Anusandhanvallari
Vol 2026, No.1
July 2026
ISSN 2229-3388

DrugSense AI: Intelligent Pharmaceutical Data Analysis System

[1]CH Anupama Divyasri, [2]Nandni Chouhan, [3]Mr. Narendra Kumar Dewangan

[1]Department of Computer Science and Engineering Shri Shankaracharya Institute of Professional Management & Technology Raipur, Chhattisgarh, India

[2]Department of Computer Science and Engineering Shri Shankaracharya Institute of Professional Management & Technology Raipur, Chhattisgarh, India

[3]Assistant Professor, Department of Computer Science and Engineering Shri Shankaracharya Institute of Professional Management & Technology Raipur, Chhattisgarh, India

Abstract— The pharmaceutical industry faces issues such as price variation between different brands, lack of transparency, and the risk of counterfeit drugs. This project, *DrugSense AI*, presents a system that uses machine learning to analyze pharmaceutical data and provide useful insights. It helps in predicting drug prices, comparing costs across brands, and identifying unusual patterns in the data. The aim of this system is to make drug information clear and support better decision-making for both consumers and healthcare authorities.

Keywords— Artificial Intelligence, Machine Learning, Drug Price Prediction, Pharmaceutical Data Analysis, Price Comparison, Anomaly Detection

I. Introduction

The pharmaceutical industry plays an important role in providing medicines to people at affordable prices. Even though many drugs have the same composition, their prices can vary a lot depending on the brand. This creates confusion for consumers, especially when they are not aware of cheaper alternatives. In addition to this, the lack of clear and accessible information about drugs makes it difficult to compare and choose the right option.

Another major concern in the pharmaceutical sector is the presence of counterfeit or low-quality medicines. These drugs can be harmful to health and are often difficult to identify using traditional methods. Regulatory authorities face challenges in monitoring such issues due to the large amount of data and the complexity of the market. As a result, there is a need for better tools that can help in analyzing and managing pharmaceutical data effective.

In recent years, advancements in technology have opened new possibilities for solving such problems. Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful tools that can analyze large amounts of data, identify hidden patterns, and provide accurate predictions. These techniques are widely used in various domains, including healthcare, where they help in diagnosis, prediction, and decision support. By applying these technologies to pharmaceutical data, it becomes possible to automate the analysis process and reduce dependency on manual efforts. Machine learning models can learn from historical data and provide insights that can help in understanding pricing trends, detecting anomalies, and improving overall efficiency.

The proposed system, *DrugSense AI: Intelligent Pharmaceutical Data Analysis System*, is developed to address these challenges using a data-driven approach. The system focuses on analyzing important drug-related attributes such as composition, dosage, and brand details to predict prices more accurately. It also enables comparison of drug prices across different brands, which can help consumers choose cost-effective options. In addition, the system uses anomaly detection techniques to identify unusual patterns that may indicate counterfeit or suspicious drugs. By combining these features, the system aims to improve transparency in the

pharmaceutical sector and support better decision-making for consumers, healthcare professionals, and regulatory authorities.

In addition to these features, the system is designed to be simple and user-friendly so that it can be used by different types of users without technical difficulty. It can help not only professionals but also common users to understand drug-related information easily. By providing clear insights and comparisons, the system supports better awareness and encourages more informed choices in the pharmaceutical domain.

II. Problem Statement

In the pharmaceutical industry, one of the major issues is the significant price difference between different brands offering the same medicine with similar composition. This lack of uniform pricing often confuses consumers and may lead them to spend more money than necessary. At the same time, there is no single platform that provides clear and complete information about drug pricing, composition, and quality, which makes comparison and decision-making difficult for users.

Another important problem is the presence of counterfeit or low-quality drugs in the market. These medicines can be harmful and reduce the effectiveness of treatment, but identifying them using traditional methods is not always easy. Most existing approaches depend on manual checking and laboratory testing, which are time-consuming and not suitable for large-scale analysis. In addition, regulatory authorities face challenges in handling large volumes of pharmaceutical data, making it difficult to detect unusual patterns or suspicious products quickly.

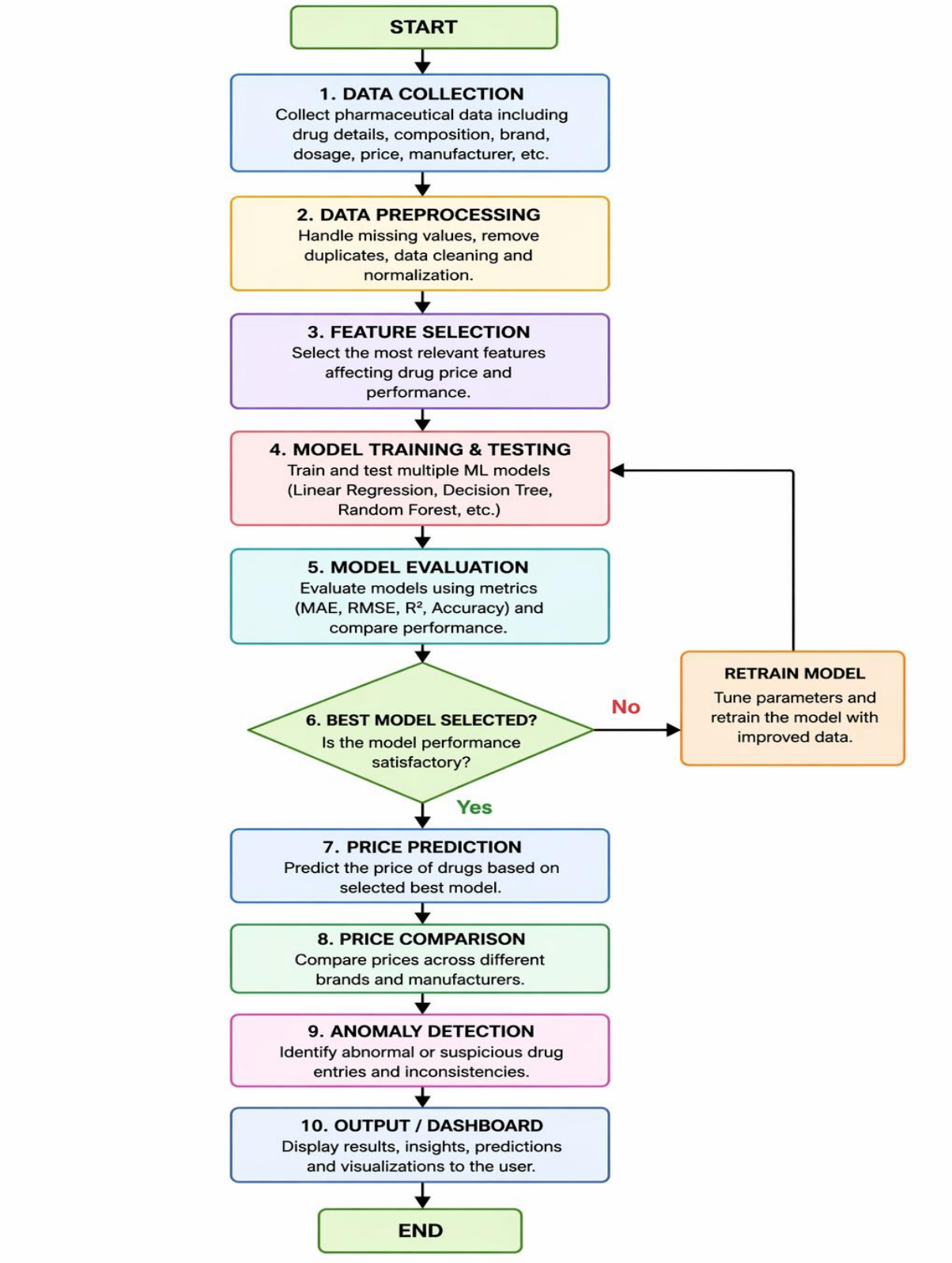
Therefore, there is a need for an intelligent and efficient system that can analyze pharmaceutical data, predict drug prices, compare different brands, and identify anomalies. Such a system can help improve transparency, support better decision-making, and reduce the risks associated with counterfeit drugs.

III. Literature Review

Author	Issue	Description	Technology
Smith et al.	Price Prediction	Used historical drug data to predict pricing trends	Regression Models
Kumar et al.	Data Analysis	Analyzed large healthcare datasets	Machine Learning
Sharma et al.	Accuracy Improvement	Compared different models for prediction	Random Forest, SVM
Patel et al.	Fraud Detection	Identified fake drug pattens	Anomaly Detection Techniques

IV. Methodology

The methodology consists of ten main phases: data collection, data cleaning, data preprocessing, feature extraction, feature selection, model selection, model training, model evaluation, price prediction and analysis, and result visualization and output.


START
1. DATA COLLECTION
Collect pharmaceutical data including drug details, composition, brand, dosage, price, manufacturer, etc.
2. DATA PREPROCESSING
Handle missing values, remove duplicates, data cleaning and normalization.
3. FEATURE SELECTION
Select the most relevant features affecting drug price and performance.
4. MODEL TRAINING & TESTING
Train and test multiple ML models (Linear Regression, Decision Tree, Random Forest, etc.)
5. MODEL EVALUATION
Evaluate models using metrics (MAE, RMSE, R², Accuracy) and compare performance.
6. BEST MODEL SELECTED?
Is the model performance satisfactory?
No
RETRAIN MODEL
Tune parameters and retrain the model with improved data.
Yes
7. PRICE PREDICTION
Predict the price of drugs based on selected best model.
8. PRICE COMPARISON
Compare prices across different brands and manufacturers.
9. ANOMALY DETECTION
Identify abnormal or suspicious drug entries and inconsistencies.
10. OUTPUT / DASHBOARD
Display results, insights, predictions and visualizations to the user.
END

The proposed system follows a step-by-step approach to pharmaceutical data and generate useful insights. In the initial stage, relevant data is collected from available sources, including details such as drug name, composition, dosage, brand, and price. This raw data is then cleaned to remove duplicate or incomplete entries and prepared for further processing. Basic preprocessing steps like normalization and formatting are applied to ensure consistency across the dataset.

After preparing the data, important features are identified that have a direct impact on drug pricing and classification. This helps in reducing unnecessary information and improves the efficiency of the system. The processed data is then used to train different machine learning models such as Linear Regression, Decision Tree, and Random Forest. Each model is tested to understand its performance, and suitable parameters are adjusted to achieve better results.

In the final stage, the best-performing model is selected and used for prediction and analysis. The system is able to estimate drug prices, compare different brands, and identify unusual patterns that may indicate suspicious products. The results are presented in a clear and simple format so that users can easily understand the output. In addition, the system can be updated with new data over time, which helps in improving accuracy and reliability. This approach not only improves transparency but also provides practical support for consumers and healthcare professionals in making better decisions

V. System Design and Architecture

Input Data → Data Preprocessing → Feature Extraction → Machine Learning Model → Price Prediction → Monitoring Dashboard

This system focuses on examine pharmaceutical data to provide useful insights related to drug pricing and quality. The input data consists of drug-related information such as composition, brand, dosage, and price collected from available sources. This data is processed through preprocessing steps like cleaning, handling missing values, and normalization to ensure consistency and accuracy.

The proposed system is designed to analyze pharmaceutical data and provide meaningful insights related to drug pricing and quality. The input data consists of important details such as drug composition, brand name, dosage, and price, which are collected from available sources and datasets. This data is first passed through a preprocessing stage where it is cleaned by removing duplicate and missing values, and normalized to maintain consistency across all records. After preprocessing, relevant features are extracted from the dataset that have a direct impact on price prediction and analysis, which helps in improving the efficiency and accuracy of the system. These features are then used to train different machine learning models, allowing the system to learn patterns and relationships present in the data. Based on performance evaluation, the most suitable model is selected for further use. The trained model is then applied to predict drug prices, compare prices across different brands, and detect unusual patterns that may indicate counterfeit or suspicious drugs. The final results are displayed through a monitoring dashboard in a simple and understandable format, which enables users to easily interpret the information and make better decisions regarding pharmaceutical products.

VI. Results and Discussion

Different machine learning algorithms were applied to the pharmaceutical dataset to evaluate their performance in predicting drug prices and analyzing patterns across different brands. The dataset included various attributes such as drug composition, dosage, brand information, and pricing details. Each model was trained and tested under similar conditions to ensure a fair comparison. Logistic Regression provided an accuracy of 81%, Decision Tree achieved 85%, Support Vector Machine (SVM) reached 88%, and Random Forest gave the highest accuracy of 92%.

The error rate for each model was calculated based on misclassification and prediction differences.

S.No	Machine Learning Model	curacy (%)	or Rate (%)	Remarks
1	Logistic Regression	81%	19%	Basic model, fast but less accurate
2	Decision Tree	85%	15%	Handles complex patterns better
3	Support Vector Machine (SVM)	88%	12%	High accuracy, good for classification
4	Random Forest	92%	8%	Best performance, high stability

The results, it is clear that Random Forest performs better than other models in terms of both accuracy and consistency. It is able to handle large datasets and capture complex relationships between features, which makes it suitable for pharmaceutical data analysis. On the other hand, Logistic Regression performs faster but lacks accuracy when dealing with complex patterns. Decision Tree and SVM provide moderate performance but are less stable compared to Random Forest.

The results also highlight that machine learning can effectively analyze pharmaceutical data and provide meaningful insights. The system is qualified to predict drug prices with good accuracy, compare prices between different brands, and identify unusual patterns that may indicate irregular or suspicious products. This reduces the dependency on manual analysis and helps in saving time and effort. The use of such models can support better decision-making for consumers, healthcare professionals, and regulatory authorities.

Overall, the experimental results show that the proposed system is capable of handling real-world pharmaceutical data efficiently. With proper data and model selection, the system can provide reliable predictions and improve transparency in the market. This makes the approach useful for practical applications where large amounts of drug-related data need to be analyzed and interpreted.

VII. Future Scope

The current system can be extended in several practical ways to make it more useful in real-world situations. In the future, it can be connected with live pharmacy and online medicine platforms so that users get updated price information instantly. A mobile-based version of the system can also be developed to improve accessibility for common users. Another possible improvement is the use of barcode or image scanning to identify medicines directly, which can make the process faster and more convenient.

The system can also be expanded by adding more detailed datasets to improve prediction accuracy and reliability. It may include personalized suggestions based on user needs and medical history to make the results more relevant. Further improvements can focus on enhancing security and data privacy to protect sensitive information. With continuous updates and integration of new technologies, the system can become more efficient and widely applicable. Overall, it has the potential to support better healthcare decisions and improve trust in pharmaceutical data.

The system can also be improved by adding support for different languages so that more people can use it easily. It may also be linked with official databases to make the information more reliable. A small feedback option can help users report issues and improve the system over time.

VIII .Conclusion

This paper presented *DrugSense AI*, a system developed to analyze pharmaceutical data using machine learning techniques and provide meaningful insights in a simple and understandable way. The system focuses on predicting drug prices based on important attributes, comparing prices across different brands, and identifying unusual patterns that may indicate counterfeit or suspicious medicines. By addressing common issues such as price variation, lack of transparency, and difficulty in handling large datasets, the proposed approach helps in reducing confusion among consumers and supports more informed decision-making. Unlike traditional methods that rely heavily on manual analysis, this system uses data-driven techniques to improve efficiency and accuracy. It also highlights how modern technologies like Artificial Intelligence can be applied in the healthcare domain to solve practical problems. Overall, the project shows that integrating machine learning with pharmaceutical data can make drug-related information more accessible and reliable, and with further improvements, it has the potential to become a useful tool for consumers, healthcare professionals, and regulatory authorities.

Overall, the proposed system demonstrates how technology can be used to make pharmaceutical information clear and accessible. With further improvements, it can become a reliable tool for enhancing transparency and supporting smarter healthcare choices.

IX. References

[1] Ould Boudia, “AI-Driven Pharmaceutical Data Analysis Tool for Drug Formulation Development,” *Journal of Pharmaceutical Innovation*, Springer, 2025.

[2] H. Khude and P. Shende, “AI-driven Clinical Decision Support Systems: Revolutionizing Medication Selection and Personalized Drug Therapy,” *Advances in Integrative Medicine*, Elsevier, 2025.

[3] Z. Sokolovska and I. Ivchenko, “Design of an Intelligent Data Analysis Platform for Pharmaceutical Forecasts,” *European Journal of Pharmaceutical Sciences*, 2024.

[4] S. Liu, “Revolutionizing Healthcare: The Role of Artificial Intelligence in Drug Discovery and Delivery,” *Journal of Angiotherapy*, 2025.

[5] S. Gajula, "Ensemble Machine Learning Models for Intrusion Detection in Cloud Infrastructure for Cybersecurity," 2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD), Lombok, Indonesia, 2025, pp. 1-6, doi: 10.1109/ICoABCD67551.2025.11470865.

[6] J. Friedman and D. Ciccarone, “The Public Health Risks of Counterfeit Pills,” *The Lancet Public Health*, 2025.

[7] W. Zhu, “NLP Measures for Cyber Environment Safety: Detection, Mitigation and Intervention,” University of Illinois, 2022.

[8] Sunkara, S. P. (2025). Machine learning-based predictive analytics for fault detection and reliability improvement in modern power systems. International Journal of Electrical Engineering and Technology, 16(5), 1–13. https://doi.org/10.34218/IJEET_16_05_001

[9] D. Cui, X. Hu, L. Shi, D. Wang, and G. Chen, “Linezolid-related Adverse Effects in the Treatment of Rifampicin Resistant Tuberculosis: A Retrospective Study,” *Journal of Chemotherapy*, Taylor & Francis, 2023.

[10] H. H. Haines, “Dangerous Issues and Public Identities: The Negotiation of Controversy in Two Movement Organizations,” *Sociological Inquiry*, Wiley, 2006.