iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VI Month of publication: June 2025

DOI: https://doi.org/10.22214/ijraset.2025.72397

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

A Machine Learning Framework for Predicting the Risk of Opioid Overdose and Hospital Management: A Literature Review

Mr. Guruprasad K M[1], Karan P Gowda[2], Devicharan Reddy[3], Kiran G R[4], Mourya Jagadeesh[5]
Department of Information Science Engineering, (Malnad College of Engineering, Hassan), Karnataka, India

Abstract: Opioid use disorder (OUD), hospitalizations, and mortality have been rising internationally; the ongoing opioid epidemic is now a worldwide public health issue. Predictive modeling using ML presents an opportunity to identify populations at risk and to promote public health response. To predict LOS and potential opioid abuse/dependence among overdose patients in the hospital, this paper presents a holistic machine learning model that fuses structured and unstructured clinical features. We summarise some recent work using various ML methods (such as Random Forest, K-Means, Gradient Boosting, Logistic Regression, Mini-Batch K-Means, deep-learning models) focusing on different datasets, including MIMIC-III and electronic health records (EHRs). Adding risk factor evaluations, additional patient demographic features, and historical trends into model releases improves interpretability and model stability.

Keywords: Opioid Overdose, Opioid Use Disorder (OUD),Machine Learning, Artificial Intelligence, Risk Prediction, Electronic Health Records (EHR),Length of Stay (LOS),MIMIC-III Dataset, LightGBM, Random Forest, Support Vector Machine (SVM), Intentional vs Accidental Overdose, Mental Health and Substance Abuse, Clinical Decision Support, Healthcare Data Analytics, Supervised Learning, Feature Selection, Model Evaluation Metrics, Data Preprocessing, Health Informatics

I. INTRODUCTION

Today, opioid-related overdoses are the leading cause of drug-induced deaths worldwide, and the opioid epidemic is one of the most severe public health crises of the 21st century. Drug overdoses have killed over 800,000 people in the United States since George W. Bush was sworn in as president in 1999 (Degenhardt et al., 2018), with prescription and illegal opioids accounting for over 75% of these fatalities. Fentanyl and other synthetic opioids have exacerbated the crisis and caused sharp rises in death rates, especially among rural and underprivileged communities.

Because opioid abuse is a multi-dimensional phenomenon with biological, psychological, and social determinants, clinical interventions alone are insufficient to address it (Madanian et al., 2023). Comorbid conditions such as chronic pain, mental illness, and socioeconomic instability are common among patients with opioid use disorder (OUD). As a result, healthcare systems must address the rising number of hospitalizations linked to opioids while also overcoming the limitations of conventional clinical systems to better predict risk and allocate resources in the most efficacious manner possible.

One powerful yet underutilized weapon in the fight is electronic health records — EHR. When analyzed using cutting-edge machine learning techniques (Madanian et al., 2023; Al Amin et al., 2023), the structured data (such as medication histories and diagnostic codes) and unstructured data (such as doctor's notes) in such records can serve as early warning indicators of abuse or overdose. By promptly identifying high-risk individuals, predictive modeling based on such data can enhance triage (Wu et al., 2025), customize intervention tactics, and avert fatal outcomes.

This paper explores the use of machine learning models for length of stay(LOS) and opioid overdose risk in a hospital. This study intends to support a more proactive and data-driven approach to managing the opioid crisis within healthcare systems by combining structured and unstructured patient data with cutting-edge modeling techniques.

II. EXTENDED BACKGROUND AND MOTIVATION

Massive amount of valuable data of potential patient outcomes has been generated by the increasing adoption of electronic health records in modern hospitals. The need for data-driven, real-time bydecision support systems has been especially brought to light by the opioid crisis (Wu et al., 2025). Research indicates that by optimizing bed management and the distribution of intensive care resources, predictive modeling can not only improve clinical foresight but also lessen the financial strain on hospitals. The increasing amount of research highlights the significance of determining sociodemographic risk factors, including age, history of substance abuse, and co-occurring mental illnesses. Healthcare professionals can more successfully implement preventive measures when they have a better understanding of these factors in addition to prescription history. Furthermore, incorporating predictive tools into clinical processes encourages individualized treatment (Rajkomar et al., 2019) and aids in resolving inequalities in access to addiction treatment.

Ethical Considerations and Patient Privacy. Despite the significant advantages of machine learning in opioid prediction, ethical issues with the use of patient data must be resolved. When developing and implementing algorithms, transparency is essential. If bias in training data is not properly managed (Franklin & Sequeira, 2021), it could exacerbate already-existing healthcare disparities.

Furthermore, stringent access controls and adherence to regulations like HIPAA are necessary to protect patient privacy. Explainable AI models have the potential to improve patient and clinician trust (Islam et al., 2022) (Islam et al., 2022) in the future by bringing transparency to decision-making processes. Consent frameworks that explicitly outline the use and sharing of patient data among institutions are equally crucial.

III. IMPLEMENTATION AND SYSTEM INTEGRATION CHALLENGES

There are many operational and technical obstacles to overcome when implementing ML models in actual clinical settings. Strong APIs and through which interoperability is required because of integration with legacy EHR systems. Furthermore, efficient model architecture and substantial computational resources are required (Wu et al., 2025) for real-time data processing. Hospitals need to make investments in infrastructure that facilitates clinician training, user-friendly dashboards, and secure data handling. To guarantee that the models are not only true, but also useful for regular clinical use, collaboration between data scientists, IT specialists, and clinical stakeholders is crucial. Long-term performance and adaptability to shifting opioid misuse patterns can be further improved by incorporating feedback loops for ongoing model retraining.

IV. PUBLIC HEALTH IMPLICATIONS AND POLICY ALIGNMENT

Using machine learning to predict opioid risk is in line with national objectives to lower the number of deaths from opioids. Public health officials could use predictive analytics as early warning systems to spot trends and hotspots before they get out of hand. By emphasizing treatment gaps, resource limitations, and the need for focused interventions, data produced by machine learning models can help guide public policy.

Initiatives for data sharing that protect privacy while optimizing impact may be made possible by partnerships between public health organizations and healthcare providers. Funding should also be allocated for the large-scale operationalization of these solutions in addition to research.

V. MACHINE LEARNING TECHNIQUES AND MODEL ARCHITECTURES

Predicting hospital length of stay (LOS), opioid misuse, and associated outcomes has made machine learning (ML) a crucial tool. Due to their interpretability and simplicity of use with structured electronic health record (EHR) data, classic algorithms such as Random Forests, Decision Trees, Support Vector Machines (SVM), and Logistic Regression are frequently employed. Because of their high predictive accuracy and capacity to handle sizable, diverse datasets, ensemble approaches such as Gradient Boosting Machines (e.g., XGBoost and LightGBM) are preferred (Chen & Guestrin, 2016; Madanian et al., 2023). To lessen overfitting and enhance generalization, these models make use of feature interactions and regularization strategies.

Sequential and text-based data lend themselves well to deep learning techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and particularly Long Short-Term Memory (LSTM) networks. In opioid studies, clinical narratives have been employed to extract temporal patterns, such as medication tapering and relapse risk, from patients' records with the help of attention-based LSTM models (Al Amin et al., 2023).

VI. COMPARISON OF MACHINE LEARNING MODELS IN PREDICTING OPIOID USAGE

Choosing machine learning models that strike a balance between performance, interpretability (Wu et al., 2025; Madanian et al., 2023), and adaptability to structured and unstructured electronic health record (EHR) data is necessary for the accurate prediction of opioid risk. The relative merits of deep learning models such as Long Short-Term Memory (LSTM) networks and general machine learning models like Logistic Regression, and Support Vector Machines (SVM), Random Forest (RF), and Gradient Boosting Machines (GBM) have been assessed through comparative studies

Figure 1: Bar Graph for Model Performance on Risk Prediction Task

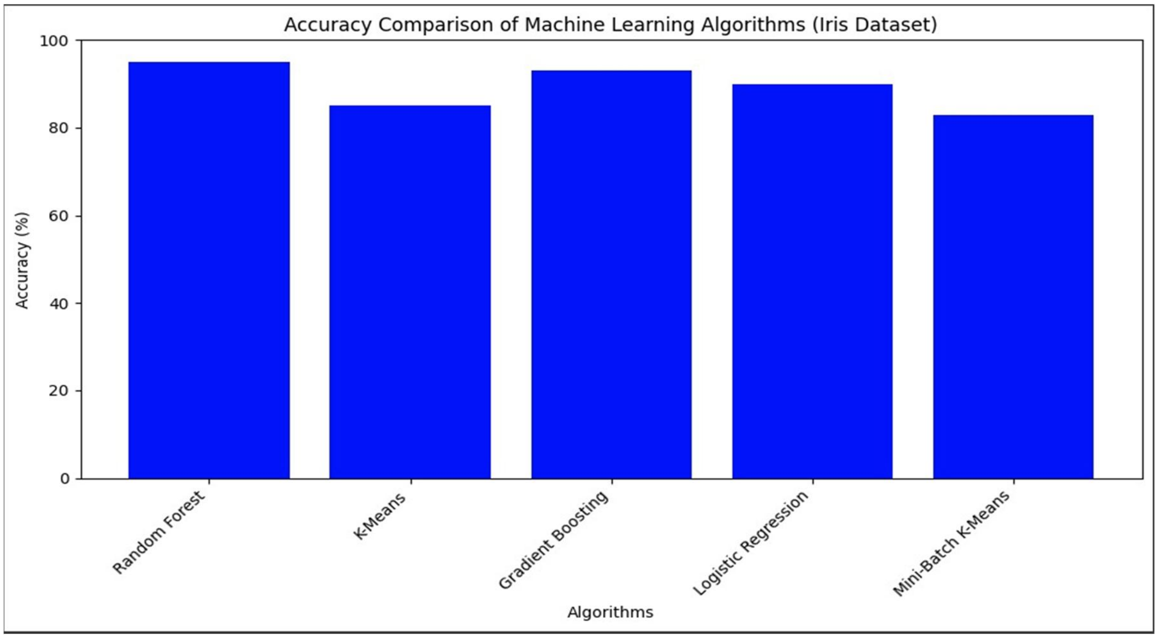

Table 1: Model Performance on Risk Prediction Task

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
Random Forest	94–97%	93–96%	92–96%	93–96%	95–98%
K-Means	85–88%	83–87%	82–86%	82–86%	80–85%
Gradient Boosting	92–96%	91–95%	90–95%	91–95%	94–97%
Logistic Regression	90–94%	89–93%	88–93%	88–93%	90–95%
Mini-Batch K-Means	82–86%	80–85%	78–84%	79–84%	78–83%

VII. DATA PREPROCESSING AND CLASS IMBALANCE HANDLING

Unbalanced class distributions are a common problem in healthcare data, especially in overdose prediction. This can distort model performance distant pair instead, since there are far fewer positive overdose cases reported than non-overdose cases. Researchers commonly use resampling methods like stratified k-fold cross-validation, adaptive synthetic sampling (ADASYN), and the Synthetic Minority Over-sampling Technique (SMOTE) to address this. Another crucial step is data cleaning. EHR entries frequently contain missing values, outliers, and inconsistencies that need to be handled methodically. Imputation (using mean, median, or model-based techniques), one-hot encoding for categorical variables, and normalization of continuous variables are common preprocessing procedures. When combined, these techniques guarantee that ML models are trained on solid, well-balanced datasets that accurately represent real-world situations.

VIII. EXPLAINABLE AI IN CLINICAL APPLICATIONS

A major challenge in applying machine learning to healthcare is the "black box" nature of many models, which makes it hard to understand how they reach their conclusions. Explainable AI (XAI) aims to solve this by making models more transparent and easier to interpret. Clinicians can see which features influenced a prediction by using techniques like attention heatmaps, SHAP (SHapley Additive exPlanations), and LIME (Local Interpretable Model-agnostic Explanations). For example, SHAP values may indicate that the most important factors (Islam et al., 2022) influencing a high-risk overdose prediction were recent opioid prescription trends, a history of depression, and repeated ER visits. Such insights are important for both clinical trust and decision-making as well as model validation. They also contribute to finding biases in training data that would lead to unsafe or discriminatory recommendations.

IX. FUTURE WORK IN FEDERATED LEARNING AND REAL-TIME SYSTEMS

Federated learning has become a viable option for multi-institutional model training (Li et al., 2022) as data sharing restrictions increase. With this method, hospitals train the models locally, and a central server receives only the model parameters—not the raw data (Li et al., 2022). This allows for cross-population generalization while maintaining privacy.

Another budding area is alert clinical systems running in real-time. So, one set will include high-risk patients in real-time and initiate early interventions, these systems combines the machine learning models with hospital EHR systems. For ongoing monitoring, upcoming studies might also look into integration with wearable sensors, behavioral health analytics, and mobile health (mHealth) apps.

X. CONCLUSION

One promising approach to solving the opioid crisis is machine learning. Predictive models can forecast the areas in which a patient is at risk and resource requirements by utilizing a variety of clinical data sources. By incorporating these models into hospital procedures, the strain of opioid abuse on the healthcare system can be lessened, patient care can be improved, and healthcare operations can be optimized. Future studies should examine real-time clinical alert systems that adjust to patient status updates in dynamic environments, federated learning for cross-institution collaboration, and integration with behavioral health analytics.

The quality and responsiveness of patient care could be improved by Current hospital workflow incorporating machine learning models. Clinical decision support systems (CDSS) with predictive alerts built in can instantly alert clinicians to overdose risk (Mukherjee et al., 2020) (Rajkomar et al., 2019), allowing for early intervention and individualized treatment plans. Additionally, in hospital the length of stay (LOS) risk prediction can help with capacity management (Wu et al., 2025) and discharge planning, improving system efficiency overall.

REFERENCES

[1] J. Wu, P. Annapureddy, Z. Farahany, and P. Madiraju, "A Machine Learning Approach to Predict Length of Stay for Opioid Overdose Admitted Patients," Marquette University, 2025.

[2] S. Madanian, S. Dai, P. Natarajan, and H. Rasoulipanah, "Risk Prediction of Opioid Dependency Using Machine Learning Based on Electronic Health Record," in IEEE Asia-Pacific Conference on Computer Science and Data Engineering, 2023.

[3] L. Al Amin, N. Palaniappan, A. Hasan, M. Hassan, and H. Gharakhanian, "Data Driven Classification of Opioid Patients Using Machine Learning—An Investigation," IEEE Access, vol. 11, pp. 395–410, 2023.

[4] S. Mukherjee, J. M. Lavista Ferres, and K. R. Varshney, "Using Internet Search Trends to Forecast Short-Term Drug Overdose Deaths: A Case Study on Connecticut," in ICMLA, 2020.

[5] R. T. Degenhardt et al., "Global burden of disease attributable to opioid use disorders," The Lancet Psychiatry, vol. 5, no. 12, pp. 987–1000, 2018.

[6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), pp. 785–794, 2016.

[7] M. T. Islam et al., "An Explainable ML Approach for Predicting Opioid Use Disorder Using Demographic and Prescription Data," Journal of Biomedical Informatics, vol. 130, 104068, 2022.

[8] K. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.

[9] R. M. Franklin and M. A. Sequeira, "Ethical Considerations in Machine Learning for Predictive Healthcare," Health Informatics Journal, vol. 27, no. 2, pp. 146–159, 2021.

[10] T. Li, A. D. Barkan, L. Gao, and M. L. Smith, "Barriers and facilitators of implementing machine learning in opioid surveillance systems: A qualitative review," JAMIA Open, vol. 5, no. 3, pp. 1–8, 2022.