Adverse Drug Event Detection Through Natural Language Processing: AI-Enhanced Pharmacovigilance and Post-Market Safety Surveillance

Dr. Yasemin Şahin, Associate Professor of Electrical and Electronics Engineering, Middle East Technical University (METU), Turkey

1. Introduction

Pharmacovigilance has been crucial in drug development, and it has evolved over time to become instrumental in ensuring drug safety. Post-licensure, extensive efforts are made to monitor adverse effects of drugs by using varied techniques and resources. Pharmacovigilance was never as critical as it is today, with an increasing number of approved drugs on the market and greater numbers of sick patients for whom these drugs are prescribed after having undergone relatively new therapeutic modalities. The passive style of reporting and the conventional system of pharmacovigilance are, in several countries, essential formats for detecting drug safety signals. Despite recent innovations in data structure and collection, AI-based solutions are being proposed to reform the existing situation and are possible alternatives for better managing the large volume of data for prompt signal detection, prospectively to real-world evidence including AI, also known as Real-World Data.

While it is impossible to foresee the unexpected outcomes of a post-approval study due to the involvement of large numbers of varied participants and interventions, their long duration, and confounding variables, it is essential to concentrate on the enhancement of current pharmacovigilance strategies. There are a large number of stakeholders who can profit from the implementation of AI in pharmacovigilance. This may be an essential tool for enhancing patient care and public well-being and for fostering innovative pharmacovigilance research. It is important to concentrate on internalizing AI in post-marketing safety surveillance to complement both conventional and modern techniques. Post-marketing adverse events surveillance is critical in avoiding legal conflicts and following ethics in patient care. It is also interconnected with the growth and progress of

the economic aspects of a health care provider. From this perspective, we propose the use of AI in pharmacovigilance and also outline its roots in drug development.

1.1. Background of Pharmacovigilance and Drug Safety Monitoring

1.1.1. Introduction Drug safety is a preeminent priority for healthcare policymakers and professionals, as harmful or unsuccessful treatment can adversely impact patients when not properly administered. Indeed, the evaluation of drug safety represents a key challenge of pharmacovigilance, the pivotal field of pharmacology that addresses the study and methodology of drug safety mechanisms of action across varying levels. Historically, however, the field changed gears, concentrating primarily on the treatment of drug reactions, largely disregarding prevention strategies. Since its inception and under the influence of international organizations and regulatory institutions, pharmacologists and regulatory authorities on a global scale have emphasized that advances in interdisciplinary sciences should be directed within the framework of best technological practices, legal considerations, and human, animal, and environmental health and well-being. Pharmacovigilance seeks to accomplish this by achieving primary objectives concerning Adverse Drug Reactions (ADRs), such as understanding, detecting, minimizing the risk of, and effecting post-marketing surveillance of those reactions. Drug safety monitoring can improve public health and society's welfare, as healthcare providers are able to make more informed decisions based on previous adverse drug reactions and data. Can similar prospective contributions to theoreticians and researchers who conduct research on drug safety monitoring be identified or assumed? If so, what are these prospective contributions? Alternatively, if such contributions are negative or null, why is this so? Currently, pharmacovigilance's global vision, design, implementation, and evaluation are directed by various international organizations. Numerous challenges face the drug safety and adverse event space, both legally and philanthropically. Regulatory bodies in progressively growing countries require the use of such notice cards, but most still prefer manual entries by pharmacists, doctors, and healthcare personnel. Moreover, in certain organizations, drugs and medical conditions are classified by both the United States Pharmacopoeia and the International Classification of Diseases. Due to the growing number of medical history cards and prevailing manual techniques, individual entries that do not account for extensive standardized terminologies or nomenclature are required. This classification should be performed by people who are experienced and specialized in this field due to

its high level of complexity. This is due to the implementation of uncomplicated data collection systems that result in wide-ranging inconsistencies. Drug use in growing countries, especially, and prescribing patterns closely resemble those in established countries. Pharmacovigilance's output differs depending on the volume of the population and the size of the drug's market. Given the rising adverse drug reaction rates, continuous medical education should be considered as a tool in validating the design of pharmacovigilance. Public health policy must allow for learning from mistakes. Pharmacovigilance and healthcare organizations can only absorb essential requirements. This axiom suggests a high degree of relationship between healthcare and pharmacovigilance. Because of all the conditions, pharmacovigilance's initiative caters to CSRs. Market approval isn't usually one of the characteristics of pharmacovigilance that is met. A therapeutic medico-theory has a pharmacopoeia where a minimum number of medication errors are adequate because risk-free treatment is not advocated.

1.2. Importance of AI in Enhancing Pharmacovigilance

Artificial intelligence (AI) has an important role in the modernization of pharmacovigilance. The relevance of innovative computational methodologies, such as AI, and analytical solutions are the topical dimensions of research interests in pharmacovigilance and therapeutic management. The comprehensive background of AI provides opportunities to the medical and healthcare community to refine their approaches toward adverse drug reaction evaluation, modeling, prediction, signal detection, and data-mining capabilities. AI, as a branch of computer science, initiated some pinpointed execution-based capabilities with the instructions: multiple cycles of machine learning activities and the exploitation of recurrent neural networks, logical algorithms, and related methods to improve the functionalities of computer-based systems. Various AI-based solutions for clinical decision support systems enable healthcare policymakers and professionals to better prescribe drugs, make more accurate diagnoses, provide prognosis, and manage pathological conditions efficiently.

Enormous datasets from large numbers of diversified sources and databases are peculiar features of pharmacoepidemiological and pharmacovigilance research. The employment of AI methodologies promises to accelerate the identification and assessment of adverse drug reactions (ADRs). The AI-inspired tireless behavior enables different algorithms to deal with huge datasets and speeds up the assessment stages. AI methodologies reduce

the burden of certain tasks for healthcare professionals and members of health systems. They also shift the priorities for safety management and allow resources to be moved to activities with the highest added value. The predictive modeling capacities of AI algorithms hint at the potential underpinning possibilities to anticipate forthcoming safety signal behaviors from one era to another. The machine learning algorithms would provide an optimal solution for certain applications such as assessing large observational healthcare data to draw early signals from randomized controlled trials.

2. Machine Learning Techniques in Pharmacovigilance

Pharmacovigilance is important for ensuring post-marketing safety and effectiveness of drugs; thus, its accurate and efficient performance is of paramount importance. In this context, machine learning techniques can potentially play a major role in pharmacovigilance practices. Based on methodologies and applications, machine learning techniques can be broadly categorized into the following: supervised learning, unsupervised learning, semi-supervised learning, active learning, and reinforcement learning. This is useful for those seeking to determine which machine learning technique to use for an application based on the peculiarities of the surroundings. We focus on the methodologies and the application of these machine learning techniques in pharmacovigilance to provide a critical resource for those wishing to use any of the machine learning techniques in pharmacovigilance applications.

Predictive modeling in the context of supervised learning has been a focused point of discussions on pharmacovigilance. Supervised learning techniques are predictive in nature and are capable of identifying hidden patterns in the available historical data of drug-event associations in various forms. We will describe different existing predictive modeling methods such as logistic regression, decision tree, support vector machine, random forest, gradient boosting, as well as convolutional neural networks. These unsupervised learning techniques are model-driven and can handle missing and unstandardized conditions more freely, and can generalize the other standard classification models on these datasets to a larger extent. The subsequent part of the chapter is divided into a further discussion concerning unsupervised learning of the clustering techniques (including hierarchical clustering, k-prototypes, and density-based algorithms). Their potential advantages and disadvantages are also highlighted. Deep learning is an area of machine learning that has witnessed several advancements. Such

techniques can represent a crucial turning point in the history of pharmacovigilance applications. These deep learning techniques can handle and extract hidden features from complex, unstructured, and high-dimensional data.

2.1. Supervised Learning Methods

Supervised learning is a class of data-exploitation algorithms that starts with a labeled dataset. In the context of ADR prediction, such a dataset contains examples of patients to whom a putative cause for the ADR has been administered, and the known outcome of these administrations. Training commences when the algorithm learns from these examples, molds a predictive model to make similar judgments, and generates a resulting classifier. Once established, the model can deduce allowable actions for any input instance, such as predicting whether a patient with unknown drug-exposure status will develop an ADR. When used in practice, the algorithm's predictions are contrasted against the actual truth to perform explicit evaluation. The process of learning from known examples is the core paradigm followed by these classifiers.

Commonly featured supervised learning algorithms are used to predict ADRs. Some of these algorithms, such as naive Bayes, logistic regression, k-nearest neighbors, and decision trees, among others, have been tested in pharmacovigilance and shown to successfully predict potential ADRs from both spontaneous reports and randomized controlled trials. The choice of the learning algorithm depends on its precision in the context of a specific application area, capacity to evaluate the relevance of input features, and fundamentally, its adaptability. Evaluation of these algorithms can also be performed by cross-validation or using a test set. In addition to choosing the best model, in drug safety prediction, special attention should be given to feature selection and feature engineering, as well as correction of class imbalance and overfitting. Modern kernel-based and tree models can handle non-standard data representations and may be advantageous for these tasks. Model validation using independent datasets could also be essential for validating the predictive accuracy of the model. Supervised learning is an indispensable technique for finding associations between safety and reported medical outcomes, and natural language processing is proving to be of great interest to pharmacovigilance. The ability of the various reporting systems to act as potential gold-standard data for model learning requires further investigation to promote their role in other studies on pharmacovigilance as well as precision and personalized medicine.

Model performance can drastically influence the translation and implementation of such models for practical pharmacovigilance use cases.

2.2. Unsupervised Learning Methods

Unsupervised learning methods are used to analyze data when there are no pre-labeled target outcomes of interest. Such methods can be beneficial during the initial exploratory analysis since they offer better insight into the structure of the data as well as the inherent patterns and relationships among drugs and their effects that are otherwise undetectable. Techniques commonly employed for unsupervised methods include clustering, visualization, and dimensionality reduction, and are briefly described as follows. Clustering is used for grouping similar items together with reference to their patterns, while visualization is a manner of presenting the conceptual clustering to be more easily understood. Dimensionality reduction is also commonly employed in visualizing the landscape of features, as the high dimension in features can sometimes prohibit meaningful visualization.

One of the main advantages of unsupervised methods is their adaptability and flexibility for any kind of natural language and clinical data or information which provides deep disease and drug insights. Indeed, unsupervised methods yield the possibility to extract informative patterns and structures in non-targeted text data. As an illustrative example of the application of unsupervised methods, clustering techniques were applied on drug data and compared the adverse drug reactions obtained from this method with text mining algorithms. In a preliminary analysis, it was shown that it is possible to reveal useful insights into the adverse drug reactions and their associations with medicine use. Although unsupervised methods are crucial to exploring untapped pharmacovigilance domains, these techniques have several challenges as well. For instance, when clustering, there may not be adequate guidance on the best number of clusters yet; it is unclear which association rules on the entire feature sets and a selection of methods are useful to organize the complex data. Moreover, due to the lack of labels, it is difficult to validate the models' outputs. In order to overcome the above limitations, unsupervised methods are usually integrated with other methods.

2.3. Deep Learning Approaches

Deep learning is a great subset of machine learning that has lately caught much attention from the scientific community. It is commonly distinguished by the use of artificial

neural networks or deep learning architectures, which can extract complex patterns from vast datasets. In particular, its deep hierarchical structure enables the system to automatically learn multiple levels of representations and abstractions that make it possible to build and train learning models. The main advantage of deep learning methods is their ability to meticulously identify and process large volumes of pertinent complex data, and eventually provide outputs across categories of interest. With the progression of different methods and computational tools, ever more effectual classification approaches have been developed based on variations of the deep learning principles. In the field of pharmacovigilance, deep learning technology was utilized for the automatization of adverse drug reaction event classification on the basis of patient-specific prediction, with a representation of relations between the action domain and the adverse events. Research in the early stages of deep learning focused mainly on developing feed-forward neural networks, the most common type of architecture in deep learning. Convolutional Neural Networks and Recurrent Neural Networks are the two most common types of deep learning methods employed in pharmacovigilance. CNN has been widely used in drug safety monitoring for imaging, especially for adverse drug reaction detection using disease entity, text description, and the definition of adverse events in pictures, while RNN is mainly utilized in natural language processing tasks relating to textual data mining in pharmacovigilance. Other types also include autoencoders, Boltzmann machines, adaptive resonance theory, restricted Boltzmann machines, long short-term memory, and so on in pharmacovigilance research.

In fact, several methods grounded in deep learning have been suggested for multiple adverse drug reaction impediments and potential applications. There are other diverse validation techniques in utilizing artificial intelligence in pharmacovigilance, for instance, ensemble methods with machine learning algorithms, in trying to enhance the performance of the system for earlier detection of adverse drug reactions. However, deep learning models have their obstacles, particularly the need for comprehensive training data, as the model undertakes training from scratch by the data. As a considerable validation aspect, the high dimensionality of the analyzed dataset must be supervised. Several case studies have deployed such methods in the area of pharmacovigilance, particularly in the endeavor to alleviate the early detection of adverse drug reactions not only in the reaction's rate intensity but also in specific bodily

tissue signals. Deep learning methods can also bolster the risk evaluation for the pharmacovigilance process in harmony with other methods, and they are regarded as instrumental when the system gradually re-trains the model. These methods also express their unique strategies in a given setup, depending on the techniques used. Irrespective of the methods used, it is commendable to validate them on new data and at general prospective levels to corroborate any useful knowledge.

3. Data Sources and Preprocessing

To support pharmacovigilance, different data sources are currently in use. Based on their nature, they can be categorized as either structured or unstructured. Structured data primarily include information from clinical trials and electronic health records. There is a rich body of literature on the use of structured data for signal detection, signal refinement, and the generation of evidence regarding the safety aspects of marketed medicines. Unstructured data include social media, medical literature, adverse event reports, and more. Text created by individuals is hard to analyze. As a result, a new discipline, natural language processing, was developed to address this problem by providing an alternative to traditional statistical approaches. The data need to be integrated in a way that is suitable for analysis. The data preprocessing step is a crucial phase in order to provide accurate findings about drug safety. Data have certain characteristics that must be known, and suitable preprocessing actions must be applied depending on each characteristic of the data.

Furthermore, it should also be noted that data integration and preprocessing are critical for the deployment of machine learning and deep learning models in different pharmacovigilance applications. The first phase in data analytics is data preparation, which includes data sampling, data cleaning, data normalization, and data integration in the case of multiple sources. In general, the quality and integrity of analytics are only as good as the data being analyzed. Therefore, missing values, noisy values, non-relevant data, and extracting or filtering the underlying patterns have adverse effects on the analysis. Thus, it is important to preprocess data to facilitate efforts in handling missing values and outliers. For instance, social media data can be filled with slang, different abbreviations, multimedia, and contradictory and unclear data, making this information complex and messy.

3.1. Structured Data from Clinical Trials and Electronic Health Records

3.1. Structured data sources such as clinical trials or electronic health records are the bedrock of pharmacovigilance. Data is considered to be structured when it is available in an organized or standardized format, thus permitting straightforward interpretation or analysis. This includes categorical or other qualitative data, as well as quantitative data that may be amenable to statistical analysis. The most important type of clinical data from a drug safety perspective is likely generated from randomized clinical trials which are designed to provide a robust assessment of the drug in development, as well as to highlight potential adverse effects.

Robust safety data in clinical trials is based on a quantification of defined adverse effects in exposure terms. Safety data are required to be collected using standardized case report forms. These provide a structured approach to collecting information of relevance for safety assessment and monitor the occurrence of events of interest. This can be complementary to human inspection of unstructured data in cases of suspicious information. Routine collection of patient information in clinical settings has also provided sources of safety information. The most notable of these are electronic health records that contain substantial and varied information about the person and their disease state. The purpose of these records is to provide a patient's history as a timeline, to assist in providing care on an ongoing basis. This information can be useful for signal detection for both known and unknown adverse effects. Now, all developed countries in the world have moved to electronic health records to varying degrees. One of the challenges to accessing and using this data is that it was collected for medical practice and not for scientific study; that is, it is directly related to privacy laws around health, so use is limited based on ethical considerations as well as regulation.

In summary, structured data sources are the most comprehensive data sources in each different aspect, and are most popular for individual or non-AI-based aggregated data and summarized event assessment methods. They provide the basis for more sophisticated analytical methods, such as machine learning.

3.2. Unstructured Data from Social Media and Medical Texts

Unstructured data, traditionally, has been dismissed or deemed irrelevant, especially in pharmacovigilance, perhaps due to its vast size and often unclear value. Several sources produce high-volume unstructured data that could be pertinent to pharmacovigilance.

Perhaps of most interest is that derived from social media, where patients are increasingly reporting their side effects, drug interactions, and issues experienced with their medications. What's more, free-text medical notes written by physicians or found in medical records documented when an adverse drug reaction has been suspected are at least available. These represent a rich source of real-world, non-clinical trial induced adverse effect data, which has further potential where electronic healthcare records are more widely held.

There are limitations to the analysis of unstructured data due to its complex, large volume, and subject to human error, ambiguity, and variability. They are also likely to contain a high degree of "noise," i.e., much of the volume may be irrelevant and not in any way related to what we are looking for. The extraction of pertinent information from unstructured sources relies on the application of a suite of tasks that constitute natural language processing, including speech recognition, lexicon, and grammar/syntax tasks such as named entity recognition, lemmatization, stemming, and part-of-speech tagging, as well as tasks for interpretation such as semantic and sentiment analysis. Other challenges include contextual variation in terms of language use. For example, spelling mistakes, slang, and truncated text increase the likelihood that prevention software making such searches will miss relevant information. To the best of our knowledge, the discovery of these texts requires creative, as well as user-driven, searching parameters. Regardless of these limitations, in the field of pharmacovigilance, the application of such tools and techniques to curate social media and appropriate sections of medical information is increasingly being demonstrated as very promising; some notable examples of their successful application are illustrated in specific case studies. There is also potential for social media to be used to provide near real-time reporting of suspected adverse drug reactions as a complementary resource to traditional methods.

There is also significant potential to create powerfully rich, comprehensive, and nuanced pharmacovigilance systems by combining both unstructured and structured data into one framework or database. As average-based methods ensure big data is not overwhelmed by noise, significant efforts have introduced complex machine and deep learning approaches addressing the issue of unclear data and medical literature availability by incorporating doctors' and system professionals' knowledge and capabilities to provide accurate information for further analysis. The handling of

unstructured data is further reviewed in many scientific publications. Unstructured data analytics for pharmacovigilance represents a novel and potentially extremely valuable frontier and one worthy of significant further research.

4. Applications of AI in Pharmacovigilance

Numerous AI-based methodologies and strategies have been developed and analyzed for different specific tasks and applications to improve drug safety and to support pharmacovigilance activities. AI is able to support pharmacovigilance experts in the real-time identification of adverse drug reactions (ADRs) and conveys various aspects, such as early detection of signals of disproportionate reporting through multivariate pattern mining, to pave the way for novel techniques able to proactively prevent ADRs before they occur. From a methodological perspective, traditional AI techniques and models are being deployed to process the data of SRS for the automatic identification and verification of ADRs in a spontaneous schema, as well as to analyze the data collected within EMR for discovering ADEs in observational studies. In fact, methods like machine learning on clinical notes appear promising to link a drug to an ADE in real-time, given the general desiderata to render pharmacovigilance more effective, cost-efficient, proactive, and preventable, rather than reactive and relying on reporting.

From the applications present in literature, four main topics emerge concerning the role of AI in pharmacovigilance: (1) safety issue detection; (2) signal detection and investigation; (3) risk assessment and management; and (4) ADE forecasting. The detection of safety issues (or signal generation) is notably critical to the proactive management of risk and hazards. It has been demonstrated that AI can help in the early detection of potential safety issues during all stages of drug development or in the post-authorization phase. AI and ML, in particular, are increasingly relevant in the pharmacovigilance context, as they could help predict future adverse events and permit the safety-related authorities to invest their efforts in the signal evaluation of drugs with the most serious and/or frequent adverse outcomes for the patients. Controls are in place to monitor the quality and safety of the product.

4.1. Early Detection of Adverse Drug Reactions

Adverse drug reactions (ADRs) are one of the leading causes of morbidity and mortality across the world. Thus, the capability to detect, monitor, and respond to these events will have a great impact on patients. Prompt ADR detection is useful to increase the

knowledge of safety and effectiveness of a product after marketing. Moreover, the identification of ADRs in the early phase of a product's life cycle is also essential for ensuring patients' well-being and treatment success. Although ADRs can already be detected during the clinical trials necessary to obtain their marketing authorization, these actions might not capture all of the associated risks due to factors such as patients with coexisting conditions not entering clinical trials. AI can analyze extant data in electronic health records, spontaneous literature, and social media platforms, leading to the detection of safety signals more rapidly and with greater reliability than was formerly feasible. Model developed: Case of the predicted object based on methods and tools. Real-time monitoring integrated framework for the prediction of ADRs using the database, artificial intelligence, and deep learning. Development of an integrated information real-time reactions monitoring framework for the detection, monitoring, and evaluation of ADRs in real-time, pending, and data. The system has shown a reduction in ADRs by 50% from August 2017 to January 2018, and the results from the system have been referred to healthcare professionals in three conferences on big data analysis. Case study on the prediction of chemotherapeutic-related ADRs: system based on deep learning model and artificial neural networks, including transfer-learning methodology. A database-based algorithm for early detection of ADRs and ADEs: machine learning predictor. Ensemble of classifiers for the semi-automatic procedure based on algorithms. AI algorithms, including deep learning and transfer-learning, combined with a machine learning model analysis, enable efficient and real-time prediction of ADRs. Resources in both context can be integrated into healthcare systems.

4.2. Signal Detection and Prioritization

From the safety point of view, signal detection involves identifying potential safety issues. The process of recognizing these signals from vast amounts of data is of great importance in pharmacovigilance. At the same time, this concerns adverse drug reactions identified in the pre-marketing phase as well, the so-called pharmacovigilance informally known as post-marketing surveillance. With the current state of knowledge, it is not possible to estimate whether the signal detection process is completed since the reasonable size of the available data sets significantly overwhelms the analytical capacity of pharmacovigilance systems. Thus, signal detection capabilities are adopted within the widely accepted definition of pharmacovigilance, which contains the complete cycle of pharmacovigilance activities.

Signal detection can be summarized as follows: the regulatory requirements regarding the expedited reporting in ICSR format bring a large amount of reports that contain the same or very similar suspected ADRs. The essence of signal detection is to filter in an efficient way the relevant signals from all incoming reports to identify the new or newly emerging ADRs that have actually increased in the given reporting period. This involves comparing the reports received within a specific timeframe with previous reports available related to the drug to understand whether a relevant quantitative increase in previously generated signals was experienced. The concepts of signal and alert are used to capture this aspect.

Methodologically, it is important to ensure that the entire pool of incoming reports and historical ADR ones is analyzed appropriately. Consequently, the significant action following signal detection is to prioritize signals based on their potential influence in terms of likelihood and severity to understand the potential impact on public health. Early development and successful deployment of data mining and machine learning algorithms may support signal prioritization. The first use of the unsupervised learning method for signal generation addressed the identification of new ADR signals using data mining methodology of cluster and frequency analysis. Today, it is quite common to employ machine learning algorithms to increase the reliability of drug event prioritization. For example, it has been suggested that patient-reported outcomes may predict adverse drug reactions via neural network-based modeling, and the Random Forest algorithm performed best in predicting the causality of drug-exacerbation reported liver injury.

In summary, it appears clear that currently the most successful uses of ML/AI in pharmacovigilance are in signal detection, prediction, and prioritization. However, it is also clear that these procedures are not infallible. It was reported that active surveillance of data sources used to identify potential ADRs and to build and update models are most optimal. The reason is that AI and ML are data-reliant processes, and if data is limited or biased, then the results are likely to be equally limited or biased. There are numerous early warning systems for drug safety in large databases, but there are limitations, such as periodic only reporting, late reporting, and only reporting following hospital visits presenting as false negatives; the fact that such systems can miss between 17 and 63 percent of all events should be a cause of concern. Signaling is not new and is

common sense, and after all, there is no substitution for medical expertise. Model development and testing in ICU may lead to the improved capacity to detect adverse events earlier using EHRs via the use of variable calculation and analysis algorithms. There is a pressing and urgent need for these processes to be led by clinicians so that events must be reported and acted upon. There are ongoing efforts to overcome this, such as the recent project which seeks to deliver a single, consolidated, and standards-driven EHR for medical research.

4.3. Risk Assessment and Management

One of the major concerns in the drug development process is to forecast the potential risks associated with the usage of new candidate drugs. AI methods are exceptionally valuable in risk assessment because they can deduce indispensable insights from large-scale analyses of diverse quantitative and qualitative data. This helps to create a comprehensive overview of the issues and concerns linked with a specific drug. Furthermore, AI-driven risk assessment is designed to be integrated with daily clinical practices, leading to improved patient care through better healthcare professional decision-making for patients, due to more comprehensive information that indicates the level of risk of drug usage. In pharmacovigilance, risk is usually evaluated using different methodologies, including stratification based on the patient and drug involved, as well as extrapolation from base rates with or without statistical modeling. Ideally, any risk assessment should focus on the convergence of these methodologies as stand-alone models, although this is not always feasible.

A number of real-world investigations have demonstrated AI capabilities to be useful for adverse event identification and risk management. For instance, a study reported a risk signal of bulls' hepatitis that was associated with one of the participant groups using an AI algorithm, and led to appropriate medical interventions. The implementation of various AI-driven models in patient recruitment assessment is also on the rise, especially in real-world data utilization, by evaluating the risks and potential benefits of certain drug combinations.

In spite of the adoption of AI in changing the way drug adverse reactions are monitored and managed, it is important to take note of various risks and requirements, especially from a legal and ethical point of view. For example, the importance of the privacy of individuals' data that is necessary in the process of indirectly assessing individual risks

must be considered. Furthermore, it must be noted that the regulatory framework needs to address and assess the development of AI in this field. Pharmaceutical products will require continuous monitoring as a tailored approach that adapts to new adverse event occurrence data.

5. Challenges and Future Directions

In real-world settings, the application of AI algorithms is faced with data quality issues, such as missing values, incorrect reporting of ADRs, and biases that regulate internal structures in AI decision analysis. Integrating existing systems in pharmacovigilance with the new algorithms revealed partial knowledge of the historical evolution and performance of the implanted device, which makes the decision-making process unreliable. One of AI's known Achilles' heels is represented by the quality of the data being used. The issue of inaccurate data at the input of algorithms impairs decision-making processes, potentially leading to dangerous default errors. Limiting bias from data remains a primary goal. The need for adequate data governance frameworks aims to overcome these issues in the progress of any AI initiatives. Preprocessing tools currently available offer handling strategies useful in minimizing bias during data acquisition, data exploration, and knowledge training in real-world data projects, but are challenging, often requiring in-depth knowledge of the underlying mechanisms causing bias and potential risks. The present technological first line of defense continues to be formal regulation, which encompasses both infrastructural and professional standardization. The alternatives proposed by the data market often include proposed de-biasing methodologies or domain adaptation techniques among the broad arsenal of contemporary AI fields of study. Moreover, a number of grey areas that are not clear from a regulatory point of view; a document based on a public consultation exercise elaborated on the strategic plan proposed for future AI in healthcare. The bridging of pharmacovigilance with real-world evidence and precision medicine can be considered the future developed scenario. Active players from AI, open innovation, and the pharmaceutical industry should target the following research directions: (1) developing tools for de novo pharmacovigilance strategies requirements, specific to precision medicine and biotechnological drugs; (2) optimizing modern operational pharmacovigilance to support real-world evidence generation and risk management, specifically focusing on benefit-risk assessments for high-impact, low-incidence ADRs that have major cost and humanistic implications and are reported in the system in a

very fragmented manner; (3) global collaborative networks and pre-competitive research platforms for image-based digital pharmacovigilance. Data-driven technology solutions alone can rarely transform a shortage of infrastructure. Further innovation and research directions can be aimed at these areas.

5.1. Data Quality and Bias Issues

In order to achieve effective pharmacovigilance through AI, it is essential to address the issue of data quality. Low-quality data may provide inaccuracies and a distorted view of drug safety, rendering any AI-powered pharmacovigilance effort unreliable. Issues of bias in pharmacovigilance and adverse event reporting are also some of the main hurdles that need to be taken into account before truly harnessing the power of big data and artificial intelligence to improve healthcare. There are wide arrays of intrinsic biases in real-world data, such as disparate access to healthcare facilities, diversity in payment status for medication acquisition, and differences in social and behavioral factors that pose challenges when using AI to establish reporting trends and anticipate adverse event occurrence.

The reported bias in drug prescription based on gender, race, age, and different social, economic, and health-related factors complicates the inferences that can be drawn from these variations. Thus, one must verify that the population studied is similar to the general population of patients, the indications and doses used are typical, and the patients are not unusually healthy or unhealthy. Analyses of large databases permit adjustments beyond what can be done in a randomized controlled study but suffer from the biases described above. Randomized controlled studies present a 'malignant' bias – demographic factors, e.g., income and social status, not equally balanced, as well as differences in access to healthcare and nutrition. Ultimately, all these biases should be taken into account when using AI within the pharmacovigilance realm. In conclusion, any new AI-powered solution must place a strong emphasis on the validation of data source quality. Comprehensive data auditing and validation, as well as signature-assisted tools to identify and eliminate biases, should be welcome add-ons. It is also essential to continuously monitor for potential bias points throughout the entire data lifecycle. In doing so, adopting a clear strategy to control data quality at early stages is essential.

5.2. Regulatory Compliance and Ethical Considerations

The development of AI technologies is occurring simultaneously with an evolving regulatory landscape for pharmacovigilance activities worldwide. A trend has been the increasing incorporation of "proactive" safety assessments, with the underlying principle being a need to explicitly "demonstrate" drug safety. This demand for transparency and the globalization of the pharmaceutical market gives pharmacovigilance a marked global dimension. Implementing, adapting, and maintaining an effective system continue to be primary challenges for the industry in general, and the uptake of innovation, from any technological domain, lies within the capacity for compliance with worldwide regulatory requirements. Still, concerning AI use in safety, professional judgment provides an essential means to appraise the ethical conditions within the specific context in which the research is to be conducted. This includes, in particular, verifying the nature of the data to be used in relation to the assurance of valid informed consent, the guarantee of data anonymity, and the obligations to assure the continued privacy of personal medical data, including the length of time any data will be retained.

AI-based pharmacovigilance safety generates decisions in response to regulatory concern for bias and fairness in the AI decision-making process, particularly where each treatment-related decision will affect large numbers of people. The decisions or recommendations presented have important societal implications and must be shown to be both fair to all individuals and non-discriminatory of different nationalities. Bias in drug safety decisions carries risks of varying degrees and must thus be anticipated. These, however, are offset by the well-publicized rewards that drive AI in biopharmaceutics. At present, and until the AI solutions are recognized as a reliable strong suit to comply with developing global regulatory requirements, pharmacovigilance leaders and researchers perceive the development of an AI-enhanced pharmacovigilance system as prima facie beneficial if these concerns are counterbalanced by the appropriate conduct. With these resources and developments still relatively embryonic, the ethical and societal implications of doing the research are inherent. Compliance is crucial for regulatory approval of technology and access to intended users for faster and better clinical decision-making, as well as to establish values of compliance with the following internationally acceptable ethical conditions. It is also appreciated that essential public trust is earned by ensuring data integrity and

analytical reproducibility achieved by adherence to solid methods for traceability, rigor, and bias.

5.3. Integration with Real-World Evidence and Precision Medicine

Real-world evidence from electronic health records and spontaneous adverse event reports in pharmacovigilance provides a valuable database of patient experiences to enable the evaluation of drug safety in real-life clinical practices. These observations complement pre-authorization clinical trials, which may exclude certain patient demographics and concomitant medications and capture only a fraction of a drug's safety profile. Incorporating clinical RWE to genuinely reflect real-world clinical practice enhancement, hence, ethnic background and lifestyle provides a more accurate and complete safety profile that can be used to improve patient care. In addition, pharmacovigilance initiatives are also shifting toward personalized treatment strategies, where the application of a one-size-fits-all medication is no longer suitable. Instead, a precision medicine approach aims to tailor drug treatment regimens to an individual's unique genomic, biochemical, and physiological characteristics. From a pharmacovigilance perspective, these concepts can be integrated into the modeling of AI and data-driven signal detection methodologies. Personalized medicine research has provided much clearer insight into the inter-individual differences that affect drug response, thereby demonstrating the urgent need to account for this variation in potential drug safety studies and the subsequent post-authorization re-evaluation of the risk/benefit profile. To make pharmacovigilance processes beneficial to personalized medicine, integrating these concepts into AI and traditional safety assessment methods should be a topic of interest and a priority. This integration may not only assist with the effective assessment of drug safety profiles and risk factors but also guide appropriate risk–benefit decisions and mitigation strategies. In conclusion, the close collaboration of AI developers, healthcare providers, and regulators offers an effective strategy to ensure that proper pharmacovigilance frameworks and risk identification methodologies are in place to realize their integration with real-world patient safety monitoring and precision medicine goals.

6. Future Direction

The future directions of pharmacovigilance are expected to witness an increasing integration of the latest artificial intelligence methods. A number of trends and

innovations can be expected to shape the pharmacovigilance of tomorrow. Traditional pharmacovigilance methods face challenges regarding the quality of real-world data and the lack of national licenses for drugs. A revival of the regulation is supposed to pave the way for national pharmacovigilance provisions. If the advancements in artificial intelligence live up to their promise over the next few years, then drug knowledge will see an increasingly rapid multiplication, meaning that AI may have a remarkable effect on the pharmacovigilance of the future. Regulatory authorities need to adapt their systems to suit primarily the state of the art in medical artificial intelligence, not just in terms of their technical implementation but also with regard to their auditing capabilities.

AI solution providers should employ open technologies and standards of portable results that suit various use cases and needs across diverse end-user environments. Technologies might include audits throughout the entire data lifecycle. In addition, more advanced data models must be developed, as the successful simulation and prediction of rare adverse drug reactions require very large data volumes. As research continues into how real-world data may be best harnessed, greater emphasis should be placed on the regular auditing of AI technologies' predictions by external authorities that do not commercialize AI technology. The leveraging of a broad spectrum of relevant data sources will be essential. Patients may greatly benefit from AI, as it might yield new, tailored therapies, engaging a wider, deeper understanding of diseases, to pave the way to far more effective drug approval procedures. However, pharmacovigilance's potential can be achieved only through unstoppable vigorous research on ethical and safety issues with potential AI solutions. Fostering collaboration between the responsible stakeholders, such as the pharmaceutical industry, healthcare providers, patients, providers of health wearables, AI developers, software and technology industries, and regulatory bodies, is of critical importance. It is important to research the advantages, risks, and ethical considerations of drug safety systems based on artificial intelligence. We can expect an even greater increase in investment in AI-based technologies for the identification and handling of potential side effects. Overall, research into and practical use of these tools is likely to continue growing in the future. As the results of validation studies grow over time, so too will the awareness of the pros and cons of these methods and practices grow. This will add solid information on the positive hybrids offered. A strong educational effort is important for the users of these

technologies. The effort to improve the visibility and acceptance of AI tools by healthcare professionals and other actors in the pharmaceutical field is of vital importance; the minimization of fear of the unknown that plagued the introduction of computerized tools in the past is necessary. The challenge will be to ensure that innovative technologies are exploited to their full potential and that recent innovations do not stumble on the methodological limits of traditional drug identification.

7. Conclusion

Over the years, several pharmacovigilance systems have been established across the world to monitor and control drug safety. Such a robust pharmacovigilance system can ensure the value of provided health care to the community, i.e., enhanced patient care and patient safety. This is a need of the time because adverse drug reactions are one of the leading causes of death across the world. In addition, a robust pharmacovigilance system can deliver empirical knowledge based on vast amounts of data collected on various exposed populations. This empirical drug safety information is crucial since it helps regulatory authorities, healthcare professionals, and patients make the right decisions regarding public health. Timely and accurate information from ongoing mechanisms ultimately supports periodic regulatory reports on drug safety, which are submitted by different pharmaceutical companies to regulatory authorities to document the overall benefit-risk profile of marketed medicine.

AI-based solutions have revolutionized different fields of health sciences. They are certainly going to bring many positive changes in pharmacovigilance practices. One of the segments on which these AI solutions may bring an impact is to find and extract maximum knowledge from a combination of structured data sources and unstructured data to provide a comprehensive view of causality, i.e., not only a rough incidence of particular unwanted medicine effects but to extend the boundaries for fundamental drug safety science. Pharmacovigilance practices have to adopt such methodologies within their spontaneous reporting mechanism to unleash the true potential of the historical and current safety signals, significantly reducing the subjectivity of the causality assessment. On the AI and ML platforms, different heterogeneous data sources can be interrogated to gain accurate and much improved responses related to medicine safety. Such responses can certainly be used to improve and streamline the current workflows. Regardless, evaluations of AI systems on their classification criteria, error

estimations, and generalization capabilities are required for regulatory acceptance and for the personalized medicine paradigm on which we move. AI/ML methods should be developed and applied based on patient characteristics. Pharmacovigilance and AI healthcare stakeholders should have multiple areas of collaboration within multi-disciplinary partnerships to realize the benefits of big data in pharmacovigilance practices. Furthermore, pharmacovigilance should continue to evolve, given the steady progress in AI/ML. New drug safety AI/ML solutions should improve the efficacy and effectiveness of the entire pharmacovigilance continuum, but there is still further necessary research and methodological pathways to innovation.