International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 010-017
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0002

Online E-Health Monitoring and Drug Overdose Prediction Using Machine Learning

Kiran G R[1], Mrs.Divya H T[2], DeviCharan Reddy R[3], Karan P Gowda[4], Mourya J[5]
[1,2,3,4,5] Information Science, Malnad College of Engineering
Email ID: *kirangr771@gmail.com[1], dht@mce.ac.in[2], charanreddyr26@gmail.com[3], gowdakaranp@gmail.com[4] mouryaj12042004@gmail.com[5]*

Abstract
The opioid crisis is growing out of proportions; this requires smart approaches to detecting patients who are likely to overdose on drugs. This system is using ML to classify heroin consumption using behavioral, psychological and demographic variables that are based on the Drug Consumption Quantified dataset. The twelve critical attributes such as dimensions of personalities and level of impulsivity are used in order to maximize prediction. A normalization of the data is done by StandardScaler followed by classification with Logistic Regression with a balanced class weight, optimized based on GridSearchCV and stratified 5-fold cross-validation. The trained model has a high classification accuracy and reliability with the visual analysis of the accuracy curves and confusion matrices that provides a thorough evaluation of the performance. The framework focuses on interpretability and the importance of mental health indicators in predicting opioid consumption, which ensures the data-grounded and explainable method of overdose risk assessment. The system offers early detection and proactive response to opioid dependency by effective preprocessing, powerful model tuning and visual representation of metrics to develop transparent and trustworthy ML-based analytics.

Keywords: *Opioid intake, mental illness, MIMIC-III database, machine learning, deep learning".*

1. Introduction

The use of opioid analgesics is usually associated with the treatment of acute and chronic pain, and the United States has very high rates of prescription. The Centers of Disease Control and Prevention (CDC) state that the costs of opioid abuse in a year amount to around 78.5 billion dollars [1], not to mention the findings of several studies that estimated the number of opioid prescriptions written to be approximately 153 million in 2019 [2]. Although these medications are used in a therapeutic manner, they include the potential of being overused because they are addictive, which creates a major health problem among the citizens [3]. A number of researches have cited that patients usually take opioids not just to relieve pain, but also because of dependency, which may result in cases of overdose [4]. Mental health disorders are known to be related to intentional overdoses, which implies that depression is a key issue that contributes to opioid misuse [5]. The statistics of the prevalence of intentional drug consumption among adolescent populations are also alarming and must be addressed in terms of early intervention [6]. Additional studies indicate that opioid use disorder (OUD) is strongly connected with other mental illnesses. It has been shown that people with psychiatric disorders are predisposed to opioid dependence, and socioeconomic variables of age and ethnicity influence the usage patterns even more [7], [8], [9], [10]. Although there is a plethora of literature that dwells upon particular demographic groups or discrete factors, there is still a lack of findings that utilize both behavioral and psychological characteristics, which prompts the necessity of predictive models that can bear in mind numerous risk determinants at the same time. In that regard, machine learning methods would offer a solution to the problem of identifying people at risk of opioids abuses. Unstructured data, such as the clinical event notes of patients, is united with structured data, including demographic, behavioral, and psychological characteristics, to gain a comprehensive picture of how they are used. Unstructured data are processed with deep learning-based natural language processing (NLP) models, such as word embeddings and attention-based LSTMs, and the knowledge distillation methods are used to migrate the knowledge obtained on structured data to improve the unstructured datasets. Incorporating mental health and socio-economic variables with opioid intake patterns allows identifying the latter early, assisting in clinical decision-making and being integrated in prevention measures that focus on dieting against the risk of overdose.

2. Literature Review

Vunikili, Glicksberg, and Subramanian [11] suggested a predictive modeling method, to determine vulnerability to substance abuse, mortality, and drug-drug interactions in patients on opioids. Their research emphasized the importance of organized electronic health record information in the interpretation of personal risk factors and forecasting negative outcomes of opioid use. The authors utilized machine learning algorithms on the tabular datasets of patients to predict the probability of substance misuse and complications and have proven that trained models can deliver action-able information to clinicians. They focused on the importance of demographic, clinical, and medication-related factors as the most important predictors, which allows making a more accurate assessment of opioid-related risks and the personalized interventions of the patients under risk. This work formed the basis of data-driven approaches to decrease opioid misuse by incorporating various health aspects in a predictive model. Yu, Beam, and Kohane [12] have given an overview of the use of artificial intelligence in healthcare, speaking about the possibility of AI to change clinical decision-making, patient monitoring, and prediction of risks. The authors described several machine learning and deep learning algorithms that can be applicable to structured and unstructured medical data. Their input described how AI models can increase predictability, especially in the mechanism of identifying vulnerable patients who can face adverse outcomes, including the abuse of medication. They underlined that AI in healthcare procedures can optimize the workflow, preventive interventions due to the possibility of monitoring adverse risk dynamics at the initial stages, and this is why they could be applicable to the prevention of opioid abuse. As indicated in this paper, knowledge of clinical domain in conjunction with machine learning can result in interpretable and reliable information to medical workers. Jiang et al. [13] reported on the history of artificial intelligence in the healthcare field, its application in the past, and the current problems and opportunities. Their argument was that structured clinical data and unstructured narrative reports were crucial in developing predictive models and that sound preprocessing and feature engineering played a major role in assisting them to achieve high performance. The authors proposed that AI systems can be useful in proactive healthcare, which comprises early indicators of chronic disease progression or negative drug effects. In the case of opioids, the chance to utilize this type of AI-based structures to determine risky or at-risk individuals implies that the demographic and behavioral data have to be analyzed. Their analysis has demonstrated the significance of the interpretable models, which will make a trade-off between predictive accuracy and clinical significance. The updated opioid risk tool was described by Cheatle et al. [14], and it is a clinically validated tool that predicts opioid use disorder among the subjects diagnosed with chronic nonmalignant pain prescribed to take opioids. The patient demographic, clinical and psychosocial data were considered in this tool in order to develop a risk score that can be used to inform prescribing habits and monitoring plans. Their research has revealed that when the array of patient data is combined with risk prediction algorithms, more vulnerable patients can be identified early and that is the key to the prevention of opioid-related complications. This work was a transition between traditional clinical scores and predictive models that are developed with the assistance of this machine learning showing that standardized scores could be enhanced with the help of machine learning and be more precise. MIMIC-extract Wang et al. [15] created an effective data extraction, preprocessing and representation pipeline on MIMIC-III database. This structure enabled researchers to effectively handle big data on critical care, which entails the incorporation of organized tables and clinical stories. The authors introduced automated preprocessing operations such as normalization, missing values, and feature representation, which minimizes interruptions in the process of using machine learning to both sophisticated datasets. Through their work, structured datasets appropriate in predictive modeling were generated and it was simpler to generate models of tasks like opioid risk prediction besides ensuring reproducibility and data integrity. The study by Calcaterra et al. [16] aimed at forecasting chronic opioid use in the future of hospitalized patients through the analysis of their past medical history. They used the structured patient data (the history of previous medications, demographics, and comorbidities) under supervised machine learning algorithms. Their results pointed to the fact that past clinical data might be used to anticipate the long-term patterns of opioid use, which would offer clinicians with a data point to engage in the process of prescribing and monitoring. The paper highlighted the significance of correct feature selection and model testing in order to obtain credible predictions, which proves that machine learning could be used to supplement clinical reasoning to avoid opioid dependence. Mensah et al. [17] suggested a proactive method of dealing with opioid crisis by making use of machine learning methods. They emphasized the role of predictive models in helping healthcare professionals to reach high-risk people prior to misuse development. Using supervised learning in structured datasets of healthcare, their study proved the ability of predictive analytics to inform prophylactic interventions, allocate resources efficiently and mitigate the effects of opioid abuse on society. Their strategy focused on applying data-based strategies to clinical practice, so that machine learning predictions would be converted into actionable decisions. Introduction of MIMIC-III database by Johnson et al. [18] is an open-source critical care database, which stores structure and unstructured patient data. It has been an indispensable source of healthcare predictive modeling that provides full demographic, clinical and laboratory information. The authors highlighted that it can be used in the development of machine learning models capable of predicting risks, stratification of patients, and patient outcome. Regarding the opioid-related research, MIMIC-III enables the combination of the histories, vital signs, and clinical notes of patients to study the trends of their usage and abuse. Nuthakki et al. [19] considered the MIMIC-III clinical notes analysis through the assistance of the natural language processing (NLP) and neural networks to reveal the diagnosis and procedure. They also revealed through their experiment that unstructured clinical narratives possess informative contents, which can be utilized to enhance predictive modeling in the case where they are combined with structured data. They employed deep learning structures to process textual data into embeddings and recall useful features that are applicable in classifying patients. Social media trends related to opioids, including Twitter, were investigated by Lossio-Ventura and Bian [20] and it was proved that substance misuse can be monitored with alternative sources of data. They suggested the application of data mining and machine learning methods to determine patterns, geolocations, and demographics regarding opioid use. According to their results, the integration of clinical data with social media analytics can give a bigger picture of the patterns of misuse, improve the early warning mechanisms and shape the methods of informing the population about the health aspects. The paper supports the importance of using a wide range of data sources as a supplement to conventional medical records to predictive model.

3. Framework and Methodology

The Preprocessing of the data is done with the help of StandardScaler that normalizes the distribution of features. The classification model uses the Logistic Regression where the class weight is set to balanced to deal with imbalance in classes. Optimization is done by applying a stratified five-fold cross-validation approach to model optimization with the help of the GridSearchCV to find the best regularization parameter (C). The whole workflow is executed as a strong scikit-learn pipeline, which allows maintaining uniform data transformation and assessment. Matplotlib can also be used to generate visual insights like accuracy curves and confusion matrices allowing one to clearly understand how the model behaves. The systematic method of opioid risk classification increases their reliability, scalability, and interpretability. The MIMIC-III database forms the start of the system architecture, and the keyword filtering is used to extract the data about the patients in the relevant format. The filtered dataset is broken down into structured data, including demographic and clinical characteristics, and unstructured data, which is represented by event notes. The structured data is subjected to statistical analysis to extract meaningful features and input them in the conventional machine learning models. Deep learning models are applied to unstructured data to convert it into a form of vectors. Lastly, the results of the both models are tested using an evaluation technique of ablation studies that determine the performance of the results and the optimization of the predictive accuracy. Figure 1 shows Proposed Architecture

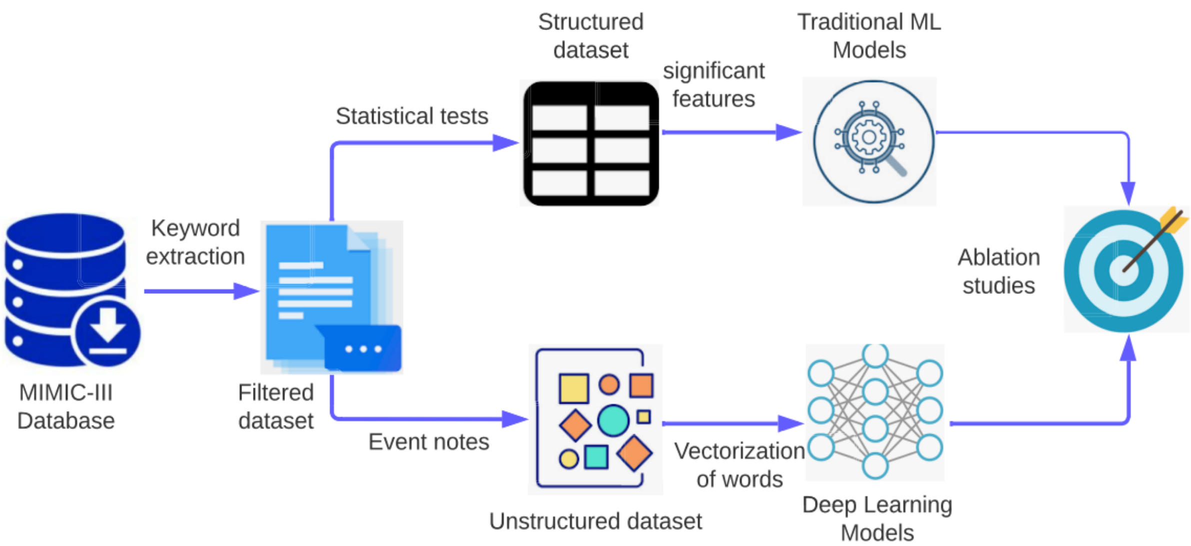

Figure 1 Proposed Architecture

4.1 Dataset Collection

The Drug_Consumption_Quantified dataset implies the behavioral, psychological, and demographic data that help to study the tendencies to the usage of drugs in people. It consists of the following attributes like Age, Gender, Education, Country, Ethnicity, and psychological characteristics like Nscore, Escore, Oscore, Ascore, Cscore, Impulsive, and SS. The data captures the rate of usage of different substances such as heroin thus binary demarcation of those who use them and those who do not can be made. Such a rich and organized dataset is useful in the extraction of features and predictive modeling in the detection of drug overdose risk.

4.2 Pre-Processing

Pre-processing phase is a stage before the machine learning process that prepares the data by cleaning, standardizing, and choosing the most relevant features, data consistency, noise reduction, and increasing the accuracy and interpretability of the model. Data Preprocessing: The data obtained is processed to be consistent and of quality. Missing or erratic values are detected and processed and numeric variables standardized using StandardScaler to

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 010-017
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0002

normalize scaling. The target variable, Heroin is coded in binary terms, that is 1 is recent users (CL4–CL6) and 0 others, which allows machine learning models to understand the result. This is done to prepare the data to be reliably trained and assessed.

Feature Selection: The dataset is then narrowed down to twelve important features in an effort to improve the model and eliminate noise. These are demographic, behavioural, and psychological variables that are Age, Gender, Education, personality scores (Nscore, Escore, Oscore, Ascore, Cscore), Impulsivity, and SS. The feature selection makes the model to concentrate on those variables that are most informative, it enhances the predictive accuracy, interpretability, and minimizes computational complexity.

4.3 Training and Testing

The data set will be divided into training and testing samples where the classification model will be developed and tested. A machine learning pipeline is created with the help of normalization using the StandardScaler and with class-weight=balanced in logistic regression to deal with the problem of class imbalance. The grid search hyperparameters are optimized by use of GridSearchCV Stratified 5-Fold Cross-Validation to identify the best performing model. This guarantees sound evaluation, strong learning as well as extrapolation to unknown data.

4.4 Algorithms/Techniques

Logistic Regression: To test whether heroin was used, the model is the Logistic Regression. It has the ability of classifying data based on a logistic function effectively separating the classes and the uneven data is weighted by the classes to ensure meaningful and accurate predictions.

$$P(y = 1 \mid X) = \frac{1}{1 + e^{-(W^T x + b)}} \qquad (1)$$

StandardScaler: StandardScaler standardizes numeric variables, such that the mean is eliminated and the variance is brought to one. This guarantees a stable range of features, does not introduce bias in training a model and enhances convergence and stability of Logistic Regression during the classification process.

GridSearchCV: The Hyperparameter tuning is automated by the gridsearchCV that cross-validates on different parameter combinations. It determines the best regularization value (C) in the Logistic Regression that will improve the model results and performance of the model on unseen data.

Stratified 5-Fold Cross-Validation: Stratified 5-Fold Cross-Validation divides the dataset into equal portions of folds to ensure the proportion of classes. It guarantees healthy model assessment, less overfitting, and has trustworthy performance measurements in different data bundles.

Fuzzy logic: The uncertainty in behavioral and psychological data is treated with the help of fuzzy logic. The fuzzy algorithm measures the imprecise characteristics such as impulsivity and neuroticism and converts them into numerical scales to improve the interpretability of the model, the accuracy of classification, and overdose risk prediction Shown in Figure 2 - 9.

4. Experimental Results

Accuracy: The accuracy of a test refers to the test being able to distinguish the patient and healthy cases. In order to determine the accuracy of a test, the calculated portion of true positive and true negative in all the cases considered should be computed. This would be mathematically expressed as:

$$\boldsymbol{Accuracy} = \frac{\mathbf{TP + TN}}{\mathbf{TP + FP + TN + FN}} \qquad (2)$$

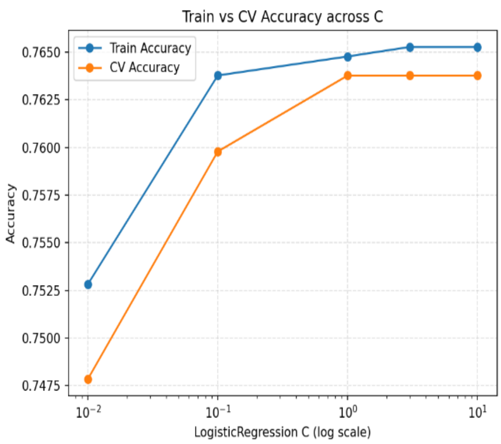

Figure 2 Accuracy Graph

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 010-017
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0002

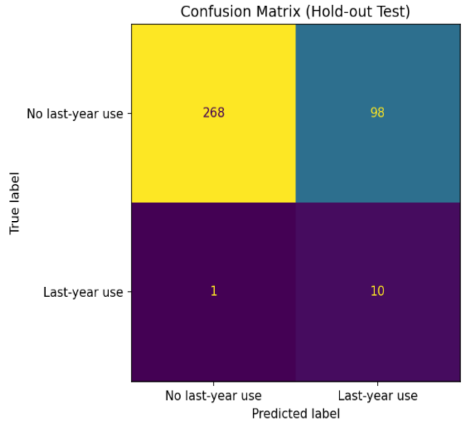

Figure 3 Confusion Matrix of Model Training

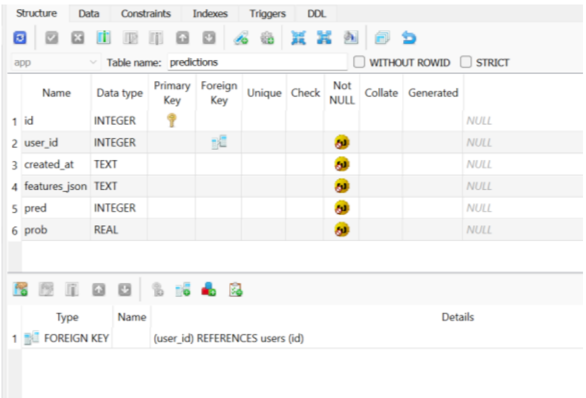

Figure 4 Database

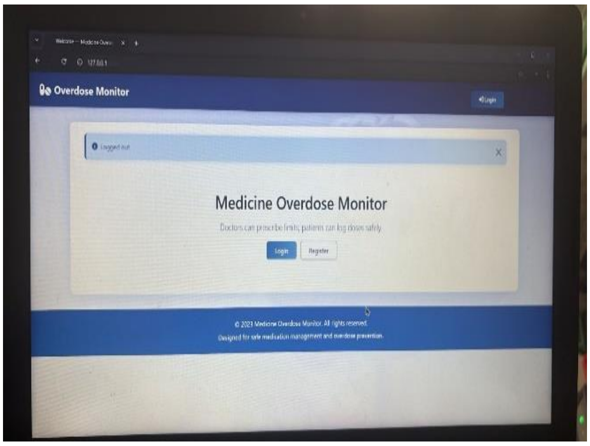

Figure 5 Home Page

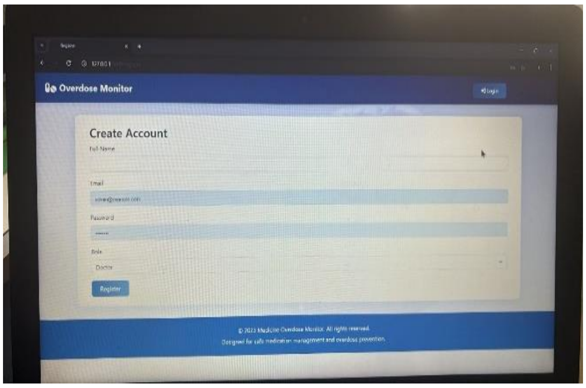

Figure 6 Registration Page

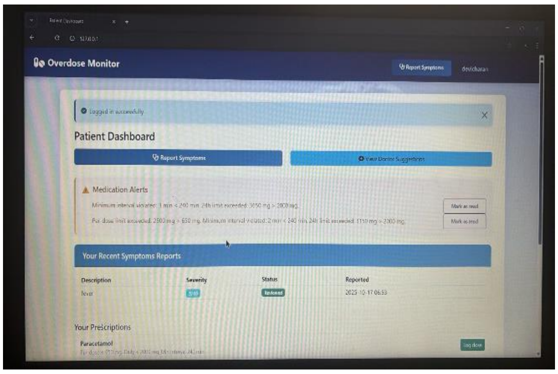

Figure 7 Patient Dashboard

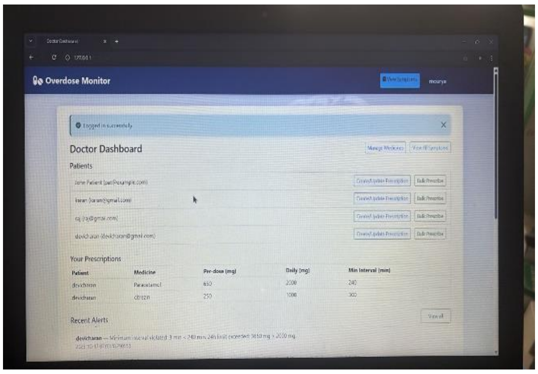

Figure 8 Doctor Dashboard

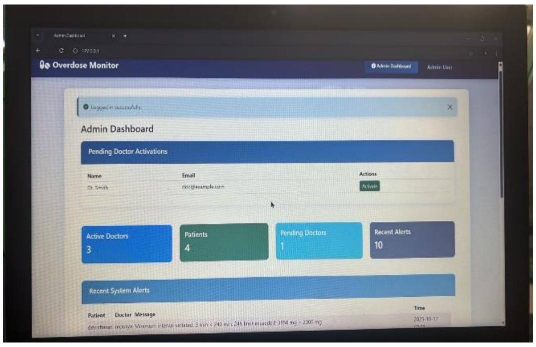

Figure 9 Admin Dashboard

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 010-017
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0002

Conclusion

The devised system is effective in screening people at risk of drug overdose with the help of a stable machine learning framework trained on the Drug_Consumption_Quantified dataset. It makes use of twelve behavioral, psychological and demographic characteristics, which guarantees a data-driven and interpretable method of predicting heroin use. The preprocessing of the data was done with the help of StandardScaler and the next steps were done with the help of Logistic Regression with the addition of class weight, class weight=balanced to overcome the imbalance of the classes. Generalization and stability of the model was maintained by hyperparameter optimization with the Stratified Five-Fold Cross-Validation used by the GridSearchCV. The accuracy curves, and confusion matrices were used to evaluate the performance of the model and they proved the strength and validity of the predictive framework in differentiating between recent heroin users and nonusers. The system focuses on the importance of mental health indicators and personality traits, which offer clear information about the factors that affect the consumption of opioids. Generally, the systematic preprocessing, feature selection, model optimization, and a comprehensive assessment develop a robust and scalable analytical framework of opioid overdose identification that contributes to the identification of risks and preventive intervention via interpretable machine learning analytics. To better the system, one could add additional datasets to represent a broad population and type of drug to have a generalized system. Better algorithms, such as ensemble or deep learning models, can be employed to refine the quality of being made predictions. The assistance of real-time monitoring and mobile application would assist receiving the risk alerts in time. In addition, social behavioral examination and mental health testing could be one worthwhile input. The system also can be scaled, interpreted and efficient to preventive intervention on the misuse of opioids via interactive exploration via visualization dashboards.

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, NIH, 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]. 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.

[8]. 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.

[9]. 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.

[10]. 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.

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

[12]. 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.

[13]. F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: Past, present and future," Stroke Vascular Neurol., vol. 2, no. 4, pp. 1–14, 2017.

[14]. M. D. Cheatle, P. A. Compton, L. Dhingra, T. E. Wasser, and C. P. O'Brien, "Development of the revised opioid risk tool to predict opioid use disorder in patients with chronic nonmalignant pain," J. Pain, vol. 20, no. 7, pp. 842–851, Jul. 2019.

[15]. S. Wang, M. B. A. McDermott, G. Chauhan, M. Ghassemi, M. C. Hughes, and T. Naumann, "MIMIC-extract: A data extraction, preprocessing, and representation pipeline for MIMIC-III," in Proc. ACM Conf. Health, Inference, Learn., Apr. 2020, pp. 222–235.

[16]. S. L. Calcaterra, S. Scarbro, M. L. Hull, A. D. Forber, I. A. Binswanger, and K. L. Colborn, "Prediction of future chronic opioid use among hospitalized patients," J. Gen. Internal Med., vol. 33, no. 6, pp. 898–905, Jun. 2018.

[17]. E. A. Mensah, M. J. Rahmathullah, P. Kumar, R. Sadeghian, and S. Aram, "A proactive approach to combating the opioid crisis using machine learning techniques," in Advances in Computer Vision and Computational Biology, Cham, Switzerland: Springer, 2021, pp. 385–398.

[18]. A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-Wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," Sci. Data, vol. 3, no. 1, pp. 1–9, 2016.

[19]. S. Nuthakki, S. Neela, J. W. Gichoya, and S. Purkayastha, "Natural language processing of MIMIC-III clinical notes for identifying diagnosis and procedures with neural networks," 2019, arXiv:1912.12397.

[20]. J. A. Lossio-Ventura and J. Bian, "An inside look at the opioid crisis over Twitter," in Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM), Dec. 2018, pp. 1496–1499.