Extracting the Characteristics of Opioid Use Disorder Patients from Clinical Notes by Natural Language Processing

Khashayar Eshtiaghi, Alex Vajiac

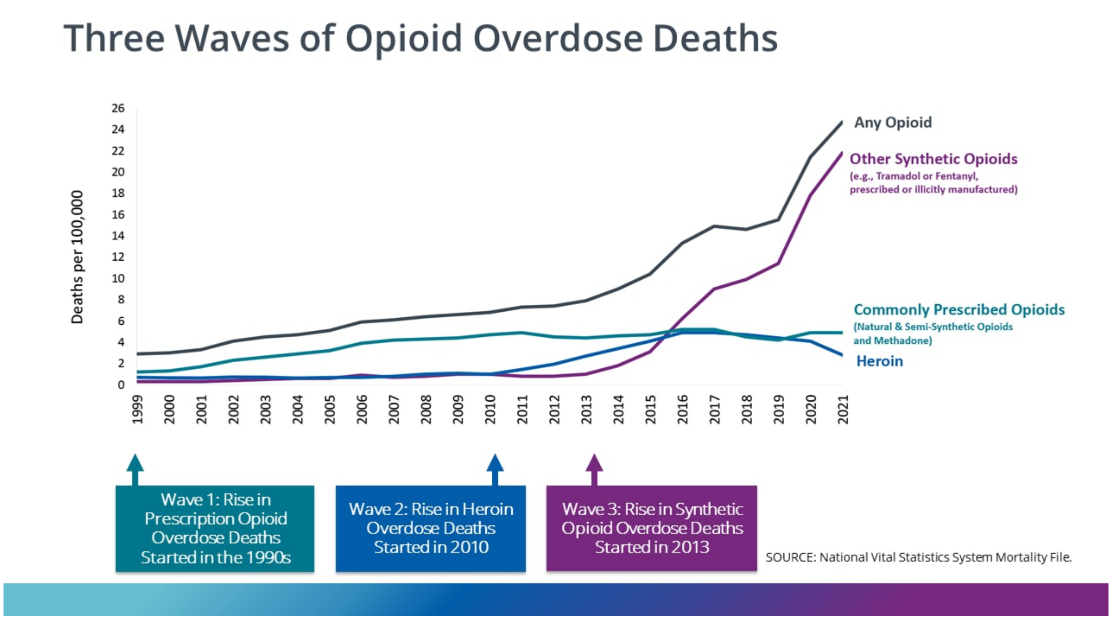
CS-670: Natural Language Processing

Spring, 2024

Background

Opioid use disorder has emerged as a major public health crisis, affecting millions of people worldwide and imposing significant social and economic burdens. The United States, in particular, has been deeply impacted, especially in the past few decades where there have been vast increases in the number of deaths due to opioid overdoses involving prescription opioids, heroin and fentanyl. Over 2.7 million individuals in the US struggle with opioid use disorder (OUD), and opioid overdose deaths in the US have nearly quadrupled from around 21,000 in 2010 to over 80,000 in 2021, which further highlights the urgency of finding research strategies to mitigate the consequences of this deadly epidemic. [1]

There are 3 major waves of increases in opioid overdose deaths in the USA. Beginning in the 1990s and 2000s, there was a steady increase in opioid overdose deaths, closely followed by an increase in heroin overdose deaths that lasted until around 2018 due to patients with opioid use disorder resorting to heroin as a way to continue to fuel their opioid addiction. However, the sharpest increase in opioid overdose deaths is attributed to deaths from fentanyl, and starting in 2013, opioid overdose deaths increased more than ever in history due to this. [2] Since both opioid and non-opioid substances are often adulterated with fentanyl, and because fentanyl is one of the most lethal opioids in history, being 50 times more deadly than heroin [3], from 2013-2021 opioid overdose deaths increased exponentially. This further reinforces the fact that it is imperative that we must research and act to combat this epidemic urgently.



Project Goals and Methodology

To help develop further strategies and policies for combating the opioid epidemic, it is imperative that we find a deeper understanding of opioid use disorder and characteristics associated with OUD patients. A key resource that can be used to understand these characteristics is electronic health records (EHRs), which are computerized databases that store medical histories, clinical information and administrative data associated with hospital patients. For this project, we used natural language processing techniques to obtain key information from the clinical notes data in the MIMIC database, which is a freely available database containing anonymized EHRs from patients of the BID Medical Center in Boston, Massachusetts.

In this project, the primary goals were to identify the most common feelings, most common comorbid disorders, and most common medications prescribed to patients with opioid use disorder, as well as to predict whether a patient has opioid use disorder based on their clinical notes and to determine which features contribute the most to identifying whether or not a patient has opioid use disorder. To do this, we first extracted the data using SQL queries to access the clinical notes from the MIMIC dataset, specifying the ICD codes (codes used by healthcare professionals to classify diagnoses, symptoms and procedures for claims processing) to filter out patients both with and without opioid use disorder. Then, we cleaned and preprocessed the clinical notes and identified key terms collected from academic resources to analyze the distribution of feelings, disorders, analgesic medications, psychiatric medications, withdrawal management medications, and adjunctive and supportive medication associated with opioid use disorder patients. Finally, we visualized the outcomes to compare patients with and without OUD and used the Naïve Bayes, Logistic Regression and Random Forest classifiers to predict whether or not a clinical note belongs to a patient with opioid use disorder.

Data Preprocessing

To preprocess the data, we first used regular expressions to remove extra whitespace, strip trailing spaces, replace newline characters with spaces, and replace '|||' with new lines. We also removed sequences of three or more underscores or asterisks. Then, we tokenized the text (i.e., divided the text into an array of words to make data analysis easier) using the word_tokenize function in Python's Natural Language Toolkit (NLTK) and filtered out stopwords, which are words in the English language with little to no meaning, as defined in the

NLTK corpus. Then we provided 5 lists of characteristics as reference. Those list are collected with the help of experts in this field. To accurately identify which words exist in the text and how frequently they occur, we used NLTK's WordNet Lemmatizer, an NLTK class with a lemmatize function to find the lemma or stem of a word, ensuring that variations such as misspellings, differences in grammatical tense, differences in uppercase and lowercase letters, spacing, and other discrepancies do not hinder word identification. We then created columns in the dataset for patient feelings, disorders, and analgesic, psychiatric, withdrawal management, and adjunctive and supportive medications by extracting the keywords for those categories from the lemmatized set in each clinical note to use in the key term analysis.

Figure 1: An example of a clinical note from the MIMIC database

```
[4]: data['aggregated_notes'][2]

[4]: 'Admission Date:  [**2120-2-18**]              Discharge Date:   [**2120-2-21**]\n\nDate of Birth:  [**2088-3-2**]             Sex:   F\n
     \nService: CARDIOTHORACIC\n\nAllergies:\nReglan / Prochlorperazine\n\nAttending:[**First Name3 (LF) 492**]\nChief Complaint:\ninability sl
     eep, racing thoughts, visual hallucinations,\nparanoia\n\n\nMajor Surgical or Invasive Procedure:\nNone\n\nHistory of Present Illness:\nTh
     e pt is a 31 yo F with a PMH of bipolar disorder brought in by\nfamily for paranoia and inability to sleep for 7 days. Pt is\nunable to gi
     ve a reliable history but reports that she has not\nslept for the past 7 nights and that she believed that someone\nwas putting up securit
     y cameras in her home. The patient\nconfronted her parents about this and was told that there was no\none in the house. She also reports t
     hat she has not been eating\nwell over the past week. She denies feeling like her thoughts\nare racing, denies increased energy, denies im
     pulsive behavior.\nPer pt\'s sister and cousin, pt has been unable to care for\nherself for a number of years. She is currently living wit
     h her\nparents but her parents left for vacation 1 wk ago and since\nthat time, pt has not slept or attended to ADLs. Denies SI/HI.\n.\nIn
     the ED, initial vs were: T98.9 HR96 BP113/73 RR18 O299RA\nShe was evaluated by psychiatry and was Sect 12. She was found\nto have an eleva
     ted tylenol level to 38.5 as well as urine tox\npostive for benzos and opiates. She was started on NAC for\nelevated tylenol level. Prior
     to admit to floor, the patient\nbecame tachycardic, dry and flushed and was given valium. She is\nbeing admitted to the ICU for hourly CIW
     A scales.\n\n\nPast Medical History:\nChronic bronchitis\n? thrush\nh/o kidney stones\nBipolar Disorder - her family reports that she has
     a long hx of\nsevere depression and bipolar d/o, and also was abusing\nprescription medications for many yrs. They report that while\nshe
     was living in\n[**State 4565**], she had multiple ER visits with various psychiatric\ncomplaints, but was never admitted. In [**Month (onl
     y) 216**], she attended a 30\nday detox/rehab program for her prescription medication abuse.\n\nSocial History:\nPt was born and raised in
     [**Location (un) 86**] area. She has 1 sister. [**Name (NI) **]\nbeen living with her parents. Per pt\'s family, pt has not been\nable to
     support herself for a number of yrs.\n\n\nFamily History:\nBiological father with EtOH abuse and other "psychiatric issues"\n- details unc

[5]: len(data['aggregated_notes'][2])

[5]: 132043
```

Figure 2: Keywords identified with patients with opioid use disorder by category (feelings, analgesic medications, psychiatric medications, withdrawal management medications, comorbid disorders, adjunctive and supportive medications)

```
[12]: feelings_list

[12]: ['Anxious',
       'Depressed',
       'Vulnerable',
       'Isolated',
       'Hopeless',
       'Stigmatized',
       'Desperate',
       'Withdrawn',
       'Fearful',
       'Distrustful',
       'Ashamed',
       'Guilty',
       'Fragile',
       'Needy',
       'Confused',
       'Agitated',
       'Overwhelmed',
       'Unstable',
```

```
[14]: analgesics_list

[14]: ['Acetaminophen',
       'Ibuprofen',
       'Celecoxib',
       'Tramadol',
       'Ketorolac',
       'Lidocaine',
       'Lidocaine (topical)',
       'Aspirin',
       'Naproxen',
       'Diclofenac',
       'Indomethacin',
       'Piroxicam',
       'Meloxicam',
       'Etodolac',
       'Sulindac',
       'Tolmetin',
       'Oxaprozin',
       'MefenamicAcid',
```

```
[15]: psychiatric_list

[15]: ['Sertraline',
       'Fluoxetine',
       'Paroxetine',
       'Citalopram',
       'Escitalopram',
       'Venlafaxine',
       'Duloxetine',
       'Desvenlafaxine',
       'Levomilnacipran',
       'Milnacipran',
       'Diazepam',
       'Lorazepam',
       'Clonazepam',
       'Alprazolam',
       'Temazepam',
       'Olanzapine',
       'Quetiapine',
       'Risperidone',
```

```
[16]: withdrawal_management_list

[16]: ['Clonidine',
       'Lofexidine',
       'Methadone',
       'Buprenorphine',
       'Naltrexone',
       'Suboxone',
       'Suboxone (buprenorphine/naloxone)',
       'Vivitrol',
       'Vivitrol (naltrexone injection)',
       'Probuphine',
       'Probuphine (buprenorphine implant)',
       'Buprenorphine patch',
       'Naloxone',
       'Gabapentin',
       'Baclofen',
       'Hydroxyzine',
       'Ondansetron',
       'Mirtazapine',
```

```
[13]: disorders_list

[13]: ['HIV',
       'AIDS',
       'Hepatitis',
       'Hepatitis B',
       'Hepatitis C',
       'Tuberculosis',
       'Skin infections',
       'Abscesses',
       'Endocarditis',
       'Septicemia',
       'Depression',
       'Anxiety',
       'Bipolar',
       'Schizophrenia',
       'PTSD',
       'Post-traumatic stress disorder (PTSD)',
       'Personality',
       'Stroke',
```

```
[17]: adjunctive_and_supportive_list

[17]: ['Multivitamins',
       'Calcium',
       'Iron',
       'VitaminD',
       'VitaminC',
       'VitaminB12',
       'Folic',
       'Folic acid',
       'Zinc',
       'Zinc supplements',
       'Magnesium',
       'Magnesium supplements',
       'Potassium',
       'Electrolyte solutions',
       'Potassium supplements',
       'Omega3',
       'Omega-3',
       'Omega-3 fatty acids',
```

Figure 3: Code used to preprocess the text using the word_tokenize function and extract keywords using the wordnet lemmatizer, and the resulting preprocessed dataset used for the project

```python
def preprocess_text(text):
    # Import required libraries
    import re
    import nltk
    from nltk.tokenize import word_tokenize
    from nltk.corpus import stopwords

    # Remove extra whitespace and strip leading/trailing spaces
    text = re.sub('\s+', ' ', text).strip()

    # Replace newline characters with spaces and '|||' with newlines, then strip leading/trailing spaces
    text = text.replace('\n', ' ').replace('|||', '\n').strip()

    # Remove sequences of three or more underscores
    text = re.sub(r'_{3,}', '', text)

    # Remove sequences of three or more asterisks
    text = re.sub(r'\*{3,}', '', text)

    # Tokenize the text into words
    tokens = word_tokenize(text)

    # Get the set of English stopwords
    stop_words = set(stopwords.words('english'))

    # Filter out the stopwords from the tokens
    filtered_tokens = [token for token in tokens if token.lower() not in stop_words]

    # Return the list of filtered tokens
    return filtered_tokens
```

```python
def extract_words(text, word_set):
    # Filter out non-alphabetic words from the text
    alpha_words = [w for w in text if w.isalpha()]

    # Create an instance of WordNetLemmatizer
    wnl = nltk.WordNetLemmatizer()

    # Lemmatize the alphabetic words and convert them to lowercase
    alpha_words_lem = [wnl.lemmatize(w.lower()) for w in alpha_words]

    # Return a set of words that are present in the given word_set
    return set(word for word in alpha_words_lem if word in word_set)
```

```
data
```

	aggregated_notes	OUD condition	Preprocessed Notes	Patient Feelings	Patient Disorders	Patient Analgesics Medications	Patient Psychiatric Medications	Patient Withdrawal Management Medications	Patient Adjunctive and Supportive Medications
0	Admission Date: [**2169-8-28**] ...	0	[Admission, Date, :, [, *, *, 2169-8-28, *, *,...	{unstable, lethargic, angry, restless, agitate...	{thrombosis, hepatitis, dizziness, anxiety, ki...	{aspirin, ibuprofen, fentanyl, morphine}	{midazolam}	{}	{multivitamin, thiamine, potassium, aspirin, f...
1	Admission Date: [**2154-12-31**] ...	0	[Admission, Date, :, [, *, *, 2154-12-31, *, *...	{withdrawn, confused, sensitive, anxious, unst...	{endocarditis, abscess, schizophrenia, hyperka...	{morphine, acetaminophen, lidocaine, aspirin, ...	{haloperidol, mirtazapine}	{haloperidol, hydroxyzine, ondansetron, mirtaz...	{multivitamin, zinc, senna, iron, magnesium, b...
2	Admission Date: [**2120-2-18**] ...	1	[Admission, Date, :, [, *, *, 2120-2-18, *, *,...	{ashamed, anxious, embarrassed, restless, depr...	{bipolar, insomnia, anemia, hyperthyroidism, b...	{acetaminophen, ibuprofen, lidocaine}	{haloperidol, diazepam, lorazepam, escitalopram}	{haloperidol, diazepam, lorazepam, escitalopram}	{folic, thiamine, metoclopramide, senna}
3	Admission Date: [**2145-8-6**] D...	0	[Admission, Date, :, [, *, *, 2145-8-6, *, *, ...	{unresponsive}	{thrombosis, pneumonia, gerd, vascular, stroke...	{fentanyl}	{}	{}	{potassium}
4	Admission Date: [**2121-4-30**] ...	1	[Admission, Date, :, [, *, *, 2121-4-30, *, *,...	{unresponsive, sensitive, suspicious}	{migraine, arrhythmia, dizziness, miscarriage,...	{}	{}	{}	{}
...	...	...	...	...	...	...	...	...	...
1995	Admission Date: [**2140-5-23**] ...	1	[Admission, Date, :, [, *, *, 2140-5-23, *, *,...	{lethargic, distant, agitated}	{obesity, schizophrenia, thrombosis, hepatitis...	{methadone, hydromorphone}	{lorazepam}	{methadone, lorazepam}	{}
1996	Admission Date: [**2121-11-21**] Discha...	0	[Admission, Date, :, [, *, *, 2121-11-21, *, *...	{passive, irritable, struggling, hesitant}	{depression, pneumonia, hepatitis}	{}	{}	{}	{iron}

Key Term Analysis

We have found significant differences in patient feelings, comorbid disorders, as well as analgesic, psychiatric, withdrawal management, and adjunctive and supportive medications between patients with and without opioid use disorder. Regarding feelings, opioid use disorder patients are significantly more likely to feel agitated, suspicious, lethargic, unresponsive and angry compared to non-opioid use disorder patients (the control group). This suggests that opioid use disorder patients tend to be more mistrustful of authority as well as less energetic and more temperamental than average. In addition, regarding comorbid disorders, opioid use disorder patients are significantly more likely to have HIV, hepatitis, anxiety, cirrhosis and bipolar disorder than the control group. This suggests that people with bipolar disorder and anxiety are significantly more likely to resort to opioid use as a coping mechanism and might also suggest that opioid use disorder might cause additional anxiety in patients. Moreover, HIV, hepatitis and cirrhosis are medical problems that are often times strongly correlated with opioid use disorder due to HIV and hepatitis often times being spread by intravenous substance use, and since opioids can cause significant liver damage it is not surprising that opioid use disorder patients are significantly more likely to have cirrhosis.

Figure 4: Comparison of feelings for OUD vs. Non-OUD patients

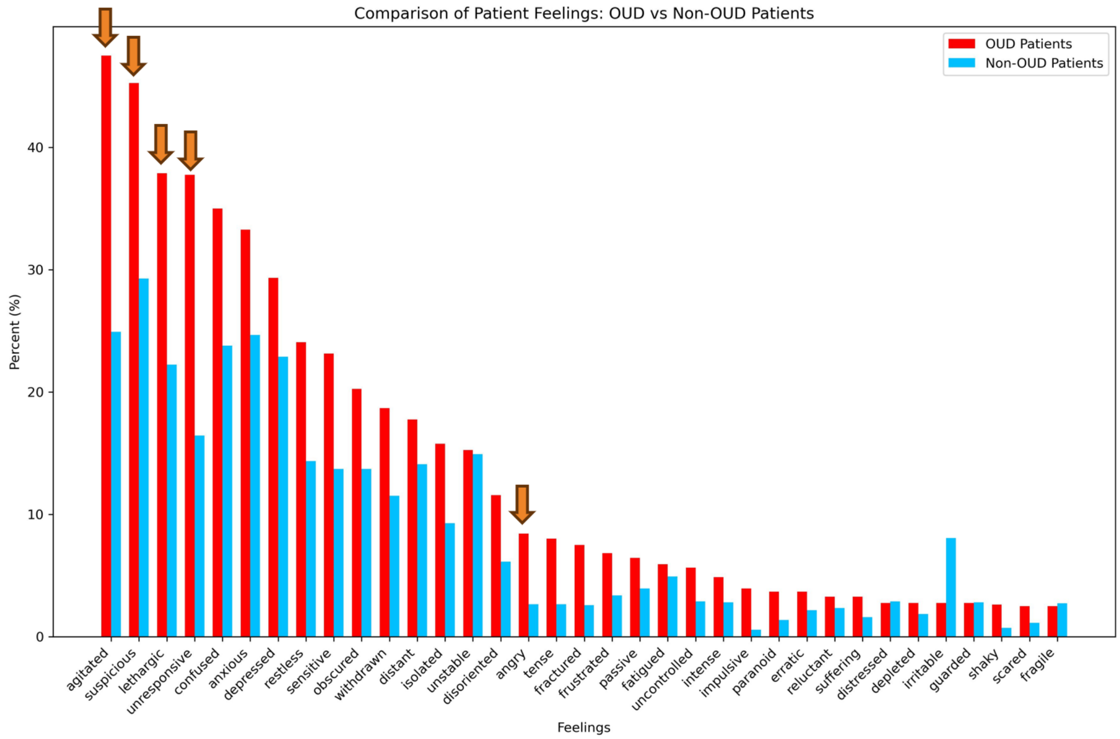
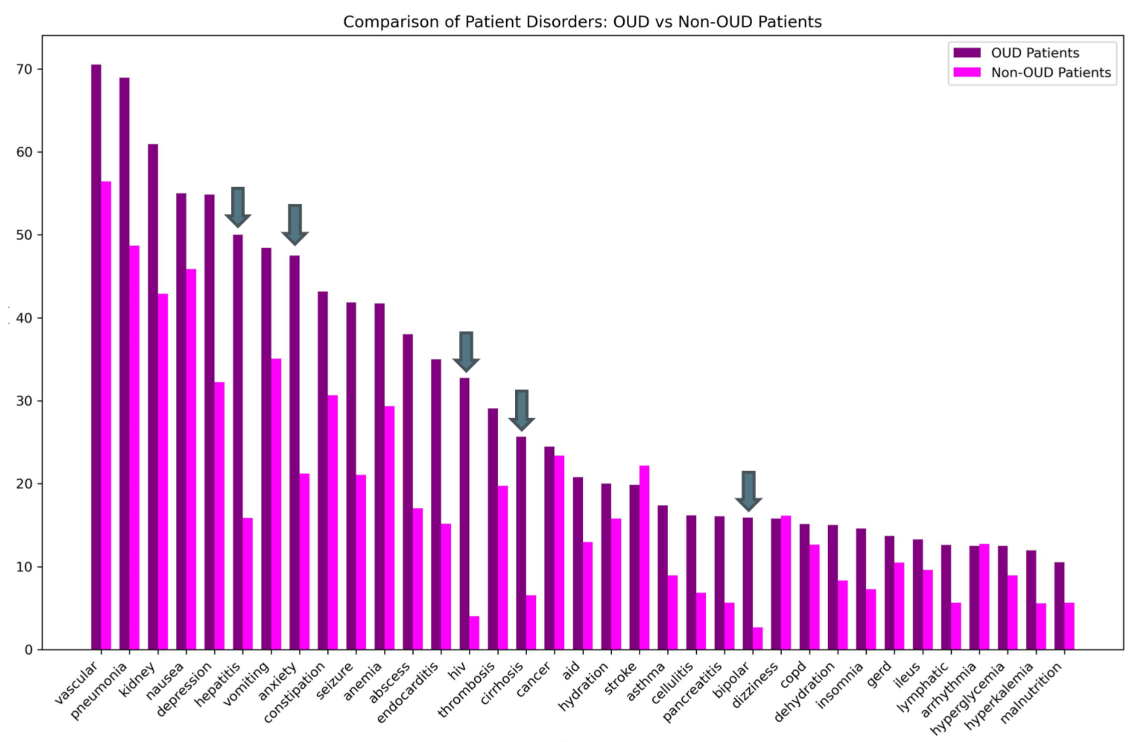
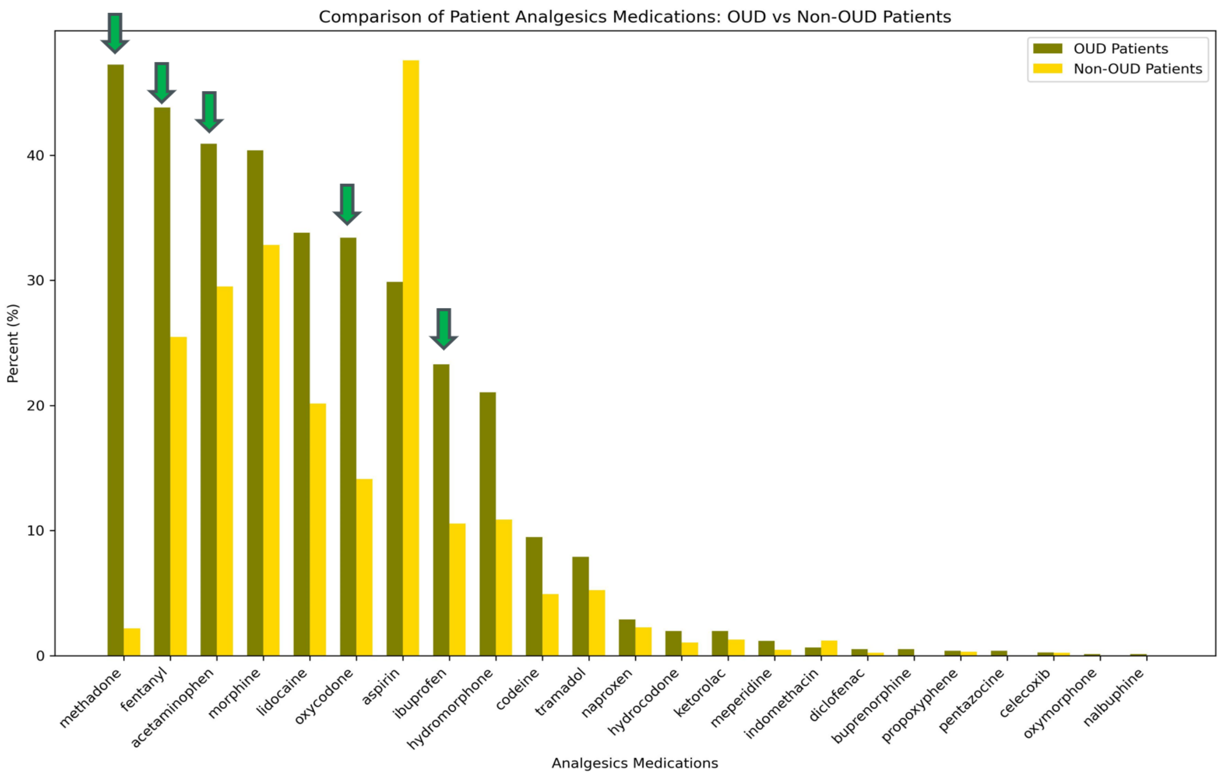


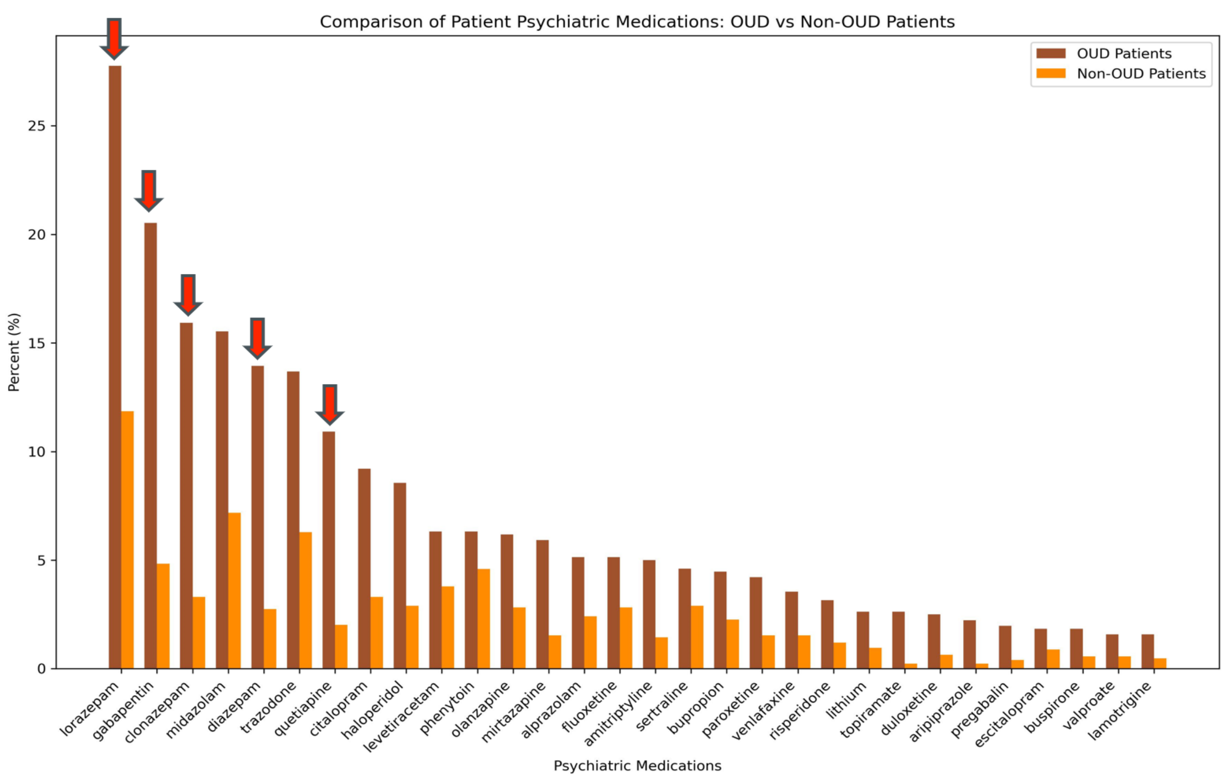
Figure 5: Comparison of comorbid disorders for OUD vs. Non-OUD patients

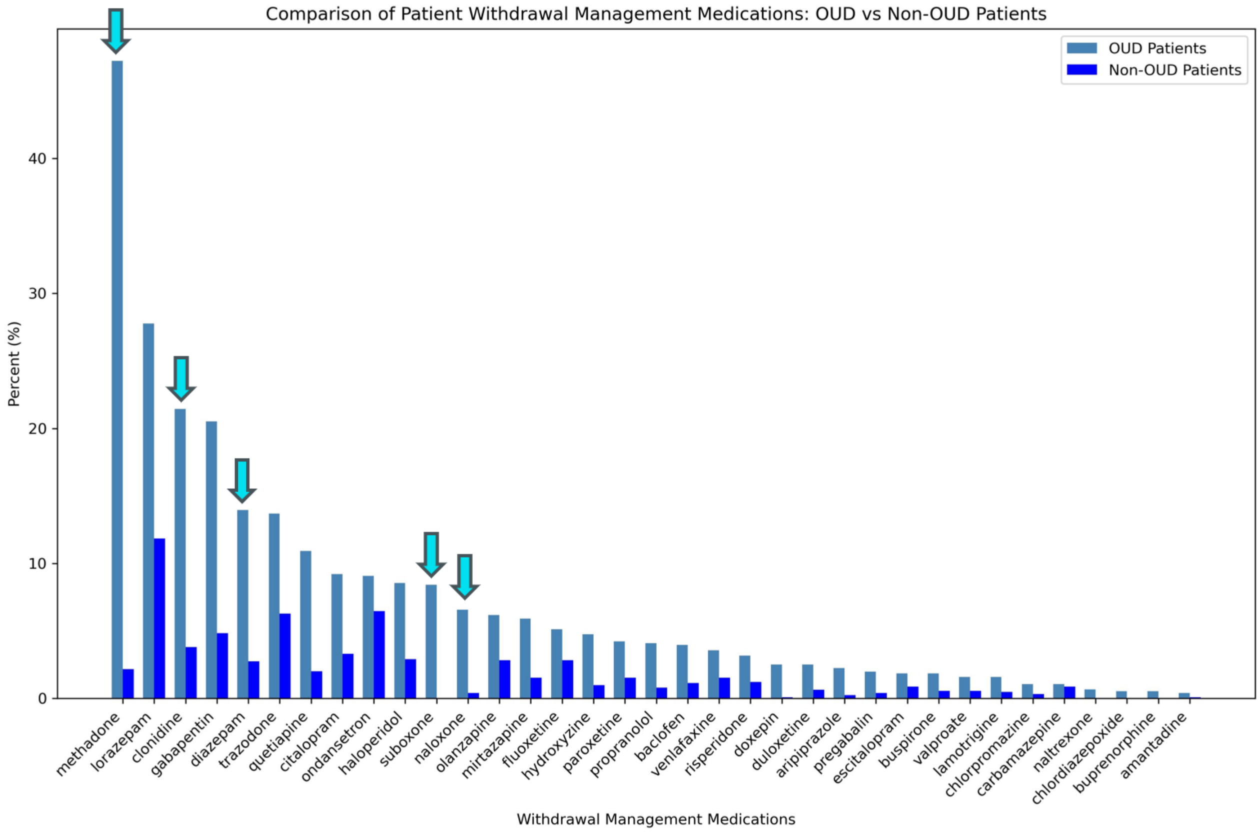


One of the most informative analyses regarding opioid use disorder treatment found in the MIMIC clinical notes databases is the analysis of the differences in prescribed medications between OUD and non-OUD patients. Regarding analgesic medications, opioid use disorder patients are significantly more likely to be prescribed methadone, oxycodone, fentanyl, ibuprofen and acetaminophen. This is because often times, opioid use disorder begins with doctors prescribing opioid medications (e.g. oxycodone, fentanyl) to patients usually for pain relief, and ibuprofen and acetaminophen are often times combined with opioid medications in many prescription opioid drugs (e.g. Percocet, Norco). Opioid use disorder then develops as a result of misuse of these medications. Regarding withdrawal management medications, opioid use disorder patients are more likely to be prescribed methadone, clonidine, diazepam, suboxone and naloxone. This makes sense since naloxone is often times used to treat opioid overdoses and suboxone, clonidine and methadone are often prescribed to ease opioid withdrawal symptoms. An interesting fact to note is that methadone is found both in the analgesic medication list and the withdrawal management medication list, which makes sense since methadone is sometimes used in analgesic settings as well. For psychiatric medications, opioid use disorder patients are more likely to be prescribed gabapentin, clonazepam, diazepam, lorazepam and quetiapine, which makes sense since these medications are generally prescribed to treat anxiety-related symptoms that are often comorbid with opioid use disorder. As aforementioned, diazepam is also on the top 5 list of withdrawal management medications, which hints that anxiety could possibly be a symptom of opioid withdrawal in addition to being comorbid with OUD. In addition, the top 5 adjunctive and supportive medications prescribed to opioid use disorder patients are thiamine, folic acid, lactulose, senna and multivitamins, which suggests that OUD patients are significantly more likely to have deficiencies in B vitamins, of which thiamine and folic acid are alternate nomenclatures for vitamin B1 and vitamin B9, respectively.

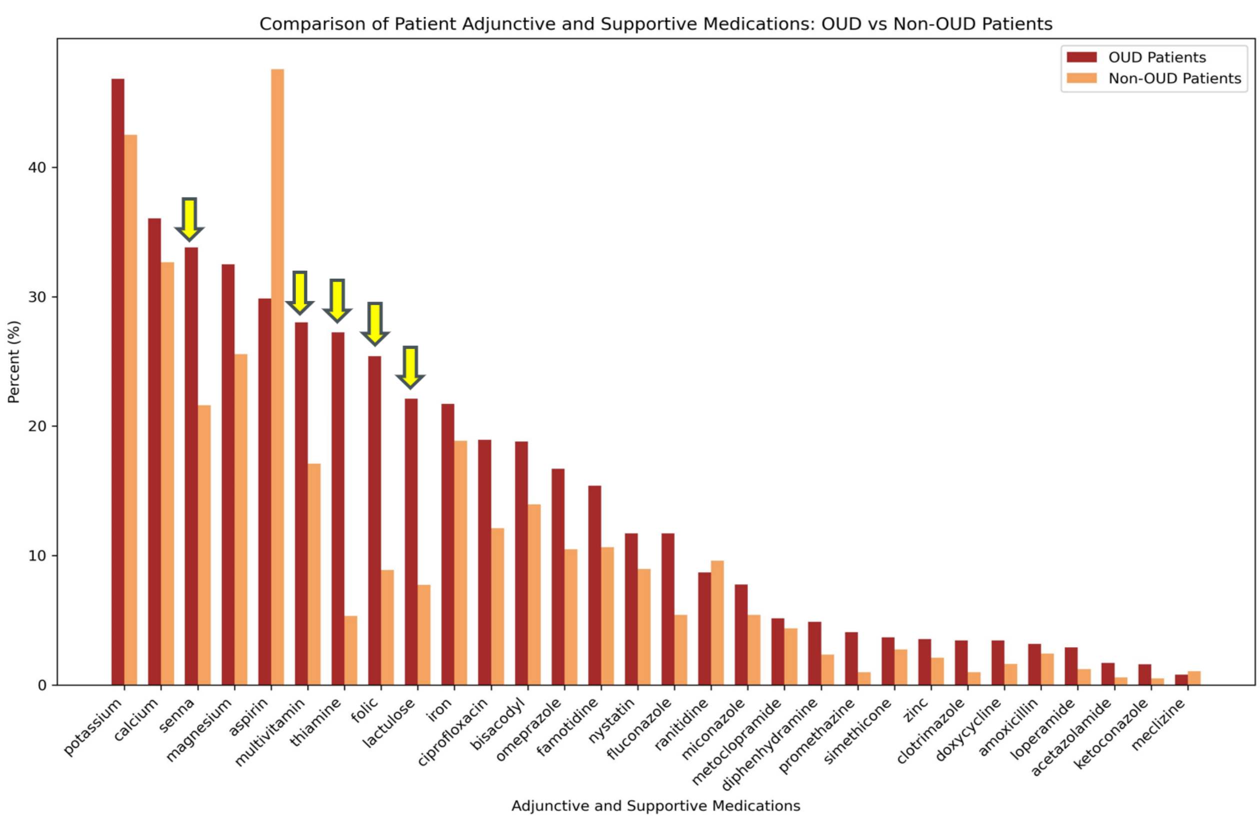
Figure 6: Comparison of analgesic, psychiatric, withdrawal management and adjunctive and supportive medications for OUD vs. non-OUD patients







Comparison of Patient Withdrawal Management Medications: OUD vs Non-OUD Patients
OUD Patients
Non-OUD Patients
Percent (%)
0
10
20
30
40
methadone
lorazepam
clonidine
gabapentin
diazepam
trazodone
quetiapine
citalopram
ondansetron
haloperidol
suboxone
naloxone
olanzapine
mirtazapine
fluoxetine
hydroxyzine
paroxetine
propranolol
baclofen
venlafaxine
risperidone
doxepin
duloxetine
aripiprazole
pregabalin
escitalopram
buspirone
valproate
lamotrigine
chlorpromazine
carbamazepine
naltrexone
chlordiazepoxide
buprenorphine
amantadine
Withdrawal Management Medications



Comparison of Patient Adjunctive and Supportive Medications: OUD vs Non-OUD Patients
OUD Patients
Non-OUD Patients
Percent (%)
0
10
20
30
40
potassium
calcium
senna
magnesium
aspirin
multivitamin
thiamine
folic
lactulose
iron
ciprofloxacin
bisacodyl
omeprazole
famotidine
nystatin
fluconazole
ranitidine
miconazole
metoclopramide
diphenhydramine
promethazine
simethicone
zinc
clotrimazole
doxycycline
amoxicillin
loperamide
acetazolamide
ketoconazole
meclizine
Adjunctive and Supportive Medications

Predictive Modeling

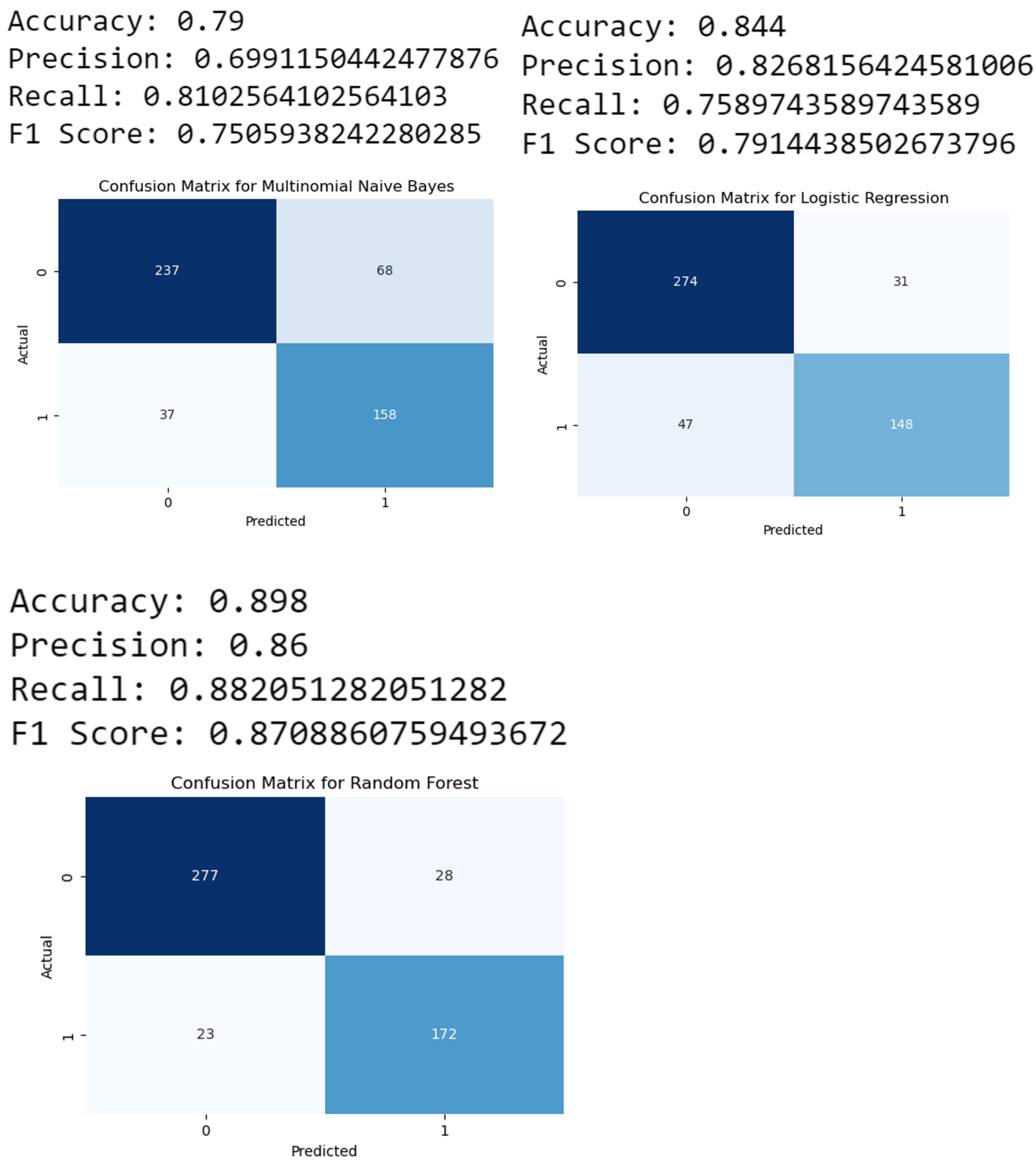
To predict whether or not a patient has opioid use disorder, we tested 4 classifiers: Naïve Bayes Classifier, Multinomial NB, Logistic Regression and Random Forest. Regarding accuracy, precision, recall and F1 score, Random Forest performed the best, with an accuracy of 89.8%, precision of 86%, recall of 88.2% and an F1 score of 87.1%. Logistic Regression performed second best regarding precision, F1 score and accuracy, Multinomial NB performed the third best regarding recall, which is the most important metric for medical research, with a recall score of 81%. Regarding interpretability, we have found that the Naïve Bayes classifier performed most effectively, as it highlighted the most important features to predict whether a patient has opioid use disorder. As expected from our previous analysis on key terms, our most informative features include the use of methadone and clonidine, which are medications that are commonly prescribed to opioid use disorder patients. Another interesting finding from the Naïve Bayes analysis is the strong relationship between opioid use disorder and PTSD (post-traumatic stress disorder). This is further evidenced by the fact that in our previous analysis on patient feelings and comorbid disorders we have found agitation, anger, suspicion, and anxiety to be common themes among OUD patients, and PTSD is often associated with these feelings and comorbid disorders.

Figure 7: Most informative features regarding classification of patients with OUD

```
classifier.show_most_informative_features()
```

```
Most Informative Features
        count(methadone) = 5                   1 : 0      =     22.2 : 1.0
          has(methadone) = True                1 : 0      =     18.6 : 1.0
            has(doxepin) = True                1 : 0      =     18.4 : 1.0
         has(topiramate) = True                1 : 0      =     17.3 : 1.0
        count(clonidine) = 4                   1 : 0      =     17.2 : 1.0
        count(methadone) = 6                   1 : 0      =     16.0 : 1.0
        count(methadone) = 3                   1 : 0      =     15.8 : 1.0
        count(methadone) = 1                   1 : 0      =     15.7 : 1.0
        count(methadone) = 8                   1 : 0      =     14.9 : 1.0
             count(ptsd) = 1                   1 : 0      =     13.1 : 1.0
```

Figure 8: Accuracy, precision, recall, and F1 score as well as confusion matrices for the Naïve Bayes, Logistic Regression and Random Forest classifiers



Conclusion

We have found that specific patterns can be identified in clinical notes belonging to patients with opioid use disorder regarding their feelings, disorders and medications that distinguish them from patients without OUD, and that some of the most important features in classifying whether or not a patient has OUD include methadone use, anxiety, bipolar disorder and PTSD. In addition, to treat opioid use disorder we have found that the most effective way is via medication-assisted treatment in which patients are prescribed both anti-anxiety psychiatric medications such as gabapentin and diazepam as well as opioid withdrawal management medications such as suboxone and methadone, and that vitamin B supplements play a role in managing opioid use disorder as well. All in all, these findings can be used to further improve diagnosis and treatment of opioid use disorder and can potentially save lives if applied in time to treat OUD patients.

Limitations and Further Steps

Limitations regarding this analysis include the need for more in-depth models that are tailored to analyzing medical notes, the need for more variables (e.g. blood pressure, BMI, history of heart disease, bilirubin) and possibly sentiment analysis to check the general health condition of patients both with and without opioid use disorder, and data sourced from more than one hospital to improve generalizability. Further steps that can be taken to address these concerns are joining other tables for the MIMIC database to fetch variables for important medical metrics, sourcing EHR data from other hospitals around the country, and developing a more advanced NLP model that is tailor-made to read medical notes as opposed to general text. Furthermore, regarding steps to combat opioid use disorder, it is imperative that underlying conditions are treated as well, so further analysis on patients with anxiety and PTSD would help prevent patients developing OUD before it is too late.

References

1. NIH. (2023). Drug Overdose Death Rates. Retrieved from https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates
2. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, August 8, 2023, https://www.cdc.gov/opioids/basics/epidemic.html
3. Fentanyl: One pill kills. Texas Health and Human Services. https://www.hhs.texas.gov/services/mental-health-substance-use/mental-health-substance-use-resources/fentanyl-one-pill-kills