International Journal for Research in Biology and Pharmacy

GREEN PUBLICATION

ISSN (Online): 2208-2093
Volume 10 Issue 04 December 2024
Available on-line at: https://ijrbp.com/index.php/bp/issue/view/560

DOI:*https://doi.org/10.69980/bp.v10i4.2500* *Received: 17-Sept-2024 Accepted: 22-Nov-2024*

INTEGRATED ANALYSIS OF PHARMACOVIGILANCE EVENT CHARACTERISTICS AND CLINICAL SAFETY INDICATORS

Dr. Marie Lindquist[1], Dr. Saad Shakir[2], Dr. Priya Bahri[3], Dr. Fatheya Al Awadi[4]

[1] Uppsala Monitoring Centre, Uppsala, Sweden
[2] Drug Safety Research Unit, Southampton, United Kingdom
[3] European Medicines Agency (EMA), Amsterdam, Netherlands
[4] Dubai Hospital, Dubai Health Authority, Dubai, United Arab Emirates

Corresponding Author
Email: marie.lindquist@who-umc.org

Abstract
Adverse drug event detection from biomedical text using computer-based algorithms is becoming more common in pharmacovigilance. Nonetheless, the uneven distribution of events could be an impediment to the detection process of some clinically significant adverse drug events. The present study aimed to analyze pharmacovigilance event characteristics and evaluate machine learning performance for distinguishing adverse events from potential therapeutic events in biomedical text. A quantitative secondary data analysis was conducted using 5,019 annotated biomedical text records. Event distribution, text length, and word count were summarized descriptively. Logistic Regression, Support Vector Machine, and Random Forest classifiers were trained using TF-IDF feature representations. Model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrices. Feature coefficient analysis was additionally performed to identify terms associated with each event class. Adverse events represented the majority of records (n = 4,485), whereas potential therapeutic events accounted for 534 records. Therapeutic-event records showed higher mean text length and word count than adverse-event records. Random Forest achieved the highest accuracy (0.882) but demonstrated very low recall (0.033). Logistic Regression produced the strongest balanced performance, with an accuracy of 0.877, a recall of 0.512, and an F1-score of 0.500. Feature interpretation identified clinically meaningful terms associated with therapeutic improvement and adverse drug reactions. Interpretability of machine learning helped in the classification of pharmacovigilance events, but the imbalance had a large effect on the model's behavior. It is important to have balanced performance and interpretability for the analysis of biomedical safety texts.

Keywords: *pharmacovigilance, adverse drug events, biomedical text, machine learning, natural language processing*

1. Introduction

Pharmacovigilance plays an integral role within healthcare systems since it helps detect, monitor, and prevent medication-related safety risks, including adverse drug reactions. Due to the rising usage of pharmaceuticals, there is an increasing need for the continuous monitoring of drug safety to improve patient outcomes and address associated medication risks (Al-Worafi, 2020). Continuous monitoring takes place after clinical trials as adverse reactions occur in the course of widespread drug use. An ever-growing amount of biomedical information contained within clinical records, reporting systems, and healthcare databases requires the use of computational techniques for pharmacovigilance purposes (Lucas et al., 2022). Advancements in technology have enabled the transformation of pharmacovigilance practices via the automated analysis of biomedical and clinical information. The application of artificial intelligence, machine learning algorithms, and computational analytics is common in adverse drug reaction detection and identification of safety signals in healthcare (Desai, 2024).

The necessity of pharmacovigilance arises from the significance of timely communication of drug reaction signals. Safety signals generated during pharmacovigilance activities contribute significantly to decision-making related to regulatory policies and medication risk evaluation (Sartori et al., 2023). The developments in technology, on the other hand, have expanded the role of biomedical informatics and healthcare IT systems as a means of enhancing contemporary pharmacovigilance activities (Yang & Li, 2025). Such advancements help to process biomedical information effectively and conduct post-marketing surveillance of drug safety. Moreover, investigations related to pharmacovigilance reveal the practical significance of analyzing the safety profiles of drugs. According to previous research, some of the adverse effects associated with medication use include cardiovascular disorders and other related complications (Vestergaard Kvist et al., 2021). Therefore, there is a need for efficient analysis techniques capable of extracting useful information from vast volumes of healthcare data.

Due to a rapid increase in the number of electronic healthcare records and biomedical text, natural language processing (NLP) is commonly applied to pharmacovigilance studies. Numerous applications of NLP include adverse event classification, clinical incident analysis, and healthcare surveillance. Earlier studies have shown that the technique holds substantial potential for improving adverse event detection and healthcare reporting systems (Young et al., 2019). Since clinically meaningful safety information can be embedded within biomedical narratives, the utilization of computational text analysis becomes essential for the extraction of pharmacovigilance data. Machine learning approaches can improve the pharmacovigilance event extraction process via the automated identification of relationships between drugs, adverse reactions, and clinical symptoms. Such approaches might help improve the efficiency and precision of adverse drug event detection compared to the traditional approach (Negi et al., 2019). Present-day developments in the field make it possible to carry out real-time monitoring, automated signal detection, and biomedical text analysis (Al-Worafi, 2023).

Healthcare records became significant sources of data for conducting an analysis of adverse drug events and monitoring the dynamics of medication safety issues among clinical populations. Significant progress in the detection of medication-related issues using biomedical text contributed to improving computational pharmacovigilance systems (Liu et al., 2019). Artificial intelligence also reveals high potential in assessing drug toxicity and carrying out predictive safety analytics within the area of pharmacovigilance (Basile et al., 2019). Likewise, machine learning systems show promising results concerning the early detection of adverse reactions and drug-induced toxicity to ensure patient safety (Panda & Mohapatra, 2024). However, pharmacovigilance practices still suffer from various limitations that must be addressed. For instance, spontaneous reporting systems often encounter issues such as underreporting, lack of sufficient documentation, differences in quality, and biases, which negatively affect safety signal validity (Alomar et al., 2020). Moreover, drug safety monitoring requires constant adjustment due to evolving risks arising from pharmaceutical product use in various clinical contexts (Mehta, 2025).

Even though computational pharmacovigilance and biomedical NLP have been extensively studied, important gaps still exist. In many studies, more attention is paid to algorithm improvement and extraction accuracy while ignoring such aspects as analytical evaluation of pharmacovigilance event characteristics and clinical safety indicators contained within structured biomedical text. At the same time, few research studies explore how NLP-based event extraction could be integrated into the analysis of contextual relationships and clinically meaningful safety indicators.

Thus, the current study was designed to conduct an integrated analysis of characteristics of pharmacovigilance events and clinical safety indicators using structured biomedical textual information. The study aims to analyze pharmacovigilance-related variables, examine clinical safety indicator relationships, and investigate event-related patterns.

2. Methodology

2.1 Research Design

The quantitative approach using secondary data analysis was used for assessing characteristics of pharmacovigilance events and machine learning classification performance in a biomedical text dataset. This paper specifically sought to analyze pharmacovigilance events and text features through statistical and computational techniques. The analyses assessed dataset distribution, text features, and classification performance by machine learning models.

2.2 Data Source

The data set used in this research was taken from the PHEE data set created by Sun et al. (2022). This data set is meant for analyzing adverse drug events from biomedical texts and clinical data and consists of data relevant to the adverse drug events, medical entities, and context. It can be used to analyze healthcare events and biomedical data.

2.3 Study Variables

The variables considered for the analysis include those that were deemed relevant to the analysis of events and biomedical text. Variables that are related to event categories, text features, and biomedical information extracted were considered in carrying out the analysis. Other variables were generated to support the statistical assessment and classification.

2.4 Data Processing and Cleaning

The dataset was preprocessed before any further analysis could be performed on it. This included renaming the variables in order to standardize them so that calculations would be made more consistently. Also, missing values, duplicate records, and formatting issues were addressed before moving forward. Text variables were cleaned and converted before extracting features.

2.5 Statistical and Machine Learning Analysis

The descriptive statistics technique was adopted to describe data attributes such as frequencies, percentages, means, and standard deviations. The analysis of the length and word count of the texts in relation to the different categories of biomedical events was carried out. Texts were classified into the various categories by logistic regression, support vector machine (SVM) and random forest algorithm using term frequency inverse document frequency. The measures for the assessment of model performances included accuracy, precision, recall, F1 score, and confusion matrix. The coefficients of significant features were identified.

3. Results

3.1 Dataset Characteristics and Distribution

The dataset comprised 5,019 biomedical events from text-annotated data that were classified into three datasets (training, testing, and validation). Adverse event types were predominant in the dataset with 4,485 entries, while 534 records belonged to the therapeutic event type category. Training had the most data points, followed by testing and validation. The distribution of the event types in the entire dataset is shown in Table 1.

Table 1. Distribution of biomedical event classes across dataset splits

Dataset Split	Adverse Event	Potential Therapeutic Event	Total
Dev	886	117	1003
Test	889	121	1010
Train	2710	296	3006
Total	4485	534	5019

The significant difference between the number of adverse events and therapeutic events implied possible classifying bias towards the majority class. However, the class ratios were fairly uniform across the training set, validation set, and test set, reflecting stable data splitting.

Figure 1 demonstrates the overall distribution of the dataset. Figure 1A displays the frequency distribution of biomedical event classes, while Figure 1B represents the proportion of adverse and therapeutic events in the training set, validation set, and test set.

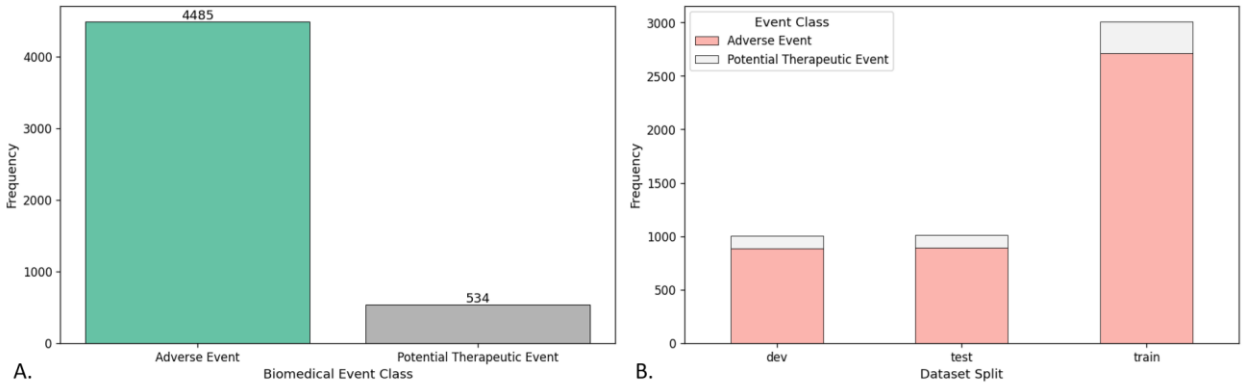

Figure 1. Dataset distribution. (A) Distribution of biomedical event classes. (B) Distribution of event classes across training, validation, and testing subsets.

The biomedical texts displayed moderate variability in length and structural complexity. From the descriptive analysis, the results indicate that adverse event records have a text length of 135.37 characters, while therapeutic event records have a higher text length of 154.52 characters. The same applies to the word count per document, which is relatively higher for therapeutic events compared to adverse events. The descriptive features of the biomedical text records are shown below in Table 2.

Table 2. Descriptive characteristics of biomedical text records by event class

Event Class	Records	Mean Text Length	SD Text Length	Mean Word Count	SD Word Count
Adverse Event	4485	135.37	61.42	18.46	9.22
Potential Therapeutic Event	534	154.52	70.90	21.36	10.66

Figure 2 shows the frequency of biomedical text length for both event types. It can be seen from the histogram that the majority of biomedical documents had around 10 to 30 words, whereas therapeutic event documents had longer texts with more variation than adverse events.

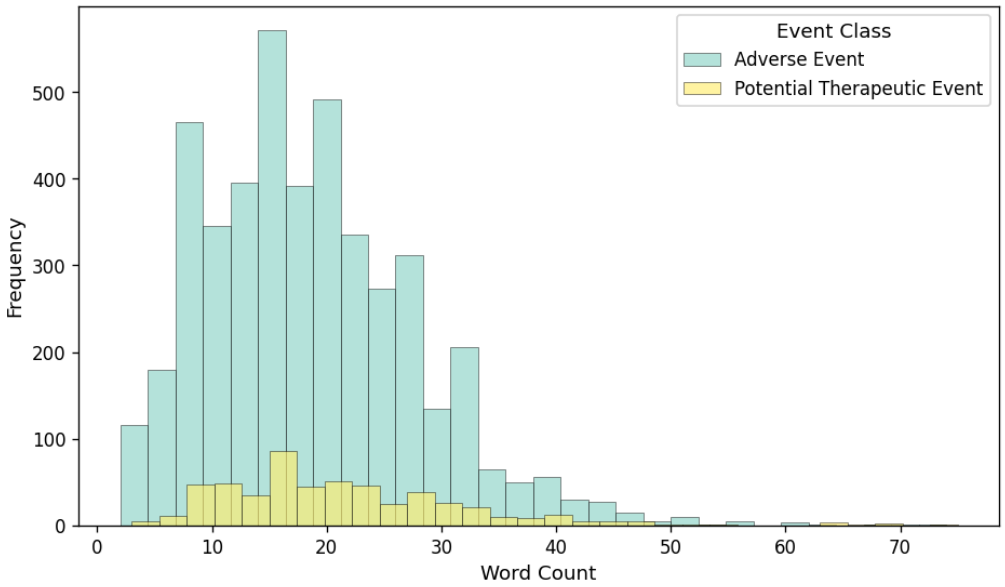

Figure 2. Distribution of biomedical text length by event class.

3.2 Machine Learning Classification Performance

Three machine learning algorithms were considered for biomedical event classification based on TF-IDF features; these algorithms are known as Logistic Regression, SVM, and Random Forest. The performance comparison of these models is shown in Table 3.

Table 3. Comparative performance of machine learning models

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.877	0.488	0.512	0.500
Support Vector Machine	0.876	0.477	0.339	0.396
Random Forest	0.882	0.667	0.033	0.063

Of the tested classifiers, the Logistic Regression classifier scored the best F1-score (0.500), hence demonstrating an excellent precision-recall tradeoff with respect to detecting therapeutic events. While the Random Forest classifier scored the best accuracy score (0.882), its recall value was relatively low, implying that it had limited sensitivity for predicting the minority class of therapeutic events.

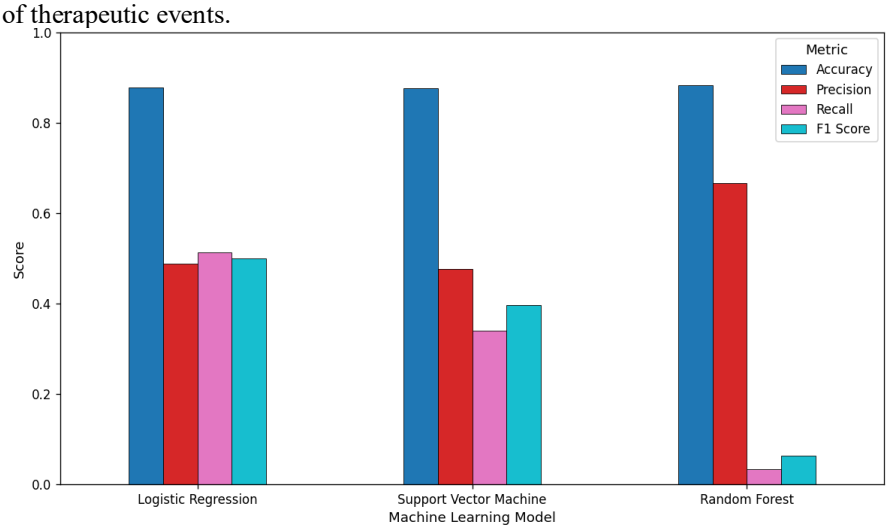

Figure 3. Comparative performance of machine learning models.

The performance analysis revealed that Logistic Regression had the most stable classifying ability in all measurement indicators. On the other hand, Random Forest had considerable performance disparity due to its high accuracy and very poor recall.

3.3 Confusion Matrix Evaluation

The confusion matrix was analyzed for the class-wise predictive ability of each of the ML models. The Logistic Regression classifier predicted 824 cases of adverse events and 62 cases of therapeutic events, while wrongly predicting 65 adverse events and 59 therapeutic events.

Figure 4 provides the confusion matrices of the classifiers used. In Figure 4A, the confusion matrix of the Logistic Regression algorithm is provided, which shows an almost equal performance in the prediction of adverse and therapeutic events. In Figure 4B, the confusion matrix of the SVM algorithm is provided, which shows a good predictive power for adverse events but poor sensitivity for therapeutic events. Figure 4C provides the confusion matrix of the Random Forest algorithm, which shows majority-class dominance.

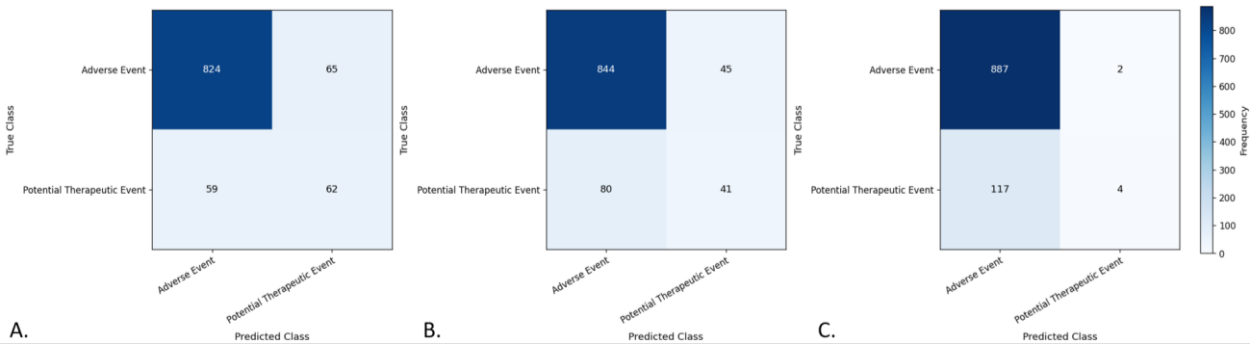

Figure 4. Confusion matrices of evaluated classifiers. (A) Logistic Regression. (B) Support Vector Machine. (C) Random Forest.

According to the confusion matrix analysis, Logistic Regression performed the best-balanced classification method, while Random Forest was significantly biased towards predicting adverse events due to imbalanced data.

3.4 Biomedical Feature Interpretation

For feature coefficients analysis, the Logistic Regression model was used to obtain the highest-ranked TF-IDF features related to each biomedical event category. The classification of therapeutic events was significantly associated with the following keywords: "resolved," "cessation," "treatment," "prescribed," and "effective." These terms indicated effective treatment and improvement of symptoms.

In contrast, the classification of adverse events was linked to the following keywords: "associated," "developed," "syndrome," "toxicity," "adverse," and "reactions." These terms were relevant to pharmacovigilance terminology for adverse biomedical events induced by medication use.

Figure 5 provides the highest-ranked TF-IDF features obtained by applying the Logistic Regression model. Figure 5A depicts the dominant features associated with therapeutic events, while Figure 5B depicts the principal features associated with adverse biomedical events.

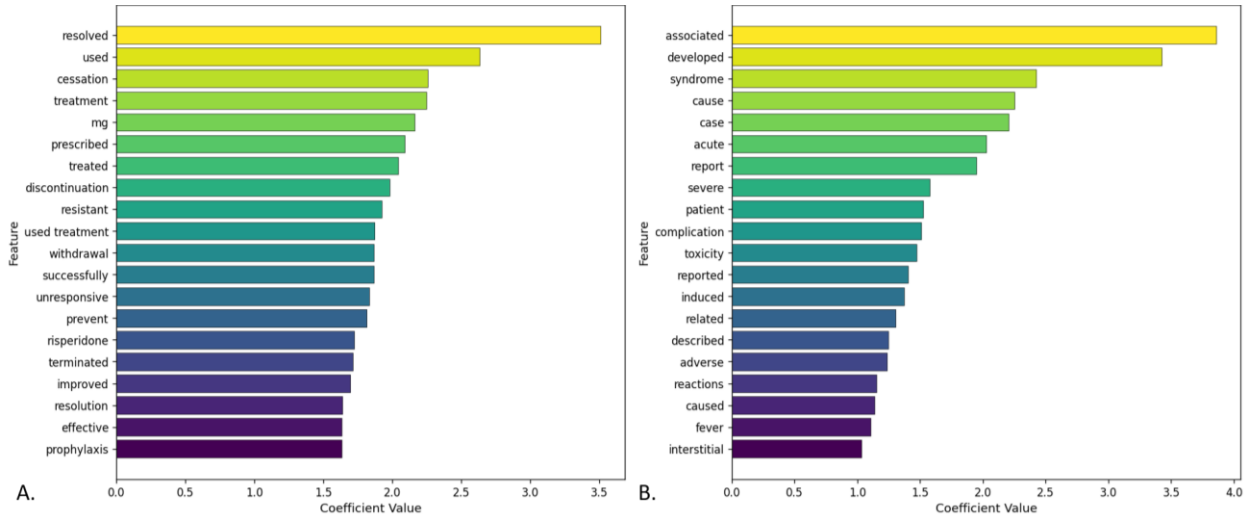

Figure 5. Top TF-IDF features identified by the Logistic Regression model. (A) Features associated with potential therapeutic events. (B) Features associated with adverse events.

In the feature interpretation analysis, it was evident that the classifier captured the clinical relevance of biomedical terms in pharmacovigilance and therapy outcomes.

4. Discussion

In the analysis of the biomedical texts, it was found that adverse-event records dominated the dataset, with potential therapeutic events making up a considerably smaller portion. This is a notable characteristic of many pharmacovigilance data sets since their distribution of clinically relevant event categories is often unbalanced. Since all dataset partitions turned out to be equally representative, there appeared an environment wherein machine learning models performed well with regard to class accuracy, yet had issues recognizing events in the minority class. The textual characteristics confirmed that potential therapeutic-event records contained longer text lengths and a higher number of words compared to the other event class. It is possible that this feature was due to therapeutic-event reports being associated with narrative contexts like the cessation, treatment process, or resolution of some symptoms. On the other hand, adverse-event records proved to be shorter and more frequent due to their association with toxicity, drug reactions, and harm-related language. Of all machine learning algorithms applied to the biomedical data set, logistic regression had the best balance between accuracy and recall, providing the best F1-scores and recall for therapeutic events. Random forest had the highest overall accuracy yet extremely low recall, implying its strong bias towards the majority event class. This once again proves the inadequacy of accuracy as the sole metric for evaluation in imbalanced datasets. Confusion matrix results supported this conclusion by demonstrating that logistic regression models outperformed the others at detecting therapeutic events. The results of feature interpretation revealed that the former was associated with terms like "resolved," "cessation," "treatment," and "effective," while the latter was characterized by such terms as "toxicity," "syndrome," "adverse," and "reactions." These results suggest that the machine learning models captured clinically meaningful information from pharmacovigilance texts.

The predominance of adverse-event records is consistent with the literature on pharmacovigilance, according to which adverse reaction reports make the largest part of drug safety information. Li et al. (2023) concluded that a large-scale adverse-event reporting system could be used to gather valuable information regarding serious drug reactions, in particular, the case of severe cutaneous adverse reactions. As for the reports of serious adverse drug reactions, they were reported to contain valuable information about such aspects as medication exposure, event seriousness, and drug interactions (Létinier et al., 2021). The better performance of logistic regression models compared to complex ones proves that traditional methods of text classification remain useful, despite their lower accuracy in some cases. While recent studies highlight the potential of deep learning and NLP to boost the performance of machine learning models, it should be noted that their predictive power largely depends on such factors as data structure and task complexity (Khemani et al., 2025). Similar observations were made concerning predictive pharmacovigilance models, which were reported to be subject to multiple interpretative caveats when measuring overall performance (De Abreu Ferreira et al., 2024).

The study's conclusions are consistent with current discussions of the application of artificial intelligence technology in pharmacovigilance. In his article, Majekodunmi (2025) emphasizes the significance of artificial intelligence solutions in increasing the capacity of pharmacovigilance and improving the processing of large amounts of information related to the safety of medicines. As he explains, the validity of AI-supported data processing should be ensured by rigorous validation of automated classification algorithms in the context of pharmacovigilance. Similarly, Martin et al. (2022) emphasize that adverse drug reaction report classification by means of AI should be validated in the context of national pharmacovigilance systems. The feature interpretation results are consistent with the findings of surveillance studies, which show that pharmacovigilance data contains information on adverse drug events if properly structured and analyzed. For instance, Sharma et al. (2021) prove the efficacy of adverse drug reaction surveillance in revealing drug safety trends through extended monitoring periods. Moreover, the current results correspond to NLP-based research on adverse-event extraction from pharmacovigilance narratives, confirming their richness in structured drug safety data (Kim et al., 2023).

The study's conclusions are of significant practical relevance for pharmacovigilance. First, class imbalance should be considered when evaluating the machine learning models because high accuracy does not necessarily imply adequate recognition of minority categories. Second, logistic regression can be recommended as an interpretable method for event classification in pharmacovigilance. Third, interpretative feature analysis can help in reviewing the narrative contexts of adverse events and therapeutic outcomes, which could save labor efforts and improve signal prioritization. Fourth, accuracy alone should not be used as a metric of machine learning performance, while precision, recall, F1-score, and confusion matrix should be included in reports. Indeed, neglecting minority categories can be harmful for pharmacovigilance. Integrating machine learning with domain reviews might ensure reliable automated pharmacovigilance workflows.

There were several limitations of this analysis. Being based on the secondary annotated dataset, it was limited by the event types and the dataset structure provided. Class imbalance adversely influenced the results as well. Additionally, this study was limited to TF-IDF-based features and traditional machine learning algorithms, which might not have accounted for semantic depth.

Further research may explore a wider range of text data sources, apply techniques of class balancing, and evaluate models, both conventional and transformer-based. Additional validation of results on independent pharmacovigilance data could provide further insights.

5. Conclusion

Automated pharmacovigilance event analysis demonstrated clear value for distinguishing adverse events from potential therapeutic events in biomedical text. Adverse-event records were substantially more frequent, creating a class imbalance that influenced model performance. Logistic Regression provided the most balanced classification outcome, achieving a stronger F1-score and therapeutic-event recall than Support Vector Machine and Random Forest models. Although Random Forest showed the highest overall accuracy, its low recall confirmed that accuracy alone is insufficient for evaluating pharmacovigilance classifiers in imbalanced datasets. Feature interpretation further showed that the model captured clinically meaningful terminology, with therapeutic events linked to improvement-related terms and adverse events linked to toxicity- and reaction-related expressions. These findings support the use of interpretable machine learning approaches as practical tools for pharmacovigilance event screening and safety narrative analysis. Future work should incorporate larger text sources, class-balancing strategies, transformer-based biomedical language models, and external validation to improve generalizability and clinical utility.

References

1. Sun, Z., Li, J., Pergola, G., Wallace, B. C., John, B., Greene, N., & He, Y. (2022). PHEE: A Dataset for Pharmacovigilance Event Extraction from Text [Data set]. Empirical Methods in Natural Language Processing (EMNLP), Abu Dhabi, UAE. Zenodo. https://doi.org/10.5281/zenodo.7689970
2. Al-Worafi, Y. M. (2020). Pharmacovigilance. In *Drug safety in developing countries* (pp. 29-38). Academic Press.
3. Lucas, S., Ailani, J., Smith, T. R., Abdrabboh, A., Xue, F., & Navetta, M. S. (2022). Pharmacovigilance: reporting requirements throughout a product's lifecycle. *Therapeutic advances in drug safety*, *13*, 20420986221125006.
4. Desai, M. K. (2024). Artificial intelligence in pharmacovigilance–Opportunities and challenges. *Perspectives in Clinical Research*, *15*(3), 116-121.
5. Sartori, D., Aronson, J. K., Norén, G. N., & Onakpoya, I. J. (2023). Signals of adverse drug reactions communicated by pharmacovigilance stakeholders: a scoping review of the global literature. *Drug safety*, *46*(2), 109-120.
6. Yang, J., & Li, F. (2025). The Impact of Technological Progress on Pharmacovigilance. *Pharmacovigilance-Facts, Challenges, Limitations and Opportunities: Facts, Challenges, Limitations and Opportunities*, 43.
7. Vestergaard Kvist, A., Faruque, J., Vallejo-Yagüe, E., Weiler, S., Winter, E. M., & Burden, A. M. (2021). Cardiovascular safety profile of romosozumab: a pharmacovigilance analysis of the US Food and Drug Administration Adverse Event Reporting System (FAERS). *Journal of clinical medicine*, *10*(8), 1660.
8. Young, I. J. B., Luz, S., & Lone, N. (2019). A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. *International journal of medical informatics*, *132*, 103971.
9. Negi, K., Pavuri, A., Patel, L., & Jain, C. (2019). A novel method for drug-adverse event extraction using machine learning. *Informatics in Medicine Unlocked*, *17*, 100190.
10. Al-Worafi, Y. M. (2023). *Technology for drug safety: Current status and future developments*. Springer Nature.
11. Liu, F., Jagannatha, A., & Yu, H. (2019). Towards drug safety surveillance and pharmacovigilance: current progress in detecting medication and adverse drug events from electronic health records. *Drug safety*, *42*(1), 95.
12. Basile, A. O., Yahi, A., & Tatonetti, N. P. (2019). Artificial intelligence for drug toxicity and safety. *Trends in pharmacological sciences*, *40*(9), 624-635.
13. Panda, P., & Mohapatra, R. (2024). Revolutionizing Patient Safety: Machine Learning and AI for the Early Detection of Adverse Drug Reactions and Drug-Induced Toxicity. *Current Artificial Intelligence*.
14. Alomar, M., Tawfiq, A. M., Hassan, N., & Palaian, S. (2020). Post marketing surveillance of suspected adverse drug reactions through spontaneous reporting: current status, challenges and the future. *Therapeutic advances in drug safety*, *11*, 2042098620938595.
15. Mehta, R. (2025). Postmarket drug safety monitoring. In *Translational Gastroenterology* (pp. 395-397). Academic Press.
16. Li, D., Gou, J., Zhu, J., Zhang, T., Liu, F., Zhang, D., ... & Liu, S. (2023). Severe cutaneous adverse reactions to drugs: A real-world pharmacovigilance study using the FDA Adverse Event Reporting System database. *Frontiers in Pharmacology*, *14*, 1117391.
17. Létinier, L., Ferreira, A., Marceron, A., Babin, M., Micallef, J., Miremont-Salamé, G., ... & French Network of Pharmacovigilance Centres. (2021). Spontaneous reports of serious adverse drug reactions resulting from drug–drug interactions: an analysis from the French Pharmacovigilance Database. *Frontiers in Pharmacology*, *11*, 624562.
18. Khemani, B., Malave, S., Shinde, S., Shukla, M., Shikalgar, R., & Talwar, H. (2025). AI-driven pharmacovigilance: Enhancing adverse drug reaction detection with deep learning and NLP. *MethodsX*, *15*, 103460.
19. De Abreu Ferreira, R., Zhong, S., Moureaud, C., Le, M. T., Rothstein, A., Li, X., ... & Patwardhan, M. (2024). A pilot, predictive surveillance model in pharmacovigilance using machine learning approaches. *Advances in Therapy*, *41*(6), 2435-2445.
20. Majekodunmi, E. A. (2025). Strengthening Drug Safety and Public Health Surveillance in the United States: The Role of Artificial Intelligence in Pharmacovigilance. *Available at SSRN 5181179*.
21. Martin, G. L., Jouganous, J., Savidan, R., Bellec, A., Goehrs, C., Benkebil, M., ... & French Network of Pharmacovigilance Centres. (2022). Validation of artificial intelligence to support the automatic coding of patient adverse drug reaction reports, using nationwide pharmacovigilance data. *Drug Safety*, *45*(5), 535-548.

22. Sharma, M., Baghel, R., Thakur, S., & Adwal, S. (2021). Surveillance of adverse drug reactions at an adverse drug reaction monitoring centre in Central India: a 7-year surveillance study. *BMJ open*, *11*(10), e052737.
23. Kim, S., Kang, T., Chung, T. K., Choi, Y., Hong, Y., Jung, K., & Lee, H. (2023). Automatic extraction of comprehensive drug safety information from adverse drug event narratives in the Korea adverse event reporting system using natural language processing techniques. *Drug Safety*, *46*(8), 781-795.