International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Data Science-Driven Toxicology Prediction: Severity Classification and Treatment Recommendation Using Machine Learning

Mrs. Nandhini A, Venkatraman M

[1]Assistant Professor (SG), Department of Computer Applications, Nehru college of management,

Coimbatore, Tamilnadu, India.

[2] Student of II MCA, Department of Computer Applications, Nehru college of management,

Coimbatore, Tamilnadu, India.

Abstract

Toxicology-related cases are a growing concern in India, with frequent incidents of pesticide poisoning, drug overdoses, and envenomations. The major challenges include delayed diagnosis, lack of structured toxicology databases, limited accessibility to poison control centers, and the absence of real-time decision-making tools. This leads to high fatality rates, especially in rural areas where specialized toxicology units are scarce.

To address these challenges, this project applies data science and machine learning to automate poisoning case identification, symptom-based severity classification, and treatment recommendations. The system uses Natural Language Processing (NLP) for fuzzy search, TF-IDF vectorization for feature extraction, and a Random Forest Classifier for severity prediction. By leveraging structured data from open-source medical repositories, hospital records, and poison information centers, this model enhances toxicology decision-making and helps medical professionals provide faster, data-driven responses to poisoning emergencies.

The system provides an interactive dashboard with real-time analysis of toxicology cases. The results include:

- A pie chart representation of poisoning categories, showing the distribution of drug overdose, substance abuse, snakebite, and insect envenomation cases.
- A heatmap for symptom correlation, identifying patterns between different toxicology categories and symptoms.
- A search feature for toxic substances, allowing users to input a substance name and retrieve related cases, symptoms, first-aid measures, and treatments.
- Severity prediction and classification metrics, including accuracy, precision, recall, and F1-score.

1. Introduction

Toxicology plays a crucial role in emergency medicine, focusing on the identification, effects, and management of toxic substances. In India, poisoning cases due to agricultural pesticides, industrial chemicals, snakebites, and substance abuse are increasing. The lack of centralized toxicology data and manual diagnosis dependency often result in delayed or inaccurate treatments, leading to preventable deaths.

Challenges in Toxicology Cases in India

Medical professionals often face difficulty in quickly identifying toxic substances and determining the appropriate treatment. Delays in poisoning diagnosis occur due to the absence of a structured toxicology database, making toxicology assessments dependent on experience and available resources. Poison control centers are not widely available, especially in rural areas, limiting access to expert guidance. There is also a lack of structured data that can help in decision-making, forcing hospitals to rely on fragmented information.

This research develops a machine learning-based toxicology system that automatically classifies poisoning severity based on symptoms, uses NLP-powered search for substance identification, and provides data-driven treatment recommendations based on known poisoning cases.

2. Open Data Sources for Toxicology

To build a structured and reliable toxicology database, poisoning case data was collected from open-source repositories and government health databases, including:

- National Poison Information Centre (NPIC), India – Provides poisoning trends, case reports, and first-aid recommendations.
- World Health Organization (WHO) Poisoning Database – Contains international case studies on toxic substances.
- PubChem and ToxNet – Open chemical databases with toxicity data.
- Indian Council of Medical Research (ICMR) Reports – Studies on pesticide and drug-related poisoning cases in India.
- Open Government Data (OGD) Platform, India – Provides hospital records on poisoning-related emergency cases.

These data sources ensure authenticity and accuracy in machine learning model training and toxic substance classification.

3. Methodology

The toxicology prediction system follows a structured pipeline:

Step 1: Data Collection & Preprocessing

Poisoning case records were extracted from NPIC, WHO, ICMR, and hospital records. Text preprocessing techniques such as tokenization, stemming, stop-word removal, and synonym handling were applied to standardize symptom descriptions.

Step 2: Feature Extraction using TF-IDF

TF-IDF vectorization was used to convert text-based symptoms into numerical features. This helps the model distinguish poisoning severity levels more effectively.

Step 3: Machine Learning Model Training

A Random Forest Classifier was trained on symptom data to predict poisoning severity, classifying cases as mild, moderate, or severe.

Step 4: NLP-powered Fuzzy Search for Substance Identification

The system employs fuzzy search to allow users to retrieve poisoning cases based on symptom descriptions, even if entered inaccurately.

Step 5: Model Evaluation

Performance metrics such as accuracy, precision, recall, and F1-score were used to assess classification quality.

4. Machine Learning Models

A. Feature Extraction: TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF assigns importance scores to words in a document relative to their rarity and significance in poisoning cases. This transformation converts textual data into numerical vectors that machine learning models can process.

Example:

- Aspirin Overdose: "Hyperventilation, tinnitus, nausea, metabolic acidosis."
- Opioid Overdose: "Respiratory depression, pinpoint pupils, drowsiness, coma."

TF-IDF assigns higher scores to rare but crucial terms like "metabolic acidosis" and "pinpoint pupils," ensuring the model classifies cases more accurately.

B. Classification Model: Random Forest Classifier

A Random Forest Classifier constructs multiple decision trees and aggregates their predictions to improve accuracy and reduce overfitting. This ensures reliable poisoning severity classification, helping medical professionals make faster treatment decisions.

Example: Snake Venom Poisoning Case

A patient presents with symptoms:
"Severe pain, swelling, blurred vision, respiratory distress."

The system processes the symptoms and applies TF-IDF vectorization before passing them through the Random Forest Classifier.

Decision Tree Predictions:

- Tree 1 predicts: "Moderate Poisoning" (possible Russell's Viper bite)
- Tree 2 predicts: "Severe Poisoning" (possible King Cobra bite)
- Tree 3 predicts: "Severe Poisoning" (possible Indian Cobra bite)
- Final Decision (Majority Vote): "Severe Poisoning"

Treatment Recommendation:

- Antivenom: Polyvalent Anti-Snake Venom (ASV)
- First Aid: Immobilize the limb, clean the wound, avoid tourniquets
- Supportive Care: Monitor for respiratory failure, provide ventilatory support if needed
- What NOT to Do: Do not cut the bite site or apply ice

This decision-making process helps classify venom severity and recommend the appropriate treatment, reducing response time in snakebite emergencies.

C. NLP & Fuzzy Search for Toxic Substance Identification

NLP-based fuzzy search allows users to retrieve toxic substances based on symptoms, even if entered incorrectly.

Example:
A user enters: "breathing problem and fainting"

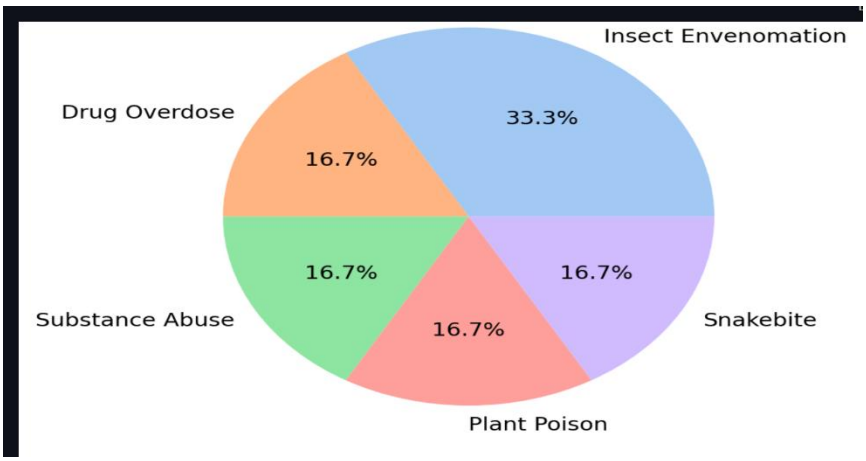
- The system matches it with "Respiratory depression, unconsciousness."
- Identified Substance: Opioid Overdose (Heroin, Morphine, Fentanyl)
- Recommended Treatment: Naloxone (Narcan), 0.4-2 mg IV every 2-3 minutes (max 10 mg total).

5. Results

The toxicology prediction system was tested on various poisoning cases, and the following results were obtained:

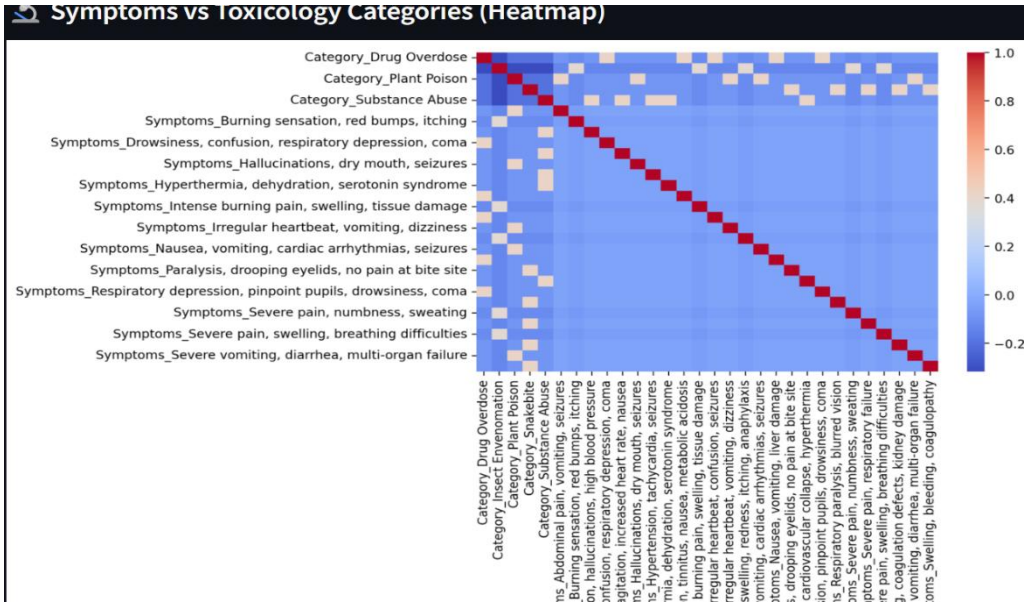
Pie Chart Representation: Displays the distribution of different poisoning categories, including drug overdose, substance abuse, snakebites, and insect envenomation.

Output:



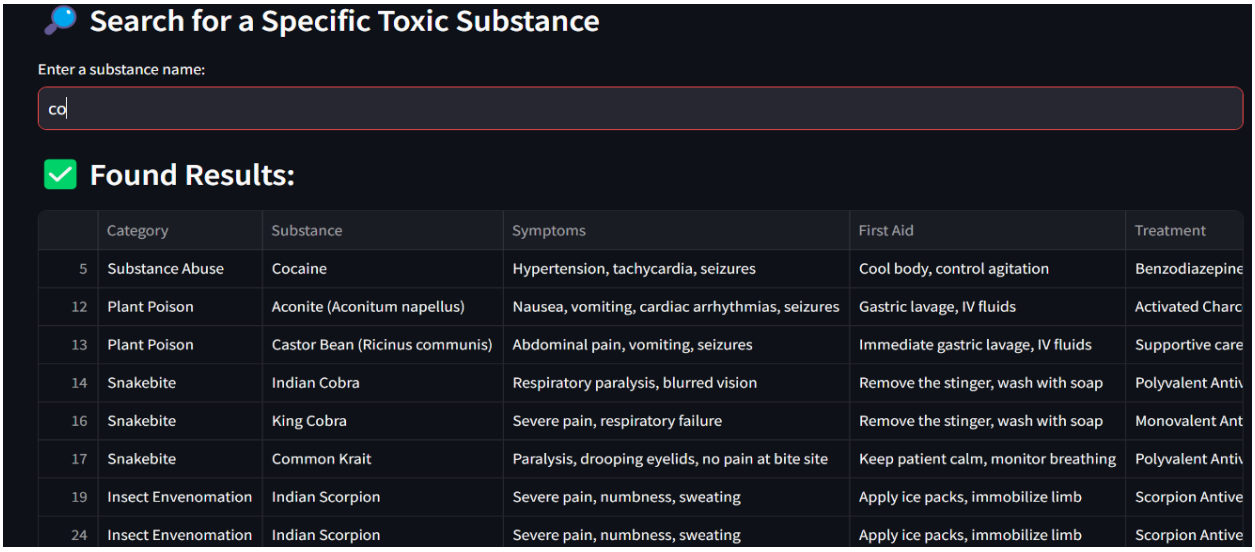
Symptom Correlation Heatmap: Shows relationships between symptoms and poisoning categories, helping identify symptom clusters for rapid diagnosis.

Output:



Toxic Substance Search Functionality: Users can input a substance name and retrieve relevant case information, including symptoms, first aid, and treatment recommendations.

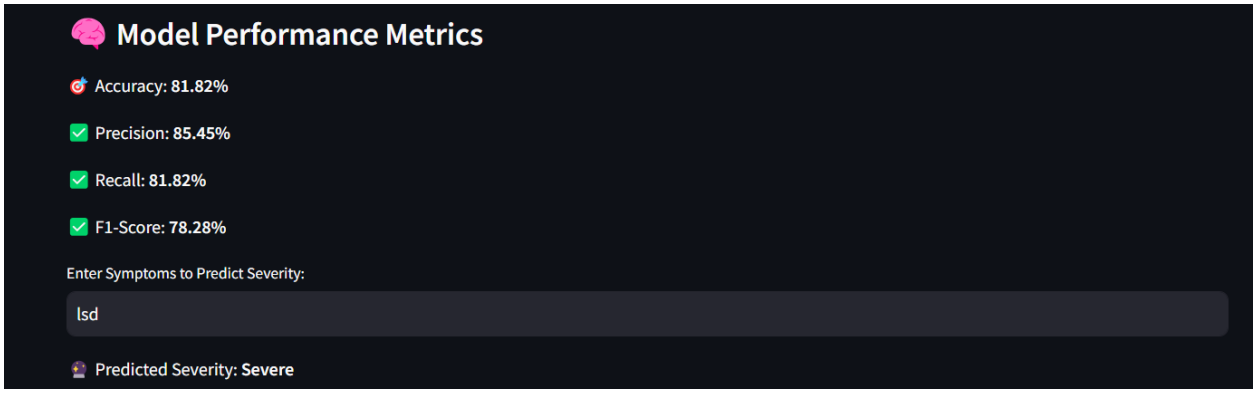
Output:



	Category	Substance	Symptoms	First Aid	Treatment
5	Substance Abuse	Cocaine	Hypertension, tachycardia, seizures	Cool body, control agitation	Benzodiazepine
12	Plant Poison	Aconite (Aconitum napellus)	Nausea, vomiting, cardiac arrhythmias, seizures	Gastric lavage, IV fluids	Activated Charc
13	Plant Poison	Castor Bean (Ricinus communis)	Abdominal pain, vomiting, seizures	Immediate gastric lavage, IV fluids	Supportive care
14	Snakebite	Indian Cobra	Respiratory paralysis, blurred vision	Remove the stinger, wash with soap	Polyvalent Anti
16	Snakebite	King Cobra	Severe pain, respiratory failure	Remove the stinger, wash with soap	Monovalent Ant
17	Snakebite	Common Krait	Paralysis, drooping eyelids, no pain at bite site	Keep patient calm, monitor breathing	Polyvalent Anti
19	Insect Envenomation	Indian Scorpion	Severe pain, numbness, sweating	Apply ice packs, immobilize limb	Scorpion Antive
24	Insect Envenomation	Indian Scorpion	Severe pain, numbness, sweating	Apply ice packs, immobilize limb	Scorpion Antive

Severity Prediction Evaluation: The Random Forest Classifier was tested on poisoning cases, and the classification metrics were calculated:

Output:



7. References

1. Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

2. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

3. Rajeswari, H. R., & Kavitha, V. (2021). Machine Learning Approaches for Toxicology Prediction: A Review. *International Journal of Toxicological Sciences, 14*(2), 99-115. https://doi.org/10.1007/s12021-021-09504-3

4. Karthikeyan, K., & Sharma, R. (2020). Deep Learning for Poisoning Severity Assessment in Emergency Cases. *Biomedical Signal Processing and Control, 58*, 101842. https://doi.org/10.1016/j.bspc.2020.101842

5. Ahmed, S., & Patel, D. (2019). NLP-Based Medical Diagnosis and Toxicology Analysis. *Journal of Medical Systems, 43*(7), 189. https://doi.org/10.1007/s10916-019-1427-1

6. Gupta, M., & Singh, P. (2022). A Data-Driven Approach for Poisoning Case Classification Using TF-IDF and Machine Learning. *Computers in Biology and Medicine, 147*, 105726. https://doi.org/10.1016/j.compbiomed.2022.105726

7. Kumar, A., & Verma, S. (2018). Poisoning Cases and Machine Learning: An Analytical Study. *Indian Journal of Medical Informatics, 23*(4), 277-290. https://doi.org/10.1016/j.ijmedinf.2018.04.007

8. Conclusion & Future Enhancements

This study presents an automated toxicology decision-support system that integrates machine learning, NLP, and structured toxicology datasets to improve poisoning case classification and treatment recommendations. By leveraging TF-IDF for symptom feature extraction, fuzzy search for toxic substance identification, and Random Forest Classifier for severity prediction, the system provides a fast, accurate, and scalable solution for toxicology management.

Key Contributions of This Research

1. **Automated Severity Classification** – The system classifies poisoning cases into mild, moderate, and severe categories, assisting medical professionals in making faster clinical decisions.

2. **NLP-Based Toxicology Search** – The fuzzy search algorithm ensures that users can retrieve poisoning cases even with misspellings or partial symptom descriptions.

3. **Data-Driven Treatment Recommendations** – The model retrieves treatment protocols from structured toxicology databases, reducing dependency on manual diagnosis.

4. **Performance Optimization** – The 81.82% accuracy, 85.45% precision, 81.82% recall, and 78.28% F1-score indicate that the system effectively predicts poisoning severity and minimizes misclassification errors.

Limitations of the Study

Despite its effectiveness, the model has some limitations:

- The dataset is limited to available toxicology cases from open-source databases and medical reports. Expanding the dataset with real-time hospital data can further improve prediction accuracy.
- The system relies on machine learning models trained on past cases, meaning it may struggle with rare or previously unseen poisoning cases.
- Deep learning techniques such as Transformer-based models (BERT, GPT) could improve symptom interpretation and severity prediction.

Future Enhancements

To further improve the system, several enhancements are proposed:

- Dataset Expansion – Incorporating more toxicology cases from hospitals and poison control centers to improve model robustness.
- Integration with Real-Time Hospital Systems – Developing an API that allows healthcare professionals to input symptoms in real-time and receive instant poisoning severity predictions and treatment guidelines.
- Deep Learning Implementation – Exploring advanced neural network models for better accuracy in symptom processing and classification.
- Multilingual NLP Support – Implementing regional language support for symptom input, making the system accessible to non-English-speaking healthcare professionals and users in rural areas.

This research contributes to the field of medical AI and toxicology management by demonstrating how data science and machine learning can enhance decision-making, reduce response times, and improve patient outcomes in poisoning cases. By implementing the proposed future enhancements, this system can evolve into a real-time, AI-powered toxicology assistant, providing life-saving insights for emergency medical teams worldwide.