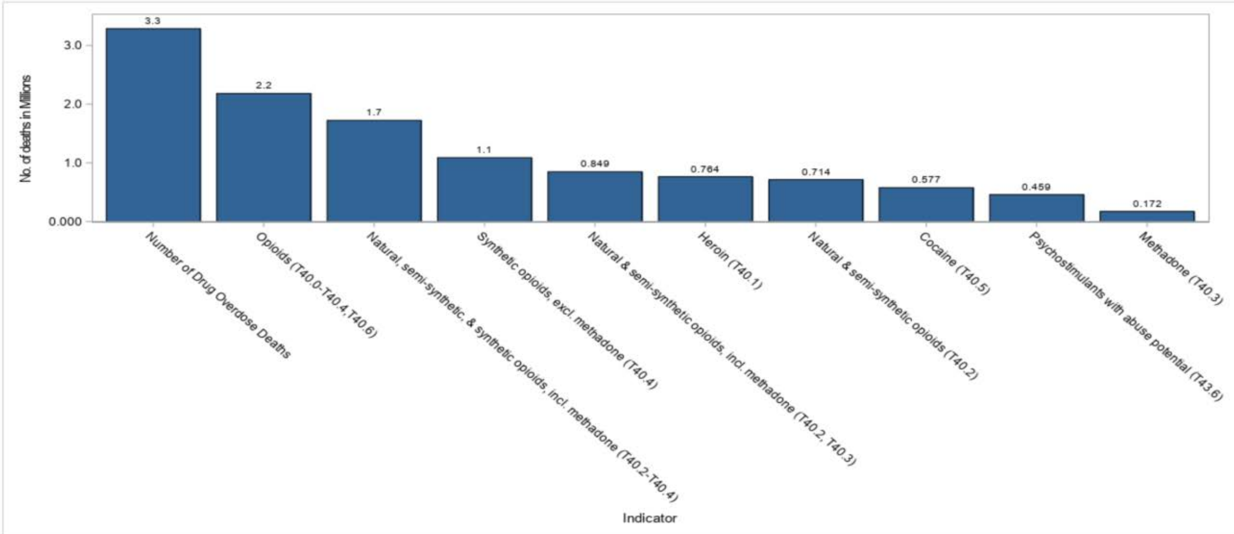


Figure2: Total number of deaths due to various drugs (2015 to 2018)

Note: Groups were not mutually exclusive as deaths could be caused by one or more than one drug.

As per the analysis, the total number of drug overdose deaths was 3.3 million for the time period of 2015 to 2018.Of these:

i. Opioids were detected in 2.2 million reported deaths.
ii. All kinds of opioids, including methadone, were detected in 1.7 million reported deaths.
iii. Synthetic opioids alone contributed to 1.1 million deaths
iv. Heroin and Natural cum Semi-synthetic drugs both resulted in over 0.7 million deaths.
v. Cocaine, Psychostimulants, and Methadone, all combined lead to over 1.2 million deaths

Then, we analyzed state-wise mortality rates owing to drug abuse.

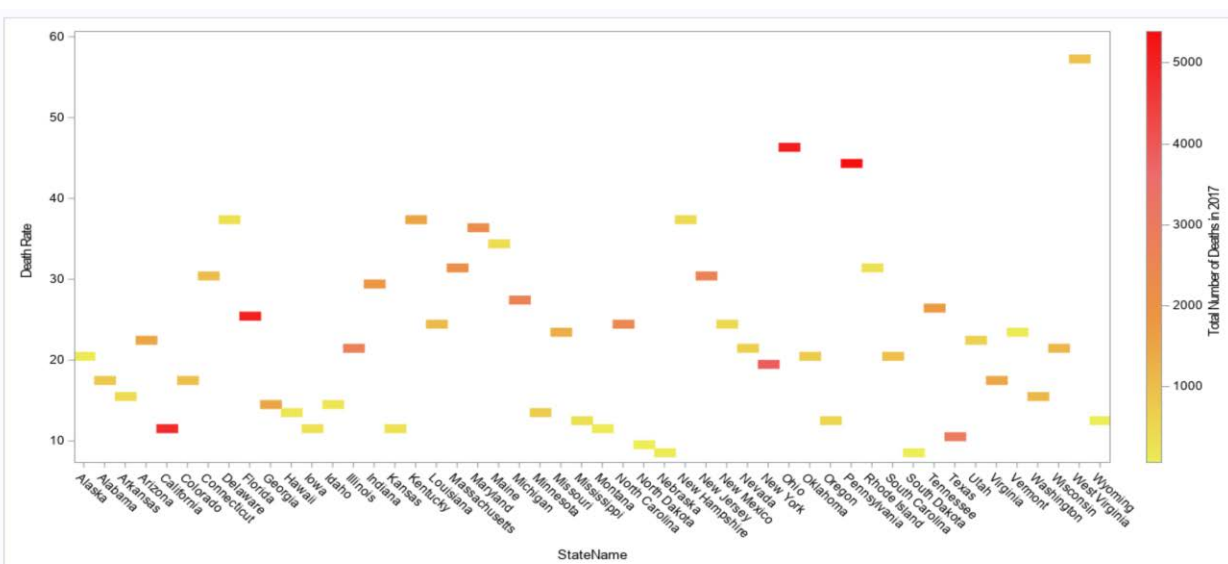


Figure 3: 2017-Drug Overdose Mortality by State

In 2017, Pennsylvania, Ohio, Florida, California, and New York were found to be the top five states where the total number of drug overdose deaths have been high. Nonetheless, counting the number of deaths would result in partial comparison due to the variance in size and population among states. So, another measure named 'death rate' was calculated as the number of deaths per 100,000 total population for an unbiased evaluation.

As per death rate measurement, West Virginia was found to have the highest death rate of 57.8. Ohio and Pennsylvania were the next states where death rates were 46.3 and 44.3, respectively. Kentucky, New Hampshire, Delaware too had a high recorded death rate of 37 for every 100,000 people.

Further, the impact of the opioid crisis across various age groups was analyzed.

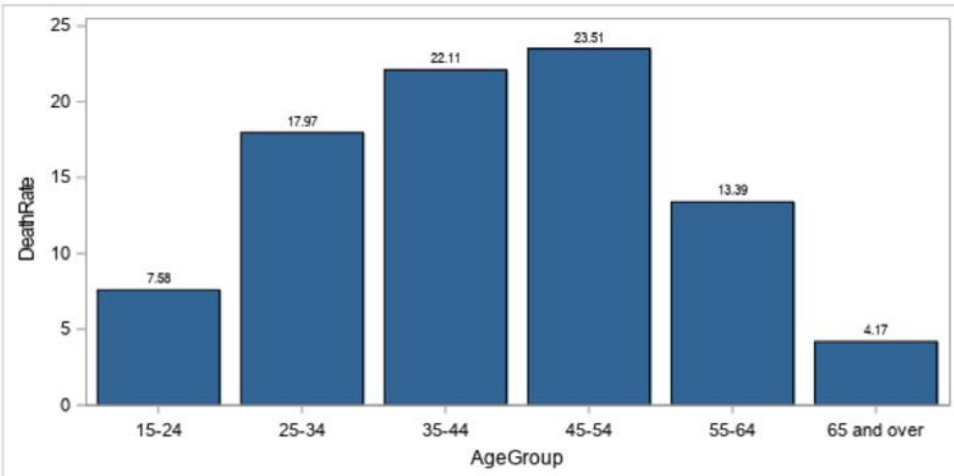


Figure 4: 2017-Drug Overdose Mortality by Age Groups

The opioid epidemic has affected older adults the most. People from 45 years old to 54 years old have the highest death rate of 24. The people from 35years old to 44 years old to have a high death rate of 22.

It was inferred from the descriptive analysis that the epidemic has impacted all age groups throughout the country. Further, the correlation analysis was conducted to understand the role of prescribing rate and insurance claims.

Correlation Analysis: Prescription rate vs drug overdose deaths

County-level prescription rates and county-level opioid overdose death rate files were combined and the Pearson Correlation test was performed to determine the association between the increased number of prescriptions and drug overdose deaths.

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations	
	OpioidPrescribingRate
DrugOverdoseMortalityRate DrugOverdoseMortalityRate	0.32649 0.0425 2513

Figure 5: Pearson Correlation Test

A positive linear correlation was found between Opioid Prescribing and Opioid Overdose Mortality Rates. So, when the number of written prescriptions rise, an increase in drug overdose deaths was associated with it. This indicated that over-prescription can be considered as one of the factors fueling the opioid crisis.

Correlation Analysis: Approved opioid claims vs prescription rate

It is widely believed that the crisis was accelerated by Insurance companies. As the number of approvals for the opioid prescription increased, so did the profit. Again, the Pearson Correlation test was performed to check whether there was any correlation between the number of approved opioid Medicare claims and the number of prescriptions. Medicare_PartD_Prescribing file for 2017 was used for this analysis.

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations	
	OpioidPrescriptionCount
OpioidClaimCount	0.44493 <.0001 16777

Figure 6: Correlation between Approved Medicare Claims and Opioid Prescriptions

A moderate positive linear relationship was found between the approved claims related to opioid medication and the number of opioids related prescribed medicines. This indicates that whenever the number of prescriptions increased, the number of approved claims also increased. In other words, opioid claims can be easily approved.

Although the outbreak had been exacerbated by over-prescription and readily accepted insurance policies, patient-level causes such as individuals who misused medication have played a similar role in the crisis. Most of the previous studies focused on controlling the prescription, however, the opioid use disorder may vary significantly among individuals. And, once the opioid abuse or dependence develops, treatment of patients becomes difficult as the tolerance level or potential overdose increases over time. Hence, it has become imperative to predict factors that could put individuals at risk of developing prescription drug use disorders.

Predictive Model1: Identify factors leading to opioid abuse

To evaluate risk factors and significant factors in individuals or groups who are more vulnerable to opioid abuse as compare to others, a cohort of 142174 patients and 100000 non-opioid patients were extracted from the inpatient sample. The dependent variable 'OUD' (opioid use disorder) was calculated such that if the patient was being diagnosed as an opioid-related disorder, then 1 was assigned to the variable. Nearly equal strata were drawn from the NIS list of patients being diagnosed for other conditions. For the non-opioid patients, 'OUD' was marked as 0. The clinical variables like 'presence of diabetes', 'alcohol disorder' etc. and demographic variables like 'Age of the patient', 'Region', 'Income' etc. were selected as the independent variables to determine the effect of the clinical and demographical characters on opioid dependence.

Further, data exploration was done to detect the potential problems in the dataset such as too many levels of categorical variables, highly skewed interval variables, etc. that could inversely impact prediction performance. We found there were some categorical variables such as 'I10_DX2',' I10_PR1', etc. that contained too many overall levels. Second, extremely skewed distributions of predictors such as 'Age',' 'I10_NDX' etc. were also detected. To overcome these problems, a transformation that had the best association with the target variable was determined and applied.

In the dependent variable('opioid-oud'), 54:46 distribution of the target variable was observed, So, no resampling (under-sampling or oversampling) was required. Almost all independent variables (except 'income level' with 2% missing) have more than 99% non-missing values. So, no imputation to handle missing was required. Data were partitioned into a 70:30 ratio for training and validation respectively. Four candidate models were selected: Decision Tree, Gradient Boosting-SVM, and HP Forest for predicting patients who are vulnerable to opioid abuse.

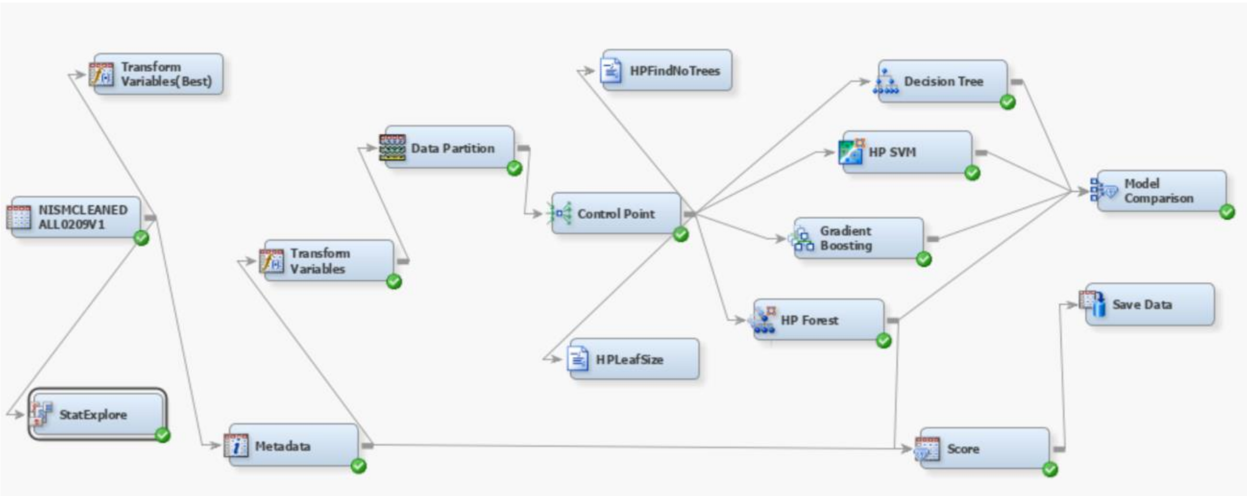


Figure 7: Model diagram in SAS® **Enterprise Miner™**

All the candidate models were trained and validated before model comparison to select the final model.

Model Assessment

The final model was selected on the basis of validation misclassification rate (the probability of being misclassified), sensitivity (the ability to correctly identify those who were identified as opioid-dependent) and ROC index.

S. No	Model	Misclassification Rate	Sensitivity
1	HP Forest	09.18%	91.10%
2.	Gradient Boosting	10.96%	88.56%
3.	Decision Tree	14.70%	73.63%
4	HP SVM	18.50%	77.58%

Table 2: Models Comparison

It was observed that the misclassification rate of the HP Forest was the least among all candidate models. Additionally, the sensitivity was the highest for the HP forest.

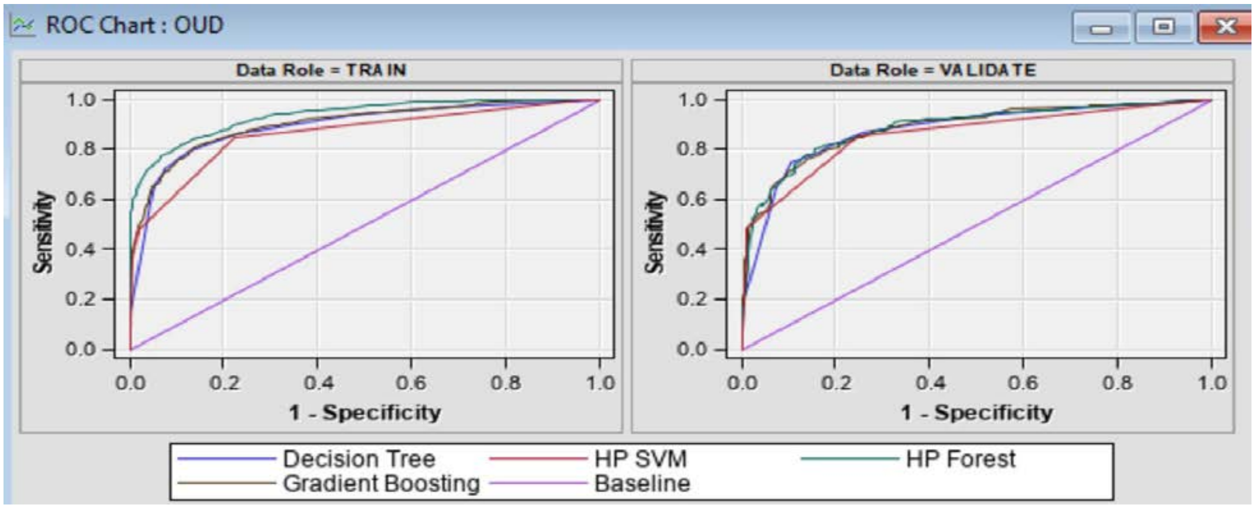


Figure9: ROC Chart from model comparison

The area under the curve (AUC) of the receiver operating characteristic (ROC) was optimum for the HP forest indicating that HP Forest was the best among four candidate models. The HP forest model had 91% accuracy and its true positive rate to correctly identify the opioid patients that are at high risk of opioid dependence was 91%.

Statistically significant factors

The below table shows variables that are statistically significant in predicting risk factors leading to opioid abuse.

Variables	No. of Splitting Rules	Train: Gini Reduction
No of drugs used	2647	0.0176
Type of opioid	2493	0.0177
Major Diagnostic Category	2109	0.0290
Major depressive disorder	2017	0.0136
Antidepressant	1660	0.0119
Alcohol	1377	0.0066
Age	849	0.1027
Benzodiazepine Use	845	0.0079
Gender	705	0.0342
Anticonvulsant	600	0.0225
Corticosteroid	587	0.0031
Chronic Pain/Back Pain	581	0.0106
Migraine	449	0.0033
Hepatitis C	417	0.0131
Median Household Income	393	0.0160
Rural or urban areas	248	0.0024
Heart Disease	110	0.0007
Race	33	0.0002

Table 3: Variable Importance

As per the above table, the number of drugs used (opioids or non-opioids) used by a person is the most significant factor. It was followed by a type of opioid and major diagnostic category. The use of antidepressants, alcohol abuse, benzodiazepine, anticonvulsants, and corticosteroids was found to be statistically significant factors leading to opioid dependence. Medical conditions: chronic or severe back pain, migraine, Hepatitis C, major depressive disorder were leading causes of opioid dependence.

Demographic variables: age, median household income, rural or urban location, gender, and race were statistically significant factors leading to opioid abuse. Males were found to be more vulnerable to opioid abuse as compared to females. Patients with median household income between $1-$43,000, were at the highest risk as compared to people with higher income groups. Rural area patients were found to be more vulnerable than people in urban areas. Native Americans and Caucasians were found to be more exposed to opioid dependence as compared to any other category.

Model 2: Risk Stratification Model

For risk stratification model, various clinical (number of diagnoses, type of drugs used, number of drugs used, etc.), medical (primary diseases along with opioid dependence, major disorders, etc.), and demographical variables (age, income group, race, etc.) were used as independent variables. The target variable was 'risk_mortality' with three classes: high-risk, medium-risk, and low-risk.

During the data exploration, some potential problems presented in the dataset like extremely skewed predictors, too many overall categorical values, were eliminated by the transformation that had the best association with the target variable. Then the data was partitioned into 70:30 split for training and validation. Multiple logistic Regression, Neural Network, HP Regression, and Decision tree were trained and validated to maximize the performance of models.

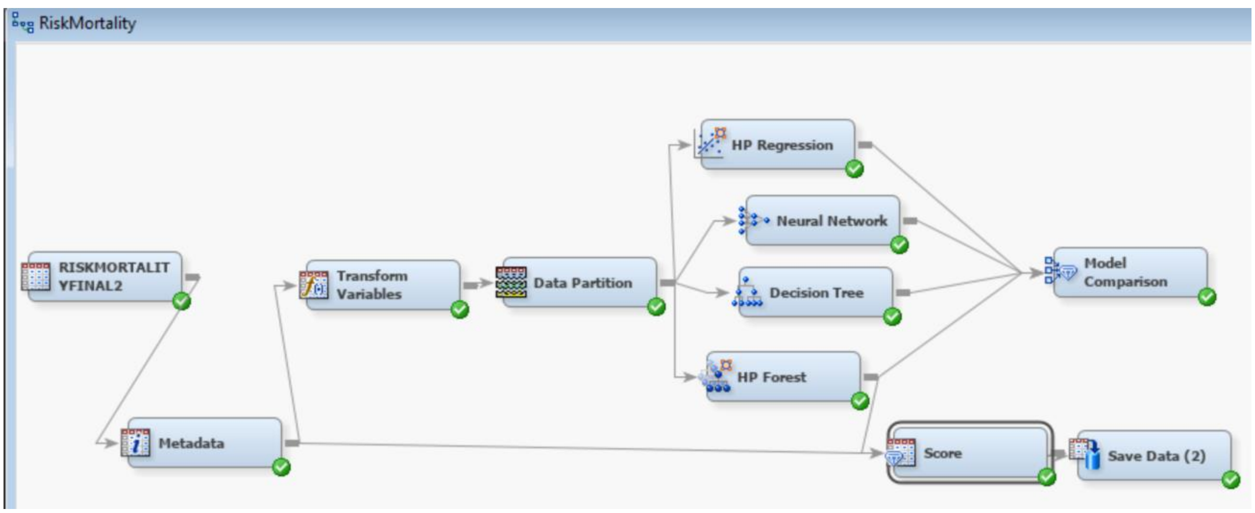


Figure8: Model2 SAS® **Enterprise Miner™**

The winning model was selected on the basis of classification accuracy, ROC Index, and averaged F1 score. HP Regression was the winning model among all selected models.

Model Assessment: Risk Stratification Model

The candidate models were assessed on the basis of the misclassification rate and ROC Index. The averaged precision, recall and F1 score were also recorded.

Sr.No	Model	Misclassification Rate	ROC Index
1	Hp Forest	12.60%	97.3
2.	Neural Network	15.97%	93.8
4	HP Regression	16.19%	81.0
3.	Decision Tree	16.21%	87.6

Table 4: Model Comparison

Hp Forest with the highest accuracy and highest Roc index was selected as the best model for classification. The averages precision, recall and F1 score of HP Model were 0.7,0.85 and 0.77 respectively

The below table describes the statistically significant variables of the model.

Variables	No. of Splitting Rules	Train: Gini Reduction
No of diagnosis	28906	0.0513
Drug version	27534	0.0089
Average dose-MME	25844	0.0203
AGE	18572	0.0090
No of ED visits	14839	0.0033
Opioid Duration	13171	0.0061
Sepsis	12698	0.0018
Respiratory Problem	12411	0.0306
Hypertension	12304	0.0052
Kidney Disease	11566	0.0066
Chronic Disease	10781	0.0011
Major Depressive Disorder	10045	0.0012
No of Procedures	9866	0.0013
Benzodiazepines	9519	0.0157
Median Household Income	8916	0.0008
Hepatitis C	8308	0.0005
Urban Rural Code	8183	0.0016
Bipolar Disorder	7211	0.0007
Alcohol Disorder	6435	0.0003
Other Non-Opioid Drugs	6099	0.0002
Race	5997	0.0004
Benzodiazepines	3526	0.0004
Antidepressants	2592	0.0001

Table 5: Important Variables

The following table shows a summary of the significant variables across the three risk groups.

Risk_Mortality	Variables	Mean	Std. dev
3-High Risk	No. of diagnosis Average dose MME No. of Procedures Age Medical Conditions	18.23 129.57 3.87 49.10 Alcohol Disorders, Major depressive disorder, Acute kidney disease, Chronic disease, Hepatitis, Hypertension, Sepsis	5.97 16.55 1.36 17.03
2-Medium Risk	No. of diagnosis Average dose MME No. of Procedures Age Medical Conditions	15.11 87 1.034 53.3 Alcohol disorder, Bipolar disorder Hypertension, Acute pain	5.45 4.89 2.07 16.42
1-Low Risk	No. of diagnosis Average dose MME No. of Procedures Age Medical Conditions	6.30 20 3.67 38.56 No substance disorder, Minor/Acute pain	5.97 8.97 2.16 12.48

Table 6: Risk Groups Summary

After scoring and data exploration, it was found that along with high-risk clinical parameters, people with lower median household income, native American, white American at the highest risk. Most of the patients with high risk have used opioids for more than 90 days.

People with medium income groups were at moderate risk. Hispanic Asian or Pacific Islanders were at moderate risk. Most of the medium-risk patients have used opioids for more than 30 days.

People with no history of substance disorder and who use the opioid with a prescribed limit are at low risk. People with lower risk have used opioids for less than 30 days.

IMPLEMENTATION

The above models can be used by doctors or prescribers as individual risk management measures for opioids before they administer opioids for any type of treatment. The opioid overuse disorder model can be helpful in predicting patients who are vulnerable to opioid overdose and therefore, the model's results could be administered to vulnerable patients upon an initial visit prior to beginning opioids for pain management. This approach will be helpful to reduce morbidity and mortality associated with opioid abuse.

The risk stratification model can be integrated into a clinical method to easily identify high-risk patients and provide physicians with real-time pre or post prescription decision-making alerts. This model can enable clinicians to take the required action and measures as per the risk level of the patient. For instance, recognize high-risk patients could warn the physician that the patient could need urgent care.

ACKNOWLEDGMENTS

The authors of this paper would like to acknowledge Dr. Zhuqi Miao, and Dr. William Paiva from the Center for Health Systems Innovation at Oklahoma State University, for their helpful guidance and support to understand the medical terms and for providing us the required data.

CONFLICT OF INTEREST

This work was conducted with data from the Cerner Corporation's Health Facts database of electronic medical records provided by the Oklahoma State University Center for Health Systems Innovation (CHSI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Cerner Corporation.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Navjot Kaur
MS in Business Analytics
Oklahoma State University
nkaur@okstate.edu

Dr. Goutam Chakraborty
SAS Professor of Marketing Analytics
Director of Masters in Business Analytics
Oklahoma State University
goutam.chakraborty@okstate.edu

Dr. Miriam Mcgaugh
Clinical Professor, Business Analytics
School of Marketing and International Business
Oklahoma State University
miriam.mcgaugh@okstate.edu

REFERENCES

1. Opioid Data Analysis and Resources at https://www.cdc.gov/drugoverdose/data/analysis.html
2. Data Sources:
   CDC: https://wonder.cdc.gov/controller/datarequest/
   https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html
   HCUP: https://www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp
   HHS: https://healthdata.ov/
   Data World: https://data.world/datasets/opioids
   Opioid & Health Indicators Database: https://opioid.amfar.org/

3. NIS File Structures . Available at
NIS Core File : https://www.hcup-us.ahrq.gov/db/nation/nis/tools/stats/FileSpecifications_NIS_2017_Core.TXT
Hospital File : https://www.hcup-us.ahrq.gov/db/nation/nis/tools/stats/FileSpecifications_NIS_2017_Hospital.TXT
Severity File : https://www.hcupus.ahrq.gov/db/nation/nis/tools/stats/FileSpecifications_NIS_2017_Severity.TXT

4. International Classification of Diseases, Tenth Revision at https://www.icd10data.com/ICD10CM/Codes

5. OSU-CHSI: https://business.okstate.edu/chsi/

6. Wide-ranging online data for epidemiologic research (WONDER). Atlanta, GA: CDC, National Center for Health Statistics; 2016. Available at http://wonder.cdc.gov.

7. Ending America's Opioid Crisis | The White House at https://www.whitehouse.gov/opioids/ https://www.whitehouse.gov/opioids/

8. HHS Acting Secretary Declares Public Health Emergency to Address National Opioid Crisis https://www.hhs.gov/about/news/2017/10/26/hhs-acting-secretary-declares-public-health emergency-address-national-opioid-crisis.html

9. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and Opioid-Involved Overdose Deaths – United States, 2013-2017. WR Morb Mortal Wkly Rep. ePub: 21 December 2018