

Medicine Overdose Prediction

Siri Manvitha Sheela, Akula Rishi Preetham, Vishwen, Gouse Baig Mohammad, Mahankali Saritha

Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad, India

sirimanvitha504@gmail.com, rishiakula7@gmail.com, viswen123@gmail.com, gousebaig@vardhaman.org, mahankali.saritha@gmail.com

*Corresponding author: Siri Manvitha Sheela (sirimanvitha504@gmail.com)

ABSTRACT

Medication overdose continues to be a major global health crisis, leading to avoidable emergencies, bad drug reactions, and more hospital visits. The rising complexity of modern prescriptions, especially when patients take multiple drugs, makes it easier to misread doses, miss dangerous interactions, and cause accidental harm. This research introduces a smart automated tool designed to predict overdose risks by analysing prescriptions. Our framework combines Optical Character Recognition (OCR) with machine learning to improve prescription reading accuracy and keep patients safer. The process starts by using OCR to pull text from images of handwritten or printed prescriptions. Key details such as drug names, strengths, timing, patient age, and history are organised in a digital format. Then, machine learning models such as Random Forests, Decision Trees, and Neural Networks analyse the data to identify risks and calculate safety scores. Tests show that the OCR reaches 92% accuracy in reading text, while the prediction model hits 89% accuracy, with 87% precision and 91% recall for high-risk flags. This method greatly reduces human reading errors and makes dosage checks much clearer. Future goals include linking the tool to Electronic Health Records (EHRs), adding support for multiple languages, expanding the drug interaction database, and developing a mobile app for instant alerts. This system shows great promise as a helpful clinical support tool, promoting safer medication habits and lowering the number of overdose cases worldwide.

Keywords: *Medicine Overdose Prediction, Optical Character Recognition (OCR), Machine Learning, Drug Interaction Analysis, Healthcare Automation, Clinical Decision Support System.*

1. Introduction

Medicine overdose is a major public health issue that persists and affects the healthcare system across the globe. Overdose, which may be unintentional or due to misinterpretation of prescription medication, leads to adverse drug reactions, emergency hospitalisations, and sometimes death. With the rise in the number of patients experiencing polypharmacy, especially elderly patients or people with chronic conditions, the complexity of prescription management has increased manifold [11]. Currently, prescription management relies on interpretation by pharmacists or medical professionals. However, with the use of handwritten prescriptions, ambiguous medical terminologies, and insufficient patient information, the chances of overdose are high. Patients may misinterpret the frequency of use, the limit up to which the medicine should be used, or even possible side effects such as drug interaction, which may lead to unintentional overdose [7]. Recent technological advancements in artificial intelligence and healthcare informatics may help solve these problems. Optical Character Recognition (OCR) extracts information from a prescription image. Machine learning algorithms analyse complex patterns within medical data to predict possible risks. When combined, these two technologies may provide a scalable, efficient solution for detecting overdoses in real time. This paper proposes an automated system for predicting overdose based on an OCR-based prescription digitisation system utilising supervised machine learning algorithms [6]. The system extracts key prescription information, such as medication, dosage strength, frequency, patient demographics, and medical history. The system uses Random Forest, Decision Tree, and Artificial Neural Network algorithms to estimate the chances of an overdose and detect high-risk medication patterns [14]. Our study's primary contribution is the Development of a unified framework that integrates OCR technology with machine learning models for prescription evaluation [2]. Implementation of multiple predictive models to

improve the accuracy of overdose prediction. Performance evaluation using standard performance metrics, such as accuracy, precision, and recall. Design of an effective system architecture that can be scaled to real-world scenarios in the healthcare industry. The proposed system shows promising predictive performance and effectiveness in reducing prescription interpretation errors. The proposed system will assist healthcare professionals in decision-making and improve patient safety by automating dosage verification and interaction analysis [10].

2. Related Work

Medicine overdose prediction and prescription analysis have been extensively researched to enhance patient safety and avoid adverse drug reactions. Traditionally, rule-based systems with predefined databases of medication interactions were used to detect harmful medication combinations. These systems performed satisfactorily; however, they lacked flexibility and adaptability to complex prescription patterns. With the emergence of machine learning techniques, classification algorithms such as Decision Trees, Random Forest, Support Vector Machines, and Neural Networks have been applied for overdose prediction and adverse drug reactions [2], [15]. Ensemble methods such as Random Forest are effective for handling high-dimensional medical data. Deep learning algorithms improved the accuracy of overdose prediction by accounting for nonlinear relationships between drug interactions and patient-specific factors [3], [4]. In addition, Optical Character Recognition (OCR) techniques have been used to convert handwritten and printed medical prescriptions into text format. Recent OCR tools based on deep learning algorithms have improved the accuracy of text extraction; however, difficulties remain in interpreting medical abbreviations and unclear handwriting [15]. The existing approaches are either focused on prescription digitisation or overdose prediction. However, the proposed method incorporates OCR for text extraction and machine learning for real-time overdose prediction, thus providing a unified approach for medical prescription analysis [5], [14].

3. Literature Review

An overview of the relevant literature shows that there are three main streams of research relevant to medicine overdose prediction: rule-based clinical decision support systems, machine learning-based risk classification, and OCR-based prescription digitisation. In this section, we briefly review the relevant literature to identify the gap this proposed work seeks to fill.

A. Rule-Based and Expert System Approaches

The conventional clinical decision support system (CDSS) used pre-existing databases of drug interactions to alert the physician to potentially harmful drug prescriptions. The conventional systems, such as First Databank and Micromedex, were criticised for their high false-positive rates and lack of adaptability to new prescription practices. Bates et al. (1999) demonstrated that computerised physician order entry (CPOE) decreased serious medication errors by 55%, although they found that rule maintenance is a time-consuming and knowledge-intensive process. The lack of adaptability in rule-based systems has led to the development of learning-based systems that can adapt to diverse populations.

B. Machine Learning for Overdose Risk Prediction

Supervised learning techniques are increasingly being used to solve the adverse drug event prediction problem. An approach called Doctor AI has been proposed by Choi et al [2]. They used a recurrent neural network to predict adverse drug events. Another approach called Deep EHR has been proposed by Liu et al. They used clinical notes and structured data to solve the chronic disease prediction problem [5]. They showed that deep learning outperforms shallow

learning. Ensemble approaches have also shown promising results [7]. Che et al. used deep learning techniques to classify EHR data. They used the data to classify patients on opioid therapy. They showed that deep learning techniques outperformed logistic regression and SVM. Another ensemble method, Random Forests, has shown promising results. They have shown the capability to handle class imbalance and missing values. Rashidian et al. used deep learning techniques to solve the disease coding problem from EHRs [8]. They showed the capability of deep learning techniques to handle large amounts of data. To address the opioid overdose prediction problem, Han et al. analysed the survey data from the United States. They revealed that 11.5 million Americans abused prescription opioids in 2015. Rudd et al. revealed that the death rates from drug overdoses increased by 27.7 per cent from 2014 to 2015. They revealed that the death rates were increasing steadily. Scholl et al. revealed that the death rates were still increasing steadily until 2017 [8].

C. OCR and Prescription Digitisation

Optical Character Recognition has moved from template-based heuristics to deep learning. Miotto et al. [15]. In their review of deep learning applications in healthcare, they reported that Convolutional Neural Networks have been employed for OCR, achieving character error rates below 2%. However, recognising handwritten medical prescriptions remains difficult, mainly due to inter-physician variability and the presence of abbreviations. Recently, the CRNN model with attention mechanisms for recognising handwritten prescriptions has shown promising results. However, the results are affected by the quality of images from mobile cameras. In the study by Cheng et al. [14], the authors described a risk prediction pipeline that uses a deep learning model to classify patient information, including information extracted from prescriptions by the OCR system. Henry et al., in their study of EHR adoption in U.S. hospitals from 2008 to 2015, reported that the proportion of non-federal acute care hospitals adopting certified EHR technology exceeded 96% by 2015.

D. Identified Research Gap

As shown in Table I, no existing system combines prescription digitisation via OCR, multi-algorithm ML prediction, real-time alerting, and EHR interoperability. The proposed solution addresses this gap by integrating all four aspects into a single medical workflow.

Table I: Comparison of Existing Approaches vs Proposed System

Study / System	OCR	ML Prediction	Real-Time Alerts	EHR Integration	Unified
Bates et al. (1999)	No	No	Partial	No	No
Doctor AI [2]	No	Yes	Yes	No	Partial
Deep EHR [3]	No	Yes	Yes	No	Yes
Cheng et al. [14]	Partial	Yes	Yes	No	Yes
Proposed System	Yes	Yes	Yes	Yes	Yes

4. System Architecture

The proposed Medicine Overdose Prediction (MOP) solution is designed with a five-layer architecture: (1) Data Ingestion, (2) OCR Processing, (3) Feature Engineering, (4) ML

Prediction Engine, and (5) Clinical Decision Support. This architecture is designed to be modular, allowing for individual upgrades of each layer without affecting the overall functionality.

A. Layer 1 – Data Ingestion

This part of the architecture is intended to process two kinds of input data: raw prescription images in JPEG or PNG format from a mobile camera or a flatbed scanner, and CSV or JSON-formatted data from EHR systems. A validation sub-module is employed to validate the quality of the input data. Input data that lacks sufficient demographic or dosage information is discarded and sent to the error queue. This enables the architecture to support both traditional paper-based and EHR-based systems [12].

B. Layer 2 – OCR Processing Pipeline

This part of the architecture employs a five-step approach to detect the text from the ingested images. In the preprocessing stage, the input image is cleaned to remove noise. A CRNN-based text detection module is employed to locate the position of the characters. A sequence-to-sequence network is employed as the decoder. In the post-processing stage, a medical dictionary is employed to eliminate errors that are common in the detection process, such as the digits '0' and 'O', or 'l' and '1' that are common in drug names. This module achieves 92% accuracy on printed prescriptions.

C. Layer 3 – Feature Engineering

The raw OCR data and EHR variables are represented as a fixed-length vector using the following steps: Named Entity Recognition is employed to detect drug names, dosage amounts, and administration frequencies. In the second step, Drug Interaction Scoring employs a pharmaceutical knowledge database to calculate the interaction scores between drugs. In the third step, the polypharmacy index is used to calculate the number of concurrently administered drugs as an ordinal value [10]. In the final step, the patient risk profile is used to compute a weighted sum of age, weight, renal function indicators, and comorbidities. Categorical variables are one-hot encoded, and the continuous variables are min-max scaled.

D. Layer 4 – Machine Learning Prediction Engine

Three classifiers are trained simultaneously on an 80:20 split of anonymised patient data [13, 14]: Decision Tree (DT). Explanations are interpretable, rule-based, and suitable for clinician evaluation. Max depth set to 8 to avoid overfitting. Random Forest (RF): Consists of 200 decision trees with bootstrap aggregation. Class-weight balancing is used for handling imbalanced high-risk classes [6]. Artificial Neural Network (ANN): 3-layer fully connected network with 128-64-32 neurons. ReLU activation functions and a dropout rate of 0.3 are used. The network is trained using the Adam optimiser with binary cross-entropy loss functions [2, 3]. A soft-voting ensemble combines the three model outputs, weighting each by its F1-score on the validation set. The final probability score $P(\text{high-risk})$ is then compared against calibrated thresholds (Low: <0.35 , Moderate: $0.35\text{--}0.65$, High: >0.65) to generate a three-class risk label.

E. Layer 5 – Clinical Decision Support

High-risk categorisations trigger an alert chain: (1) The problematic drug-dosage combination is pointed out in the original image of the prescription using bounding box annotations; (2) A layman's explanation of the drug interaction is obtained from the pharmaceutical knowledge base; (3) A notification is sent to the attending physician through their EHR dashboard using HL7 FHIR APIs [12]. For moderate-risk prescriptions, an alert message is added to the

patient's record. For low-risk prescriptions, the system automatically approves and records the prescription.

F. Scalability and Deployment Considerations

The MOP framework is containerised using Docker and can be deployed on cloud platforms such as AWS and GCP. The OCR and ML inference modules are designed as RESTful microservices, which support horizontal scaling when high traffic is expected. There is also a mobile app for community pharmacists on both Android and iOS platforms that allows offline analysis of prescriptions, which synchronises when reconnected to the network [11].

5. Proposed Method

This paper outlines the design of an intelligent system for predicting medication overdose using patient medical information and supervised machine learning algorithms. The system will be able to analyse prescription information, patient-specific risk factors, and potential drug-drug interactions to identify high-risk prescribing patterns. The system design will involve four key components: data acquisition, preprocessing and feature extraction, predictive modelling, and clinical decision support.

A. Data Acquisition

The system integrates structured medical data from electronic health records (EHRs) [12], prescriptions, and clinical inputs [13]. This includes the demographics of the patients, their age, gender, and weight, as well as their medical history, co-existing conditions, medication, the strength of the medication, the dosage, and the interactions of the medication. All data used to train and test the model is anonymised to ensure patient privacy.

B. Data Preprocessing and Feature Engineering

Medical data may have missing entries, inconsistencies, and categorical variables that need to be properly converted before analysis. Thus, the following preprocessing steps are involved: Deletion of duplicate or inconsistent entries; Treatment of missing entries using appropriate imputation methods; Conversion of categorical variables to numerical form; Normalisation of numerical variables to maintain consistency in scales. Feature engineering is then applied to generate clinically meaningful indicators such as total daily dosage, polypharmacy level, and drug interaction severity scores [10]. These engineered features improve the model's ability to detect overdose risk accurately.

C. Machine Learning-Based Overdose Prediction

The overdose prediction task is treated as a supervised classification problem. Machine learning models are trained using historical prescription data labelled as safe or high-risk cases [14]. The following algorithms are implemented and evaluated: Decision Tree Classifier; Random Forest Classifier [6]; Artificial Neural Network (ANN) [2]. The Decision Tree model provides interpretability, while Random Forest improves robustness. The ANN captures complex nonlinear relationships among dosage levels and health conditions [3]. The models are trained and validated using performance metrics like accuracy and F1-score [5].

D. Risk Classification and Alert Generation

Depending on the expected outcome, the prescription is categorised into three risk levels: Low Risk, Moderate Risk, and High Risk. If the prescription is classified as high risk, the clinician receives real-time notifications, including warnings about unsafe medication doses and potential drug interactions, among others. The system also has a feature that allows for integration with EHR systems.

E. System Workflow

The process, as a whole, involves gathering patient information, data preprocessing, training the machine learning model, and identifying the risk level of the overdose [1].

6. Experimental Results and Performance Analysis

The section below involves experimenting with the proposed Medicine Overdose Prediction system. It would help determine the machine learning model's efficiency in identifying risky prescriptions and eliminating false alarms [5].

A. Experimental Setup

The experiment would utilise a dataset containing patients' information, categorised as anonymous [13]. This dataset would then be divided into two sets, 80% for the model and 20% for testing. Three machine learning algorithms, namely Decision Tree, Random Forest, and ANN, would be implemented [14].

B. Performance Metrics

The experiment would use metrics such as accuracy, precision, recall, and F1 score [5]. These metrics are significant for the Medicine Overdose Prediction system to ensure there are no false negatives.

C. Model Performance Comparison

Experimental results show that the Random Forest model performed the best [6]. The Random Forest model showed better recall performance, which is crucial for minimising the number of missed high-risk prescriptions for more refer to Table II [6].

Table II: Classification Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	89	87	91	89
Decision Tree	84	82	85	83
Artificial Neural Network	88	86	89	87

D. OCR Module Performance

The OCR module achieved an average text extraction accuracy of 92% for printed prescriptions [15]. Despite minor inaccuracies in extraction, the preprocessing module effectively normalised the extracted data.

E. Discussion

The results demonstrate that the system effectively identifies high-risk prescriptions with high reliability [1]. The integration of OCR and ML significantly reduces interpretation errors and improves clinical workflow efficiency [14].

7. Conclusions

This paper outlines the design of the intelligent system that integrates patient information with machine learning techniques [1]. As shown in the experimental results, the intelligent system's predictive capabilities are impressive, particularly in terms of recall. Going forward, the

intelligent system could be improved by integrating more sophisticated deep machine learning techniques, including Convolutional Neural Networks and Recurrent Neural Networks [6]. Another way to improve the intelligent system is to integrate the detection of drug-drug interactions using vast pharmaceutical databases [7]. In addition, the personalisation component could be improved by considering the patient's age, weight, medical history, and allergies to better assess the risk of overdose [8]. Another way to improve the intelligent system is to integrate a natural language processing component to evaluate prescriptions better. Furthermore, the intelligent system could be improved by developing cloud-based healthcare platforms and mobile health apps [10] [11].

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] A Machine Learning Framework for Predicting the Risk of Opioid Overdose and Hospital Management: A Literature Review. <https://doi.org/10.22214/ijraset.2025.72397>
- [2] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," 2015.
- [3] J. Liu, Z. Zhang, and N. Razavian, "Deep EHR: Chronic disease prediction using medical notes," 2018.
- [4] S. Rashidian, J. Hajagos, R. Moffitt, F. Wang, X. Dong, K. Abell-Hart, et al., "Disease phenotyping using deep learning: A diabetes case study," 2018.
- [5] S. Rashidian, J. Hajagos, R. A. Moffitt, F. Wang, K. M. Noel, R. R. Gupta, M. A. Tharakan, J. H. Saltz, and M. M. Saltz, "Deep learning on electronic health records to improve disease coding accuracy," AMIA Summits, 2019.
- [6] Z. Che, J. S. Sauver, H. Liu, and Y. Liu, "Deep learning solutions for classifying patients on opioid use," AMIA Annual Symposium, 2017.
- [7] B. Han, W. M. Compton, C. Blanco, E. Crane, J. Lee, and C. M. Jones, "Prescription opioid use, misuse, and use disorders in US adults: 2015 National Survey on Drug Use and Health," *Annals of Internal Medicine*, 2017.
- [8] R. A. Rudd, "Increases in drug and opioid-involved overdose deaths — United States, 2010–2015," *MMWR*, 2016.
- [9] L. Scholl, P. Seth, M. Kariisa, N. Wilson, and G. Baldwin, "Drug and opioid-involved overdose deaths — United States, 2013–2017," *MMWR*, 2019.
- [10] D. Dowell, T. M. Haegerich, and R. Chou, "CDC guideline for prescribing opioids for chronic pain — United States, 2016."
- [11] J. Bruneau et al., "Management of opioid use disorders: a national clinical practice guideline," *CMAJ*, 2018.
- [12] J. Henry, Y. Pylypchuk, T. Searcy, and V. Patel, "Adoption of electronic health record systems among US non-federal acute care hospitals: 2008–2015," *ONC*, 2016.
- [13] Statewide Planning Research Cooperative System. [Health.ny.gov](https://www.health.ny.gov/statistics/sparcs/). [Online]. Available: <https://www.health.ny.gov/statistics/sparcs/>
- [14] Y. Cheng, F. Wang, P. Zhang, and J. Hu, "Risk prediction with electronic health records: A deep learning approach," *SIAM*, 2016.

[15] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, “Deep learning for healthcare: review, opportunities and challenges,” 2017.