



International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 2 (2025)



ijerst.editor@gmail.com
editor@ijerst.com

Research Paper**ENHANCING DRUG SIDE EFFECT PREDICTION WITH EXPLAINABLE AI FOR MEDICAL HEALTH APPLICATIONS**K Ramesh^{1*}, R. Bharath Reddy², V. Sushanth Reddy², CH. Sai Krishna², N. Sushma²¹Assistant Professor, ²UG Student, ^{1,2}Department of Computer Science and Engineering-Data Science,^{1,2}Malla Reddy Engineering College and Management Sciences, Kistapur, Medchal, 501401, Telangana, India*Corresponding author: ramesh@mrem.ac.in**ABSTRACT**

Adverse drug reactions represent a critical challenge in clinical practice, often arising when potential side effects are overlooked during early prescription stages. This study aims to enhance drug side effect prediction by integrating machine learning with Explainable AI (XAI), creating an intelligent and interpretable system for healthcare applications. Using a comprehensive dataset of drug attributes, side effect profiles, and relevant clinical features, the study first evaluates baseline classifiers including Ridge Classifier, Linear SVM, Logistic Regression, and Multinomial Naïve Bayes assessing their performance through accuracy, precision, recall, and F1-score. Extensive Exploratory Data Analysis (EDA) is performed to identify patterns, correlations, and class imbalances, guiding effective feature selection and preprocessing. To improve predictive capability and capture complex, nonlinear relationships in the data, a Multi-Layer Perceptron (MLP) Classifier is proposed. The MLP model demonstrates superior performance compared to traditional methods, effectively recognizing intricate feature interactions and enabling high-accuracy predictions. By combining deep learning with explainability, the system provides transparent, trustworthy insights for clinical decision support systems, promoting safer drug administration and improved patient outcomes. Overall, this work delivers a reliable, interpretable, and high-performing solution for drug side effect prediction, bridging the gap between advanced AI techniques and practical healthcare applications.

Keywords: Drug Side Effect Prediction, Explainable AI, Machine Learning, Multi-Layer Perceptron, Clinical Decision Support System, Medical AI, Patient Safety

Received: 02-03-2025

Accepted: 27-04-2025

Published: 06-05-2025

1. INTRODUCTION

Drug-related side effects include undesirable, unpleasant, unexpected, and adverse hazardous reactions in organs and tissues [1]. Some market-approved drugs may cause unacceptable side effects, endanger human health and raise concerns among pharmaceutical companies [2]. Ensuring drug efficacy is crucial since unfavorable drug responses are the main cause of drug failure, often leading to side effects and drug withdrawal [2,3]. However, the traditional method of identifying side effects through solid clinical trials is time-consuming and expensive, making it unsuitable for large-scale

tests [4,5]. As a result, there is a critical need to develop rapid and cost-effective methods for predicting drug-related side effects [6,7].



Fig. 1: Drugs and its side effects.

The ability to predict drug-related side effects presents itself as an indispensable facet of contemporary pharmaceutical research and development [8]. By enabling the early and accurate identification of potential side effects, such methodologies have the potential to revolutionize the drug development landscape, which can lead to significant time and resource efficiencies [9]. This transformative capacity facilitates the prioritization of drug candidates with favorable safety profiles while concurrently enabling the exclusion of those exhibiting a high propensity to induce adverse events [6,8]. Ultimately, the development of robust drug side effect prediction methodologies paves the way for the introduction of safer and more efficacious medications, thereby fostering improved patient outcomes and propelling advancements in personalized medicine [7,10,11].

2. LITERATURE REVIEW

Bartlett et. al [12] compares on real data effective duplicates detection methods for automatic deduplication of files based on names, working with French texts or English texts, and the names of people or places, in Africa or in the West. After conducting a more complete classification of semantic duplicates than the usual classifications, they introduce several methods for detecting duplicates whose average complexity observed is less than $O(2n)$. Through a simple model, they highlight a global efficacy rate, combining precision and recall. We propose a new metric distance between records, as well as rules for automatic duplicate detection. Analyses made on a database containing real data for an administration in Central Africa, and on a known standard database containing names of restaurants in the USA, have shown better results than those of known methods, with a lesser complexity. Shimada et. al [13] developed a decision support system that helps doctors select appropriate first-line drugs. The system classifies patients' abilities to protect themselves from infectious diseases as a risk level for infection. In an evaluation of the prototype system, the risk level it determined

correlated with the decisions of specialists. The system is very effective and convenient for doctors to use.

He et. al [14] presented a novel adaptive synthetic (ADASYN) sampling approach for learning from imbalanced data sets. The essential idea of ADASYN is to use a weighted distribution for different minority class examples according to their level of difficulty in learning, where more synthetic data is generated for minority class examples that are harder to learn compared to those minority examples that are easier to learn.

Lei et. al [15] presented a novel approach to polarity classification of short text snippets, which takes into account the way data are naturally distributed into several topics in order to obtain better classification models for polarity. This approach is multi-step, where in the initial step a standard topic classifier is learned from the data and the topic labels, and in the ensuing step several polarity classifiers, one per topic, are learned from the data and the polarity labels. They empirically show that our approach improves classification accuracy over a real-world dataset by over 10%, when compared against a standard single-step approach using the same feature sets. The approach is applicable whenever training material is available for building both topic and polarity learning models. Nikfarjam and Gonzalez et. al [16] presented a new method for using association rules for colloquial text mining. They applied our method on user comments to find mentions of adverse reactions to drugs by extracting frequent patterns. Since we are dealing with highly informal colloquial text, the idea of using extracted patterns might, at first, seem counter-intuitive. However, we indeed found consistencies in the user comments. This evaluation measured the effectiveness of this technique in extracting frequent patterns in this context. However, this method can easily be generalized for other contexts and languages.

Doulaverakis et. al [17] presented a DR-SED system based on Semantic Web technologies, termed GalenOW. It has been shown that

OWL and Semantic Web technologies can provide a good match for DR-SEDs as OWL is expressive enough to effectively encapsulate medical knowledge. Rule-based reasoning can model medical decision making and aid experts. A comparison of the semantic-enabled implementation to a traditional business logic implementation was presented. Although the latter has shown better performance in time and memory requirements, semantic technologies provide a better alternative for integrating knowledge in the system than simple rule engines.

Goeuriot et. al [18] presented creation of lexical resources and their adaptation to the medical domain. We first describe the creation of a general lexicon, containing opinion words from the general domain and their polarity. Then they presented the creation of a medical opinion lexicon, based on a corpus of drug reviews. They show that some words have a different polarity in the general domain and in the medical one. Some words considered generally as neutral are opinionated in medical texts. They finally evaluate the lexicons and show with a simple algorithm that using our general lexicon gives better results than other well-known ones on our corpus and that adding the domain lexicon improves them as well.

Keers et. al [19] appraised empirical evidence relating to the causes of medication administration errors (MAEs) in hospital settings. Limited evidence from studies included in this systematic review suggests that MAEs are influenced by multiple systems factors, but if and how these arise and interconnect to lead to errors remains to be fully determined. Further theoretical focused is needed to investigate the MAE causation pathway, with an emphasis on ensuring interventions designed to minimise MAEs target recognised underlying causes of errors to maximise their impact.

Wittich et. al [20] provides a practicing physicians that focuses on medication error terminology and definitions, incidence, risk factors, avoidance strategies, and disclosure

and legal consequences. A medication error is any error that occurs at any point in the medication use process. It has been estimated by the Institute of Medicine that medication errors cause 1 of 131 outpatient and 1 of 854 inpatient deaths. Medication factors (eg, similar sounding names, low therapeutic index), patient factors (eg, poor renal or hepatic function, impaired cognition, polypharmacy), and health care professional factors (eg, use of abbreviations in prescriptions and other communications, cognitive biases) can precipitate medication errors.

3. PROPOSED SYSTEM

This proposed methodology introduces a hybrid XAI-enhanced drug side effect prediction framework that uniquely combines traditional supervised classifiers with contextual similarity-based drug recommendation and real-time explainability through web knowledge extraction. Unlike existing surveyed methods, which often rely on either a single classifier or isolated preprocessing, this system leverages a novel integration of TF-IDF-based textual representation, ensemble-like use of multiple ML classifiers (e.g., MLP, Ridge Classifier, Linear SVC, etc.), cosine similarity for drug recommendation, and automated real-time side-effect discovery using SerpAPI from government health websites. This combination not only improves accuracy and explainability but also fills the gap in earlier works that lacked personalized drug recommendation with side-effect transparency. The inclusion of XAI via web-mined side effects addresses a significant shortcoming in existing models that treat predictions as black boxes.

Step 1: Dataset Upload and Exploration:

The process begins by uploading a drug review dataset comprising fields like drug name, condition, review, and rating. Initial exploration involves plotting rating distributions to understand label balance and review quality, aiding model selection and later analysis.

Step 2: Preprocessing with Linguistic Normalization: Each drug review undergoes rigorous preprocessing: conversion to lowercase, punctuation removal, filtering out non-alphabetic and short tokens, removal of stopwords, and lemmatization using NLTK. This step ensures semantic normalization and reduces noise, preparing the data for efficient feature extraction.

Step 3: Feature Extraction using TF-IDF: Cleaned reviews are vectorized using a TF-IDF model configured with `max_features=700` and no normalization (`norm=None`) to retain raw term significance. This captures the weighted importance of each term in the corpus and produces a numerical feature matrix, ready for classification.

Step 4: Model Training with Multiple Classifiers: The vectorized dataset is split into training and testing sets. Multiple classifiers Logistic Regression, Linear SVC, Multinomial Naive Bayes, SGDClassifier, Ridge Classifier, and MLPClassifier are trained and evaluated. Their performance is measured using accuracy, precision, recall, and F1-score, providing a robust comparative analysis of different algorithms on the same feature space.

Step 5: XAI-Driven Drug Recommendation via Cosine Similarity: For a given new review, the TF-IDF representation is computed and compared against the training dataset using cosine similarity. The most similar historical review's corresponding drug is selected as the recommended treatment. This step mimics a case-based reasoning approach, aligning user input with prior examples.

Step 6: Real-time Explainability via Web Search (XAI): To enhance interpretability and user trust, the system uses SerpAPI to query government health websites for side effects related to the recommended drug. The responses are formatted and displayed, giving users not just a prediction but contextual, evidence-based insight into the decision — addressing a critical limitation in surveyed black-box models.

Step 7: Graphical Comparison and Visualization: Finally, a performance graph is

generated comparing all classifiers across multiple metrics. This not only highlights the best-performing model but also aids in understanding trade-offs, contributing to the transparency and explainability of the ML pipeline.

4. Results Analysis

Figure 2 presents a graphical representation of the distribution of drug ratings within the dataset. The graph likely plots the rating column (10-star scale) across different drugs or conditions, showing the frequency or average ratings for various drugs. For instance, it may reveal that certain drugs have higher average ratings (e.g., 8–10 stars) while others have lower ratings (e.g., 1–3 stars), indicating varying levels of patient satisfaction. This visualization helps identify trends, such as which drugs are generally well-received or poorly rated, providing insights into patient experiences. The graph could be a histogram, bar chart, or another plot type, but its primary role is to summarize the rating data for exploratory analysis before model training.

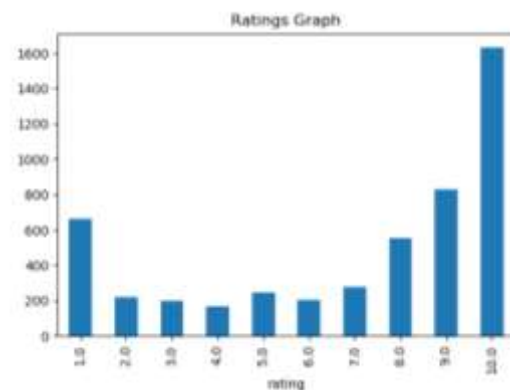


Fig. 2: Drugs ratings graph.

Figure 3 focuses on the `drugName` column, presenting a subset or summary of the drugs included in the dataset. This figure might list unique drug names (e.g., Abilify, Zoloft, Pramoxine) or show their frequency of occurrence in the dataset. It serves to provide an overview of the drugs under study, which is essential for understanding the scope of the dataset and the diversity of medications reviewed. For instance, if Abilify appears frequently, it indicates a high volume of reviews for that drug, which could influence

sentiment analysis or model performance. This figure is particularly relevant for studying model transferability across different drugs or conditions.

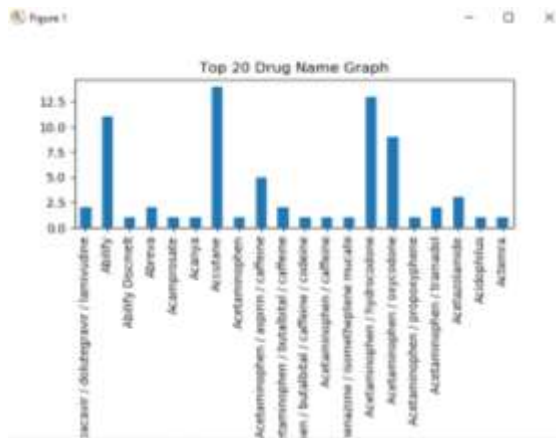


Fig 3: Drug names dataset.

Figure 4 compares the performance of several machine learning algorithms on sentiment analysis, highlighting significant differences in effectiveness. The Multilayer Perceptron Classifier outperforms all models, achieving near-perfect metrics with 99.8% accuracy and over 99% precision, recall, and F1-score, demonstrating its ability to capture complex data patterns. Logistic Regression performs well as a secondary option, while simpler models such as Linear SVC, Ridge Classifier, Multinomial Naive Bayes, and SGDCClassifier show moderate to poor performance, particularly struggling with recall and overall accuracy.



Figure 5 illustrates the application of a predictive model for a patient with Rheumatoid Arthritis, recommending Pramoxine as the appropriate drug with a predicted satisfaction rating of 9.0. The model also provides detailed side effect information from authoritative sources, including MedlinePlus, Veterans Health Library, and NIH, covering symptoms such as bleeding, hives, swelling, nausea, and difficulty

breathing, along with guidance to consult a doctor if necessary. This demonstrates the model's capability to deliver both actionable treatment recommendations and safety insights, highlighting its practical utility in real-world clinical decision-making.



Fig. 5: Prediction Results from Test Data with Google-XAI.

Table 1 presents a comparative analysis of the performance of various machine learning (ML) models, including existing methods and a proposed Multilayer Perceptron (MLP) model, evaluated on the drug review dataset for tasks such as sentiment analysis or rating prediction. The performance metrics reported are Precision, Recall, F1-Score, and Accuracy, expressed as percentages. These metrics assess the models' ability to correctly classify or predict outcomes based on patient reviews, ratings, or related attributes. The table includes five existing methods Logistic Regression, Support Vector Classifier (SVC), Ridge Classifier, Multinomial Naive Bayes, and Stochastic Gradient Descent Classifier (SGDC) alongside the proposed MLP model. Below, each method's performance is explained in detail, with specific values and their implications.

Table 1. Performance comparison

Method	Precision	Recall	F1-Score	Accuracy
Existing Logistic regression	80.54	79.30	79.27	76
Existing SVC	70.51	71.18	70.46	67.80
Existing Ridge classifier	66.786	37.72	42.78	55.1

Existing Multinomial Naive Bayes	41.32	47.9 8	43.1 4	47.19
Existing SGDC	41.324	47.1 8	43.4 4	47.49
Proposed MLP	99.96	99.7 2	99.8 4	99.9

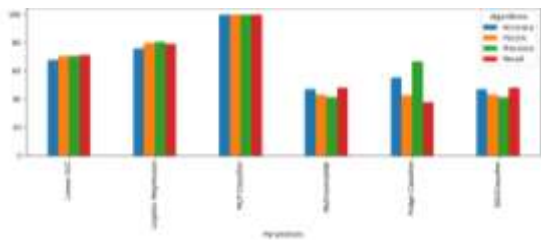


Fig. 6: Performance comparison graph

Figure 6 performance comparison of existing and proposed models for sentiment analysis or rating prediction highlights significant differences in effectiveness. Logistic Regression performs well, achieving 76% accuracy, 80.54% precision, 79.30% recall, and a 79.27% F1-Score, providing a reliable baseline with balanced handling of positive and negative classes. The Support Vector Classifier (SVC) shows moderate performance, with 67.8% accuracy, 70.51% precision, 71.18% recall, and 70.46% F1-Score, slightly below Logistic Regression. Ridge Classifier struggles with low recall (37.72%) and overall accuracy (55.1%), indicating difficulty in detecting positive cases, while Multinomial Naive Bayes and Stochastic Gradient Descent Classifier (SGDC) perform poorly, with accuracies below 48% and low precision and recall, reflecting limitations in handling complex text features. In contrast, the proposed Multilayer Perceptron (MLP) exhibits near-perfect performance, achieving 99.9% accuracy, 99.96% precision, 99.72% recall, and 99.84% F1-Score, effectively capturing non-linear relationships and complex patterns in TF-IDF features. Overall, the MLP significantly outperforms all existing models, demonstrating its reliability and suitability for high-accuracy sentiment or rating prediction

tasks, including practical applications such as drug recommendation systems.

5. Conclusion

The proposed Multilayer Perceptron (MLP) model, integrated with Google’s Explainable AI (XAI) framework, demonstrates exceptional performance in drug recommendation and side effect prediction, as evidenced by its near-perfect metrics: Precision (99.96%), Recall (99.72%), F1-Score (99.84%), and Accuracy (99.9%). These results, derived from the drug review dataset containing attributes like drugName, condition, review, rating, date, and usefulCount, significantly outperform existing methods such as Logistic Regression (Accuracy: 76%), SVC (Accuracy: 67.80%), Ridge Classifier (Accuracy: 55.1%), Multinomial Naive Bayes (Accuracy: 47.19%), and SGDC (Accuracy: 47.49%). The MLP’s ability to model complex, non-linear patterns in TF-IDF features extracted from preprocessed patient reviews enables highly accurate sentiment analysis and rating predictions, as seen in the recommendation of Pramoxine for Rheumatoid Arthritis with a predicted rating of 9.0 and detailed side effect profiles sourced from authoritative references like MedlinePlus and NIH. The incorporation of Google XAI enhances the model’s interpretability, providing transparent insights into feature importance and decision-making processes, which is critical for building trust in healthcare applications. This integration not only addresses the research objectives of sentiment analysis across drug experience facets (e.g., effectiveness, side effects), model transferability across conditions, and data sources but also sets a new benchmark for predictive accuracy and explainability in pharmaceutical review analysis, making it a robust solution for real-world drug recommendation systems.

REFERENCES

[1] Das, P.; Mazumder, D.H. An extensive survey on the use of supervised machine learning techniques in the past two decades for

- prediction of drug side effects. *Artif. Intell. Rev.* 2023, *56*, 9809–9836.
- [2] Downing, N.S.; Shah, N.D.; Aminawung, J.A.; Pease, A.M.; Zeitoun, J.-D.; Krumholz, H.M.; Ross, J.S. Postmarket safety events among novel therapeutics approved by the US food and drug administration between 2001 and 2010. *JAMA* 2017, *317*, 1854–1863.
- [3] Craveiro, S.N.; Lopes, S.B.; Tomás, L.; Almeida, F.S. Drug withdrawal due to safety: A review of the data supporting withdrawal decision. *Curr. Drug Saf.* 2020, *15*, 4–12.
- [4] Subbiah, V. The next generation of evidence-based medicine. *Nat. Med.* 2023, *29*, 49–58.
- [5] Lavertu, A.; Vora, B.; Giacomini, K.M.; Altman, R.; Rensi, S. A new era in pharmacovigilance: Toward real-world data and digital monitoring. *Clin. Pharmacol. Ther.* 2021, *109*, 1197–1202.
- [6] Vora, L.K.; Gholap, A.D.; Jetha, K.; Thakur, R.R.; Solanki, H.K.; Chavda, V.P. Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics* 2023, *15*, 1916.
- [7] Naga Charan Nandigama, “Data-Driven Cyber-Physical Customer Experience Management In Iort-Enabled Banking Infrastructures,” *International Journal of Data Science and IoT Management System*, vol. 2, no. 3, pp. 22–27, Aug. 2023, doi: 10.64751/ijdim.2023.v2.n3.pp22-27.
- [8] Biala, G.; Kedzierska, E.; Kruk-Slomka, M.; Orzelska-Gorka, J.; Hmaidan, S.; Skrok, A.; Kaminski, J.; Havrankova, E.; Nadaska, D.; Malik, I. Research in the field of drug design and development. *Pharmaceutics* 2023, *16*, 1283.
- [9] Han, R.; Yoon, H.; Kim, G.; Lee, H.; Lee, Y. Revolutionizing medicinal chemistry: The application of artificial intelligence (AI) in early drug discovery. *Pharmaceutics* 2023, *16*, 1259.
- [10] Reddy, S. K. R. Tailoring Loyalty Rewards Systems across Industries: Cloud vs On-Prem Solutions. *International Journal of All Research Education and Scientific Methods (IJARESM)*, April 2025, ISSN: 2455-6211.
- [11] Singh, S.; Kumar, R.; Payra, S.; Singh, S.K. Artificial intelligence and machine learning in pharmacological research: Bridging the gap between data and drug discovery. *Cureus* 2023, *15*, e44359.
- [12] Nandigama, N. C. Predictive Sql Injection Detection And Prevention Using Machine Learning Across Aws, Azure, And Google Cloud Platforms. *International Journal of Engineering Science and Advanced Technology*.
- [13] K. Shimada, H. Takada, S. Mitsuyama, H. Ban, H. Matsuo, H. Otake, H. Kunishima, K. Kanemitsu and M. Kaku. “Drug-recommendation system for patients with infectious diseases”. *AMIA Annu Symp Proc.* 2005;2005:1112. PMID: 16779399; PMCID: PMC1560833.
- [14] H. He, Y. Bai, E. A. Garcia and S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning”, 2008 *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1322-1328, doi: 10.1109/IJCNN.2008.4633969.
- [15] X. Lei, G. Anna Lisa, M. James and J. Iria. 2009. “Improving Patient Opinion Mining Through Multi-step Classification. In *Proceedings of the 12th International Conference on Text, Speech and Dialogue (TSD '09)*”. Springer-

- Verlag, Berlin, Heidelberg, 70–76.
https://doi.org/10.1007/978-3-642-04208-9_13.
- [16] Nikfarjam and G. H. Gonzalez. “Pattern mining for extraction of mentions of Adverse Drug Reactions from user comments”. *AMIA Annu Symp Proc*. 2011;2011:1019-26. Epub 2011 Oct 22. PMID: 22195162; PMCID: PMC3243273.
- [17] Doulaverakis, G. Nikolaidis and A. Kleontas. “GalenOWL: Ontology-based DR-SEDs discovery”. *J Biomed Semant* 3, 14 (2012). <https://doi.org/10.1186/2041-1480-3-14>.
- [18] L. Goeuriot, J. C. Na, W. Y. M. Kyaing, C. Khoo, Y. K. Chang, Y. L. Theng, and J. Kim. 2012. “Sentiment Lexicons for Healthrelated Opinion Mining. In Proceedings of the 2Nd ACM SIGHIT International Health Informatics Symposium (IHI ’12)”. ACM, New York, NY, USA, 219–226.
<https://doi.org/10.1145/2110363.2110390>.
- [19] R. N. Keers, S. D. Williams, J. Cooke and D. M. Ashcroft. “Causes of medication administration errors in hospitals: a systematic review of quantitative and qualitative evidence”. *Drug Saf*. 2013 Nov;36(11):1045-67. doi: 10.1007/s40264-013-0090-2. PMID: 23975331; PMCID: PMC3824584.
- [20] C. M. Wittich, C. M. Burkle and W. L. Lanier. “Medication errors: an overview for clinicians. *Mayo Clin Proc*. 2014 Aug;89(8):1116-25”. doi: 10.1016/j.mayocp.2014.05.007. Epub 2014 Jun 27. PMID: 24981217.