CLASSIFICATION OF OPIOID PATIENTS BASED ON DRUG USAGE USING MACHINE LEARNING

**1Dr. Ch Raja Ramesh, 2 A Sumanjali, 3A Shreya, 4D SaiPraneeth, 5 Dr. P Muralidhar Sharma, Associate Professor**

**1Associate Professor, Department of Data Science, 2UG Scholar ,3UG Scholar ,4UG Scholar Students, Department of CSE, Vignan Institute of Institute of Technology and Sciences, Deshmukhi, Hyderabad.**

**Abstract:** There have been more drug overdoses in the past few years because of the opioid problem. A number of ways have been created to guess how often doctors will prescribe opioids. But because the problem is so complicated, these methods are not yet accurate enough. The way that opioid-dependent people are categorized from stable data sources needs to be reliable and honest. Most of the studies that have been done in the past have looked at the mental health of opioid users in connection to how much they use. Some of the newest deep learning techniques, like the attention and knowledge distillation mechanism, are not used in these works. These would help them get better results. This study classified drugs using machine and deep learning. MIMIC-III organized and unstructured data was used to identify purposeful and accidental opioid users. The 41 characteristics that are used are based on machine learning and deep learning to identify the drug users. We were able to accurately predict both intentional and accidental users from organized datasets better than from unstructured datasets. A good test accuracy is also reached by using a distilled body of knowledge that comes from combining the two machine learning algorithms CNN+LSTM. Ablation analysis gives us new opioid patient data for our investigation.

Keywords: opioid use disaster, MIMIC-III, logistic regression.

1. **INTRODUCTION**

Opioid medicines usually treat severe or chronic pain. Doctors prescribe drugs to many people, especially in the US. Opioids are a type of painkiller drug that is heavily overdone because it is easy to become addicted to. A number of studies have shown that people take these drugs not to ease their pain, but because they need them. An overdose can also happen this way. [16] In healthcare, DL and ML are becoming more common. But the present opioid risk assessment tools aren't good enough when it comes to being able to predict what will happen and automatically analyzing the patient's past data. Also, doctors should be given tools that let them figure out how likely it is that a patient will abuse pills before giving them to them. ML can help investigate opioid misuse, a medical issue that impacts people and businesses. In our study, we use ML to forecast the misuse of opioids based on both organized and unstructured data. Previous research by Barkley and Shin [5] discovered a higher prevalence of depression among individuals who intentionally overdose on opioids. Other studies [6, 7] indicated that the number of youths who intentionally take drugs is alarming. Prince [8] linked drug use to delusions. According to Jones and McCanceKatz [9], opioid use disorder (OUD) is associated with mental illness. Drug usage may cause mental illness [10, 11], which needs further study. Most of the above research examines a specific opioid issue, such as age or demography [12, 13, 14]. Age, race, health, and demographic aspects in our database make it a useful source for problem information. Previous studies did not contextualize patient event notes and medical histories using NLP.

Researchers identified purposeful and unintentional opioid users using MIMIC-III structured and unstructured data. They used classic machine learning and deep learning to recognize users trying to attain goals in 455 patient instances.

Most of the earlier studies didn't look at how opioid intake classification affects the mental health of people who use them. Also, these studies don't use the newest DL -based techniques to get more useful results, such as attention and information distillation mechanisms. Machine Learning (ML) and Deep Learning (DL) techniques were used for classification.

1. **LITERATURE REVIEW**

**Co-occurring substance use and mental disorders among adults with opioid use disorder:**

Background: People with opioid use disorder (OUD) and severe illness or substance abuse are more likely to get sick and die. Addressing these problems simultaneously improves treatment and health outcomes. Some of the studies inspected about prevalence of co-occurring disorders, their demographics, and OUD patients. This hinders focused, resourced policies and therapeutic solutions. It works: About half of persons aged 18–64 with OUD had both a mental and a drug use disorder, according to the 2015–2017 National Survey on Drug Usage and Health. The participants received therapy for both their opioid addiction and mental health problems. Multivariable logistic regression was used to analyze the data. Results: 23.6% to 29.4% of persons with OUD had alcohol use disorders and 8.6% to 13.0% had methamphetamine use disorders. 64.3 percent had an AMI and 26.9% had a SMI. In the past year, around 24% of individuals with opioid use disorder (OUD) and acute myocardial infarction (AMI) received mental health and substance therapy. Thirty-nine percent of OUD/SMI individuals received both treatments. Finally, adults with OUD commonly have mental and substance issues. We must make it easier for this group to access full-service models for mental and drug use issues.

**Mental disorder and opioid overdose: A systematic review:**

The study describes the latest evidence on opioid overdose and mental illness among opioid users. Methods: The Open Science Framework published the protocol. The review question was based on the PECOS paradigm. We found and carefully read studies from North America, Europe, the UK, Australia, and New Zealand that came out between January 1, 2000, and January 4, 2021. We did this by searching computer databases, looking at citations, and calling experts. There were evaluations of the risk of bias. The lumping method was used to put the data together. The Results Out of the total of 6512 records that were looked over, 38 were chosen to be included. There is a link between at least one type of mental illness and opioid overdose in 37 of the 38 studies that made up this review.[20] The strongest evidence is for internalizing disorders in general and mood disorders in particular. Anxiety disorders are next, with only weak evidence supporting the link between thought disorders (such as schizophrenia and bipolar disorder) and opioid overdose. A link between any disorder and overdose was also found to have some proof. In conclusion Almost all of the studies that were looked at found a link between mental disorder and overdose. Having a mental disorder is linked to both deadly and non-fatal opioid overdoses, but the direction of the link is still not clear.

**Using machine learning to predict opioid misuse among U.S. adolescents**:

This study is to determine teenage misuse of opioid in the United States using three different ML methods. They analyzed data from a national survey conducted between 2015 and 2017, which included over 41,000 youths aged 12 to 17. The prediction models were created using ANN, distributed random forest, and gradient boosting. Penalized logistic regression and ML forecast models were tested. We measured prediction accuracy using AUROC and AUPRC. Our major prediction measure was the AUPRC because it is better for evaluating binary classifiers on unbalanced outcome variables than the AUROC. Overall, 3.7% of U.S. youths and young adults (n = 1521) misused opioids. All four models made nearly the same amount of accurate predictions (AUROC scores 0.809–0.815). Distributed random forest had the highest prediction accuracy at 0.172, followed by penalized logistic regression at 0.162, gradient boosting machine at 0.160, ANN at 0.157. ML can be useful in predicting uncommon occurrences such as teenage misuse of opioids, especially when the outcome being predicted is heavily imbalanced.

**Predictive modeling of susceptibility to substance abuse, mortality and drug-drug interactions in opioid patients:**

: Opioids, painkillers, are being studied . Long-term opioid use causes addiction. They may cause mental disease, muscle pain, depression, panic attacks, and other issues. This study provides prediction algorithms that use a patient's prescription history to predict opioid misuse and death. Drug combinations are especially risky with opioids. Methods and Tools: LR using L2 regularization and Extreme Gradient Boosting were trained on a public MIMIC-III dataset. These models divided patients into two opioid abuse risk categories. We investigated potential pharmacological interactions using K-Means clustering, an unsupervised approach. The F1 score for predicting opioid misuse in patients is 76.64%, which means the accuracy is approximately 77.159%.However, this model has an F1 Score of 94.4%, indicating 94.35% accuracy. These can help identify opioid abuse causes. They can also monitor prescription filling to reduce opiate abuse. Conclusions and thoughts Results demonstrate that the enhanced model may help identify opioid abusers before they start. ML algorithms can identify and link patient risk factors for opioid overdose or misuse. This can reveal gaps and dishonest prescription writing. Our unforeseen discovery discovered that insulin may react poorly with opioids, which can cause issues in diabetics. Long-term opioid users may need extra insulin to improve their insulin function. This supports previous studies on a related problem. We share our prediction models and software code under the MIT License so they can be used elsewhere.

**Predictors of transition to heroin use among initially non-opioid dependent illicit pharmaceutical opioid users: A natural history study:**

Background: A study found that an individual who uses illicit opioids are more likely to start using heroin. Young people who use illegal opioids but have never been addicted to heroin are also at risk of transitioning to heroin. The study used respondent-driven sampling to identify participants and conducted structured interviews every six months for 36 months. Cox regression is used to determine predictors of heroin usage. The study found that the estimate of heroin useage over 36 months was 7.5%, with an annual commencing use rate of 2.8% for all white participants. The average time someone used heroin before admitting to it was 6.2 years. Age was found to increase the addictive nature of opioids, and predictors of heroin usage included usage of opioids to get high and using opioids in non-oral forms. This study provides valuable insights into young adolescents using illegal opioids and their potential transition to heroin. Early on, these youths are not opioid-dependent. The results suggest what to aim for when creating the urgent protection initiatives, we need.

1. **METHODOLOGY**

In our research, we used ML and DL techniques to classify medications. We analyzed both organized and unstructured MIMIC-III data to identify individuals who intentionally or accidentally used opioids. We find the database tables that are useful for our study and choose 41 traits that are useful for our study. We can tell which patients are on purpose by looking at the keywords and their medical background. This way, we can label our information as "YES" or "NO" for opioid intake. We then generated a structured dataset and trained it with AdaBoost, LR, SVM, XGB, and RF Classifiers across 10 cross validation rounds. The model is strengthened by adding an unstructured sample. We use DL-based NLP algorithms because it is challenging to train an unstructured dataset. We train disorganized data with CNN, simple LSTM, and attention-based LSTM models. The knowledge distillation process creates a composite model. The structured dataset shows the network with the most capacity, and we transfer knowledge to the less-capable unstructured dataset. Finally, we employed distilled knowledge to assess the accuracy of organized and unstructured datasets alone and when combined by these two models. We do the different ablation studies to see how well the suggested model works when these parts are changed.

**Benefits:**

1. We looked into more types of ML and DL methods. We might be able to get better accuracy and performance with this wider range of model options in our work.

2. We do a good job of combining unstructured data, like patients' experiences and stories, which often have useful context and insights. This addition gives us a fuller picture of the opioid abuse situation and might help us make more accurate predictions.

3. To process and pull out meaningful patterns from unstructured data, we used NLP methods based on DL. Taking this method lets us get a fuller picture of the patients' situations and might help us make more accurate predictions.

4. In order to enhance the model, we utilized a technique called knowledge distillation to transfer information from a more organized dataset to a less organized one. This exchanging and merging of knowledge has the potential to result in more precise predictions and understandings.

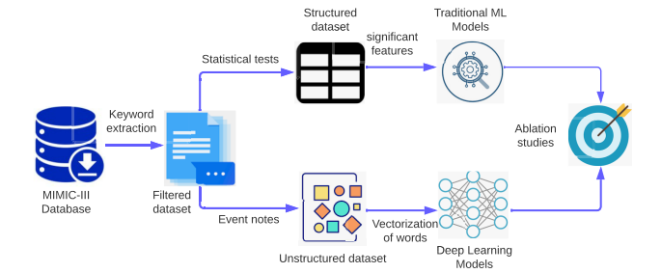


Fig 1 System Architecture

**DataSet**

MIMIC III partial data set is available in the public domains by using this data generated the data set with 41 fields in that some of the sample fields are ethnicity, diagnosis, discharge, location, insurance, gender, marital status, mental status, length of stay in ICU and length of stay in ward and death time. Some Sample fields data set is shown in the following table 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ethnicity | Diag\_  nosis | Discharge  location | insurance | Gen der |
| BLACK /AFRICAN AMERICAN | 54 | HOME HEALTH CARE | Medicare | F |
| UNKNOWN/ NOT SPECIFIED | 26 | DEAD / EXPIRED | Private | F |
| UNKNOWN /NOT SPECIFIED | 54 | DEAD / EXPIRED | Medicare | F |
| WHITE | 27 | SNF | Medicare | F |
| WHITE | 5 | DEAD /EXPIRED | Medicare | M |
| WHITE | 61 | REHAB / DISTINCT PART HOSP | Medicare | F |
| WHITE | 0 | SNF | Medicare | F |

Table. 1

Results

This paper works with CNN, LSTM and Hybrid (CNN+LSTM) model to classify the opioid patients on the provided dataset presented in table 1. The results are tabulated in the table 2, 3 and 4 and the same results are plotted in the figures 2, and 4.

CNN and LSTM models individual attains good accuracy for training but fails to attain that level in the testing scenario, this results that there may be chance of overfitting but by combining both the models, it was clear thatthe results with the hybrid model performs well when compared with other models.

CNN model

|  |  |  |
| --- | --- | --- |
| Epoch | Test Accuracy | Train Accuracy |
| 1 | 0.6167 | 0.6667 |
| 2 | 0.7 | 0.7333 |
| 3 | 0.7 | 0.73333 |
| 4 | 0.7 | 0.93333 |
| 5 | 0.8668 | 1 |
| 6 | 0.7167 | 0.9333 |
| 7 | 0.7117 | 0.8667 |
| 8 | 0.78 | 1 |
| 9 | 0.73 | 1 |
| 10 | 0.74 | 0.9333 |

LSTM

|  |  |  |
| --- | --- | --- |
| Epoch | Test Accuracy | Train Accuracy |
| 1 | 0.7097 | 0.7067 |
| 2 | 0.6615 | 0.7067 |
| 3 | 0.7507 | 0.73333 |
| 4 | 0.7242 | 0.93333 |
| 5 | 0.6903 | 0.889 |
| 6 | 0.726 | 0.9333 |
| 7 | 0.7816 | 0.8667 |
| 8 | 0.75 | 0.898 |
| 9 | 0.765 | 0.898 |
| 10 | 0.7646 | 0.9333 |

LSTM+CNN

|  |  |  |
| --- | --- | --- |
| Epoch | Test Accuracy | Train Accuracy |
| 1 | 0.7197 | 0.789 |
| 2 | 0.73 | 0.825 |
| 3 | 0.77 | 0.79 |
| 4 | 0.74 | 0.788 |
| 5 | 0.6903 | 0.769 |
| 6 | 0.73 | 0.88 |
| 7 | 0.8 | 0.8667 |
| 8 | 0.8145 | 0.898 |
| 9 | 0.798 | 0.898 |
| 10 | 0.8124 | 0.9333 |

1. **CONCLUSION**

The use of opioids by both young and old people is a problem around the world. We used the MIMIC-III dataset to build a dataset. From a total number of 26 database tables, we chose mainly 41 features for our dataset. Later, we found traits that were linked to opioid intentional YES/NO users. There is a strong link between three important traits that we found. We used this method to create a tabular dataset that did a good job of predicting how drug users would use it. We also made a DL model that can guess how opioid users will use drugs based on information about their past behavior (event notes). We got a 93% success rate with our tabular model when we used random forest classifiers. After some time, we used our DL tools (1D CNN, LSTM attention) to get 66% accurate data from patients' unstructured history data. We got an overall accuracy of 76.44% when we used the information distillation mechanism of the tabular model instead of the DL model. We found some interesting links between the mental health problems of users and the apps. There are many ways to make our studies even better. We might add more data to our set, which could mean more work by hand to find people who use opioids.

**6. REFERENCES**

[1] A. Maguire, Addressing America’s Hidden Killer: The Opioid Crisis, Penn State Univ. Site, USA, 2021.

[2] Opioid Overdose Crisis. National Institute on Drug Abuse, National Institutes of Health, Bethesda, MD, USA, May 2020.

[3] C. A. Shaheed, C. G. Maher, K. A. Williams, R. Day, and A. J. McLachlan, ‘‘Efficacy, tolerability, and dose–dependent effects of opioid analgesics for low back pain: A systematic review and meta–analysis,’’ JAMA Internal Med., vol. 176, no. 7, pp. 958–968, 2016.

[4] D. Dowell, T. M. Haegerich, and R. Chou, ‘‘CDC guideline for prescribing opioids for chronic pain—United States, 2016,’’ JAMA, vol. 315, no. 15, pp. 1624–1645, 2016.

[5] J. Barkley and J. Shin, ‘‘Analysis of intentional and unintentional drug overdose deaths occurring in Rhode Island, 2016–2019,’’ Rhode Island Med. J., vol. 105, no. 1, pp. 46–48, 2022.

[6] B. Han, W. M. Compton, E. B. Einstein, J. Cotto, J. A. Hobin, J. B. Stein, and N. D. Volkow, ‘‘Intentional drug overdose deaths in the United States,’’ Amer. J. Psychiatry, vol. 179, no. 2, pp. 163–165, Feb. 2022.

[7] K. Pringle, S. Caupp, J. Shi, K. K. Wheeler, H. A. Spiller, M. J. Casavant, and H. Xiang, ‘‘Analysis of intentional drug poisonings using Ohio poison control center data, 2002–2014,’’ Clin. Toxicology, vol. 55, no. 7, pp. 652–658, Aug. 2017.

[8] J. D. Prince, ‘‘Correlates of opioid use disorders among people with severe mental illness in the United States,’’ Substance Use Misuse, vol. 54, no. 6, pp. 1024–1034, May 2019.

[9] C. M. Jones and E. F. McCance-Katz, ‘‘Co-occurring substance use and mental disorders among adults with opioid use disorder,’’ Drug Alcohol Dependence, vol. 197, pp. 78–82, Apr. 2019.

[10] J. van Draanen, C. Tsang, S. Mitra, V. Phuong, A. Murakami, M. Karamouzian, and L. Richardson, ‘‘Mental disorder and opioid overdose: A systematic review,’’ Social Psychiatry Psychiatric Epidemiology, vol. 57, pp. 1–25, Nov. 2021.

[11] A. S. B. Bohnert and M. A. Ilgen, ‘‘Understanding links among opioid use, overdose, and suicide,’’ New England J. Med., vol. 380, no. 1, pp. 71–79, Jan. 2019.

[12] D.-H. Han, S. Lee, and D.-C. Seo, ‘‘Using machine learning to predict opioid misuse among U.S. adolescents,’’ Preventive Med., vol. 130, Jan. 2020, Art. no. 105886.

[13] M. Ramya Vunikili, B. S. Glicksberg, and L. Subramanian, ‘‘Predictive modeling of susceptibility to substance abuse, mortality and drug-drug interactions in opioid patients,’’ Frontiers Artif. Intell., vol. 4, Dec. 2021, Art. no. 742723.

[14] R. G. Carlson, R. W. Nahhas, S. S. Martins, and R. Daniulaityte, ‘‘Predictors of transition to heroin use among initially non-opioid dependent illicit pharmaceutical opioid users: A natural history study,’’ Drug Alcohol Dependence, vol. 160, pp. 127–134, Mar. 2016.

[15] K.-H. Yu, A. L. Beam, and I. S. Kohane, ‘‘Artificial intelligence in healthcare,’’ Nature Biomed. Eng., vol. 2, no. 10, pp. 719–731, Oct. 2018.

[16] Ch Raja Ramesh et al. Cluster optimization for boundary points using distributive progressive feature selection algorithm. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(2):3320–3328, 2021.

[17] Ch Raja Ramesh, NK Prasad, KP Bhuravarjula, and N Krishna. Customer relationship management system with the uscm-akd approach of d3m. International Journal on Computer Science & Engineering, 1:2036–2040, 2010.

[18] Shirin Bhanu Koduri, Loshma Gunisetti, Ch Raja Ramesh, KV Mutyalu, and D Ganesh. Prediction of crop production using adaboost regression method.In Journal of Physics: Conference Series, volume 1228, page 012005. IOP Publishing, 2019

[19] Muralidhar Pantula and KS Kuppusamy. A machine learning-based model to evaluate readability and assess grade level for the web pages. The Computer Journal, 65(4):831–842, 2022.

[20] Anusha Penugonda and Muralidhar Pantula. Assessment on multiple aspects of online book recommendation systems. In Intelligent Manufacturing and Energy Sustainability: Proceedings of ICIMES 2022, pages 625–633. Springer, 2023.