Using Electronic Medical Records and Clinical Notes to Predict the Outcome of Opioid Treatment Program

Wanting CUI1, Fatemeh SHAH-MOHAMMADI and Joseph FINKELSTEIN
Icahn School of Medicine at Mount Sinai, New York, NY, USA

Abstract. Opioid addiction is a serious public health problem in the US, and this study aimed to explore how natural language processing (NLP) can be used to identify factors that contribute to distress in individuals with opioid addiction, and then use this information along with structured data to predict the outcome of opioid treatment programs (OTP). The study analyzed medical records data and clinical notes of 1,364 patients, out of which 136 succeeded in the program and 1,228 failed. The results showed that several factors influenced the success of patients in the program, including sex, race, education, employment, secondary substance, tobacco use, and type of residences. XGBoost with down sampling was the best model. The accuracy of the model was 0.71 and the AUC score was 0.64. The study highlights the importance of using both structured and unstructured data to evaluate the effectiveness of OTP.

Keywords. Machine Learning, Natural Language Processing, Unstructured text, Electronic Health Record

1. Introduction

Opioid use disorder is a growing public health concern in the US, leading to thousands of deaths and high financial costs [1,2]. Methadone and buprenorphine are effective treatments for opioid dependence that can be provided at opioid treatment programs (OTPs). Patient data at OTPs is recorded in both structured and unstructured formats, with unstructured clinical text notes containing a significant amount of information not found in structured data [3]. Natural language processing (NLP) is an effective method for extracting this information accurately and efficiently. Despite many studies being conducted on opioid addiction treatment and prevention, only a small number of studies have utilized information from clinical text notes [4,5]. Thus, using this information could potentially improve the accuracy of outcome prediction models. This study aims to explore how NLP can be utilized to identify factors contributing to distress in individuals with opioid addiction by analyzing their clinical text notes. This information, combined with structured data, could be used to predict OTP outcomes more accurately, which will help improve the efficacy of OTPs by providing a more accurate prediction.
understanding of the factors that contribute to distress and outcomes in individuals with opioid addiction.

2. Method

The data used in the study was obtained from the Opioid Treatment Program (OTP) of the New York State Office of Addiction Service and Supports (OASAS). The information collected pertained only to patients who underwent treatment at MSHS. It is a de-identified dataset that combined data from various sources, including electronic medical records (EMR), OPT admission and discharge records and notes from various healthcare providers. It contains 48,249 records and 31,685 unique patients. 9,511 patients have records of notes.

In a previous study [6], we developed a CLAMP NLP pipeline that identified areas of distress in patients’ notes. Thus, we optimized this pipeline to accommodate larger sample sizes and used it to count the number of times patients addressed their current distress in terms of mental problems, social problems, legal problems, health problems and family problems accordingly. We further defined a variable ‘overall distress’ by calculating the geometric mean of the 5 types of distress.

For the analytic dataset, we identified patients with valid admission and discharge records and patients with valid discharge status. We only included patients with complete information on age, race, sex, education, employment, primary substance of abuse, and criminal history. We used progress notes and patients were required to have a minimum of 5 notes. Notes from healthcare providers such as: social workers, physician assistants, nurses, medical doctors and vocational rehab counselors were included. After these criteria were applied, 1,364 patients were included in the dataset. To evaluate the effectiveness of the OTP, we defined those patients who “completed treatment: all treatment goals met” and patients who “completed treatment: half or more goals met” as success. We defined patients who didn’t complete the treatment with no goals met as failed. We further combined 13 variables which includes demographic information, socio-economic status, living arrangements, substance abuse history, tobacco usage and number of arrests from the structured EMR data and the 6 distress factors extracted from notes to use for analysis. We first performed descriptive analysis and then constructed machine learning models to predict the OTP outcome. We used XGBoost as our base model, because in the previous study tree-based models performed the best compared to logistic regression and SVM [7]. Due to the highly imbalanced nature of the dataset and the relatively small sample size for successful patients, we employed the down-sampling method. We further explored the sampling ratio. We performed parameter tunings and performed 3-fold cross validation to find the best parameters. The predictive dataset was split randomly into a 70% training set and a 30% testing set. To account for the imbalanced nature of the data, the Area under the ROC curve (AUC) was chosen as the evaluation metric. The best-tuned models for each algorithm were applied to the testing set, and their accuracy and AUC were calculated accordingly.

3. Results

There were 136 patients who succeeded in the program and 1,228 patients who failed the program that had medical and progress notes. According to table 1, identifying mental
problems, legal problems and health problems from clinical notes were significant factors in differentiate whether the patients will be successful in the treatment programs. Patients who failed the program talked more about mental problems and legal problems compared to the successful patients. In contrast, patients who were successful mentioned health problems more.

Table 1. Descriptive statistic of distress extracted from clinical notes.

<table>
<thead>
<tr>
<th></th>
<th>Fail (n = 1228)</th>
<th>Success (n = 136)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>std</td>
</tr>
<tr>
<td>Mental problem</td>
<td>7.33</td>
<td>8.73</td>
</tr>
<tr>
<td>Social problem</td>
<td>1.10</td>
<td>0.37</td>
</tr>
<tr>
<td>Legal problem</td>
<td>2.85</td>
<td>5.86</td>
</tr>
<tr>
<td>Health problem</td>
<td>6.69</td>
<td>8.10</td>
</tr>
<tr>
<td>Family problem</td>
<td>1.01</td>
<td>0.14</td>
</tr>
<tr>
<td>Overall distress</td>
<td>2.15</td>
<td>1.21</td>
</tr>
</tbody>
</table>

In categorical variables, sex, race, ethnicity, education, employment, secondary substance, tertiary substance, tobacco usage and type of residences were all significant factors that determined the successfulness of patients in the programs. The successful patients were 67.65% male and 32.35% female, while the failed patients were 75.98% male and 24.02% female. White patients made up 36.76% of the successful group and 31.51% of the failed group, while black patients made up 20.59% of the successful group and 24.59% of the failed group. Hispanic patients were 39.71% of the successful group and 46.01% of the failed group. Patients with high school degrees, patients who were employed and patients who were homeless but lived in shelters had higher proportions of success. Over 97% of patients enrolled in the program used Heroin as the primary substance and there was no significant different between the 2 subsets. However, patients with no other drug uses performed better than those with a secondary and tertiary abused substance.

Figure 1. Model performance based on different sampling ratio.

We further developed machine learning models to identify whether a patient would be able to complete the treatment successfully, so that healthcare providers could provide additional personalized help for individuals who needed them. The dataset was highly imbalanced. Less than 10% of patients had completed the program successfully, whereas 90% of patients did not complete the treatment. Thus, we under sampled the failed group at different sampling ratios in the training dataset and compared the performance of each model. According to Figure 1, as more unsuccessful samples were included in the training dataset, although the model accuracy increased, the AUC score suffered a
diminishing return. The models with high accuracy were categorizing all patients as unsuccessful. Thus, the best model was an XGBoost model with down sampling and the failed to success sample ratio was 6:5 (multiple = 1.2). 266 out of 367 of failed patients and 24 out of 43 success patients were predicted correctly. The accuracy of the model was 0.71 and the AUC score was 0.64.

4. Discussion

Despite having an imbalanced dataset, a tree-based model with down sampling was a feasible approach to predict the outcome of opioid treatment. By tuning the multiplier, keeping the training dataset relatively balanced and using AUC score as the performance matrix, the model was able to identify both successful and failed patients in the dataset. Moreover, using a tree-based model allowed researchers to determine the most important features that are associated with treatment success. The top 5 important features were substance usage, tobacco usage, gender, age and race. These features align with the results obtained from the descriptive analysis. However, we have very limited samples for patients who succeeded in the program while also having records of their notes. Thus, in future studies, we will aggregate information from more platforms to increase the number of samples of patients who complete the treatment successfully. We will explore alternative machine learning methods and sampling techniques to increase the model accuracy.

5. Conclusion

XGBoost with down sampling was the best model to predict the treatment outcome. In addition, the success of patients in the programs was found to be influenced by several significant factors, including mental, legal and health distress, sex, race, ethnicity, education, employment, secondary substance, tertiary substance, tobacco usage and type of residences. Thus, aggregating structured and unstructured data was a crucial step to predict the effectiveness of OTP.

References