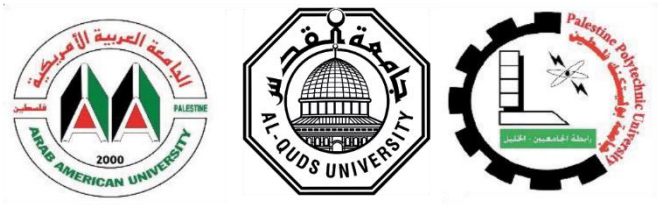
Deanship of Graduate Studies

Al- Quds University



Modeling Adverse Events Detection as Time Series Using LSTM and Labeled Data Fusion

Rasha Zaki Mustafa Assaf

PhD Dissertation

Jerusalem - Palestine

1446 H/ 2025

Modeling Adverse Events Detection as Time Series Using Lstm and Labeled Data Fusion

Prepared by:

Rasha Zaki Mustafa Assaf

Supervisor: Dr. /Prof.: Amjad Rattrout

Co-Supervisor: Dr: Rashid Jaiousy Dr:Mohammed Khalilih

A thesis Submitted in Partial fulfillment of the requirements for the degree of PhD in EIT Requirements for the PhD In Engineering of Information Technology Joint PhD Program- AAUP, PPU and QU Deanship of Scientific Research and Graduate Studies

Jerusalem – Palestine

1446 H / 2025

Al- Quds University
Deanship of Graduate Studies
PhD In Engineering of Information Technology

PhD Approval

Modeling Adverse Events Detection as Time Series Using Lstm and Labeled Data Fusion

By: Rasha Zaki Mustafa Assaf
Registration No: 21912560

Supervisor: Dr. /Prof.: Amjad Rattrout

Co-Supervisor: Dr: Rashid Jaiousy Dr:Mohammed Khalilih

PhD thesis submitted and accepted, Date: 27/05/2025

The names and signature of the examining committee members are as follow:

Head of Committee Dr. /Prof.: Amjad Rattrout Signature……..

Internal of ExaminerNael Abu Halaweh Signature……..

External Examiner Rachida Fissoun Signature ……..

External Examiner Khaled ebn abd Al-Salam Signature ……..

Dedication

I dedicate this thesis to the cherished memory of my beloved parents, Aisha and Zaki.

To my dear mother, Aisha, whose unconditional love, endless encouragement, and profound wisdom shaped who I am. Though she is no longer with us, her spirit continues to guide and inspire me every day. She instilled in me the values of perseverance, determination, and a deep respect for knowledge. Her unwavering belief in me, even when I doubted myself, was the light that carried me through the darkest moments.

To my beloved father, Zaki, whose strength, generosity, and quiet sacrifices formed the foundation of my journey. Though he too has passed on, his presence remains alive in my heart. His dedication, patience, and support—often shown without words—have left a lasting mark on my path. He taught me the value of hard work, humility, and integrity.

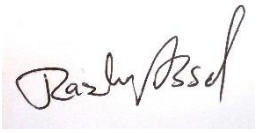
Although they are no longer here to witness this achievement, I feel their presence in every step of this journey. I carry with me their dreams, their lessons, and their love. This thesis stands as a testament to their enduring influence and the sacrifices they made so that I could reach this milestone.

With eternal love, gratitude, and reverence, Dedicated in loving memory to Aisha and Zaki..

Rasha Zaki Mustafa Assaf

Deliration:

I certify that this thesis submitted for the degree of PhD, is the result of my own research, except where acknowledged, and that this study (or any part of the same) has not been submitted for a higher degree to any other university or institution.



Rasha Zaki Mustafa Assaf

Date: 27/05/2025

Acknowledgments

I want to take a moment to express my deepest gratitude to my family. Your unwavering support, love, and encouragement have been my strength through every challenge and success. Thank you for always believing in me, for your patience, and for the sacrifices you have made to help me reach this point. I am truly blessed to have you by my side.

To my teachers, I am incredibly grateful for your guidance and wisdom. Your dedication and passion for teaching have inspired me in ways I cannot fully express. Thank you for pushing me to do my best, for sharing your knowledge, and for shaping my journey with your invaluable lessons.

This journey has been possible because of all of you, and I am forever thankful.

Rasha Zaki Mustafa Assaf

Abstract

Adverse events (AEs) , characterized as undesirable and unintentional outcomes of medical treatments, pose significant challenges in healthcare. Accurate prediction of these events is crucial for enhancing patient safety, optimizing resource allocation, and improving overall health- care outcomes. Traditional methods, such as statistical analyses and early machine learning techniques, often fail to capture complex, nonlinear relationships of medical data. This dissertation explores the application of advanced machine learning models, particularly Long Short- Term Memory (LSTM) and transformers, to predict adverse events and outcome using Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS), a database containing information on drug safety, and Medical Entities in Digital Records Annotated (Med- DRA) datasets. The dissertation focuses on integrating these comprehensive datasets, followed by data cleaning and pre-processing to ensure data accuracy and reliability. Latent Dirichlet Al- location(LDA) and clustering, a statistical method for topic modeling was used to address the complexity of classifying over 50,000 adverse event categories, reducing the number of classes to 10. The study further investigates the efficacy of sequence-to-sequence models, such as the Transformer, in predicting adverse events. Results indicate that Transformer-based models out- perform traditional machine learning algorithms in predicting adverse events, demonstrating improved accuracy and robustness. The sequence-to-sequence models achieved an accuracy of 92.81%, with a precision of 93.22%, a recall of 91.89%, and an F1 score of 92.55%. The integration of self-attention mechanisms enhances these models by allowing them to capture complex patterns and relationships within the data—insights that traditional approaches often overlook s. We implemented a Graph Neural Network (GNN) and evaluated precision and re- call across multiple categories, yielding low overall F1 scores of 0.36 for term type classification and 0.42 for FAERS reaction classification.

Keywords: seq2seq, Adverse Drug Events, FAERS, LSTM, MedDRA, Graph Convolutional Networks

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADR Adverse Drug Reaction AE Adverse Event

ANN Artificial Neural Network

BERT Bidirectional Encoder Representations from Transformers CBOW Continuous Bag of Words

CSV Comma-Separated Values EHR Electronic Health Record

FAERS Adverse Event Reporting System GCNs Graph Convolutional Networks GNN Graph Neural Network

GRU Gated Recurrent Unit

HLGT High-Level Group Term (in MedDRA hierarchy) HLT High-Level Term (in MedDRA hierarchy)

ICD International Classification of Diseases LDA Latent Dirichlet Allocation

LLT Lowest Level Terms

LOINC Logical Observation Identifiers Names and Codes LR Logistic Regression

LSTM Long Short-Term Memory

MedDRA Medical Dictionary for Regulatory Activities MLMachine Learning

NDC National Drug Code

PT Preferred Term (in MedDRA hierarchy) RF Random Forest

Seq2Seq sequence-to-sequence

t-SNE t-distributed Stochastic Neighbor Embedding XML eXtensible Markup Language

FDA Food and Drug Administration

Chapter One
Introduction

1.1 Introduction

In this chapter, we explore the prevalence and impact of adverse drug reaction ADRs and ad- verse events (AEs) within healthcare systems. We begin by defining AEs and ADRs [1][2], distinguishing between preventable and unpreventable cases, and examining their consequences on patient safety and healthcare trust. We then review multiple studies conducted in Middle Eastern countries, highlighting statistical findings, causes of AEs, and the role of healthcare professionals in mitigating these risks. A key discussion revolves around pharmacovigilance efforts across the Arab world, analyzing the extent of implementation, challenges, and gaps in safety reporting mechanisms. The chapter further investigates patient safety culture in hospitals, emphasizing how reporting transparency and non-punitive responses influence AE management. Finally, we discuss emerging strategies for reducing AEs, including the adoption of artificial intelligence, machine learning, and Predicting and avoiding possible health issues by means of electronic health records . These insights provide an all-encompassing understanding of the existing healthcare challenges and potential solutions in minimizing adverse events.

1.2 Problem Statement

Despite the tremendous scientific and technological development within the medical field, including drug discovery for new diseases, the carry out of complex and surgical operations, many medical errors are still being committed, putting patients in danger and sometimes causing death. Adverse events are categorized as non-serious or serious; unexpected or expected. AEs that occur during treatment may include drug side effects, injury, psychological

injury, dis- ability or permanent damage, congenital abnormality/birth defect, trauma, or death. AEs could be the result of a drug-to-drug interaction or the patient receiving incorrect drugs for his/her medical condition. Table 1.2 shows a few AE categories of AEs. AEs will decrease confidence in the healthcare system from patients and their families because of the suffering it inflicts on them.

Table 1.1: Adverse Events Examples

Potential AE	**Unpreventable AE**	**Preventable AE**
Adverse events, such as hospital-acquired/nosocomial infectionthat could have been avoided.	The harm that cannot be avoided because it happens as a result of an unforeseeable complication.	A patient suffers damage as a result of a preventable human or system mistake.

The prediction of AEs in healthcare is a critical challenge due to the complexity and variability of medical data. Traditional statistical methods and early machine learning techniques have limitations in capturing the non-linear and complex correlations inside huge datasets [3, 4, 6, 7]. These methods often fail to provide accurate predictions, leading to missed opportunities for preemptive interventions that could enhance patient safety and optimize healthcare resources. The emergence of advanced machine learning models, particularly Transformers, offers a promising solution [5]. However, the effective application of these implemented models in predicting adverse events using comprehensive datasets like FAERS and MedDRA remains underexplored [8]. This study tries to fill these gaps by developing and validating machine learning models for adverse event prediction, thereby improving prediction accuracy and healthcare outcomes. Table 1.2 contains the list of notations and symbols that will be used throughout this document.

Table 1.2-A: List of Notions/Symbols

Notion/Symbol	Description
M	Ordered sequence of medications prescribed for a patient.
R	Set of adverse event reactions.
m_i	Medication data given at time step i.
d	Drug name.

Table 1.2-B: List of Notions/Symbols	
Notion/Symbol	Description
di	Standardized drug portion/dosage.
r_i	Standardized drug route.
r_j	Standardized medication RC.
r_k	The k-th adverse event reaction.

For the scope of this work, we define the input M as the ordered list of medication data m_i, where m_i represents the prescribed medication for the patient at the i-th step.
Formally, the input is defined as:

$$M = m_0, m_1, m_2, \ldots, m_k$$

The term m_k refers to the medication data at the k-th step or time point in the sequence of medications. where:

$$m_i = (d, di, r_i, r_j)$$

The goal is to predict the set of reactions R.The term r_n refers to the n-th adverse event reaction in the set R:

$$R = \{r_1, r_2, r_3, \ldots, r_n\}$$

where the subscript does not impose an ordinal relation between reactions within the same set. This proposed formulation renders the problem as a variation of a classification model, where the input is the sequence of medications M and the output is a set of reactions R.
In our dataset, there are over 50,000 unique reactions with an average of 20 reactions per case. where each r corresponds to a potential adverse event, mapped to a standardized Med-DRA ontology code. One should underline the fact that R's components are not temporally or positionally ordered. Even though the predicted adverse reactions do not follow a specific order, we use a sequence-to-sequence (Seq2Seq) model because the input -the order in which drugs are given - plays a crucial role in determining which side effects might -AE occur. The model can capture intricate temporal correlations inside the pharmaceutical sequence and estimate the prediction of reaction sets during decoding by using a Seq2Seq framework based on LSTM networks.
The decoder employs a multi-label classification technique [106], using a sigmoid activation over the expected ADEs, in order to conform to the unordered nature of the output. This

enables the model to use the decoder's sequential modeling capabilities during training, but instead of producing a tight sequence, it can provide a set of reactions.
This formulation supports the creation of predictive tools for pharmacovigilance and individualized treatment planning by allowing the application of potent neural architectures to a therapeutically useful task. .

1.3 Challenges to adverse event prediction in the middle east

Adverse events present a challenge for healthcare systems worldwide and are particularly underaddressed in the Palestinian healthcare system. Every day in Palestine we witness injuries as a result of adverse event reactions. There are many examples of medical errors that one can encounter on a daily basis such as "cholecystectomy wound that did not close" and "a woman died of blood-sucking cotton pads in her body during childbirth" . AEs and ADRs are being studied in Palestine from a legal point of view in terms of who is legally responsible, and material compensation [9][10]. However, there are no mechanisms in place to prevent this from happening. Research on adverse events in the Middle East is challenging. Yet, it is very critical especially as the number of AEs and ADRs is on the rise and there is a lack of accountability. Designing and implementing machine learning-based adverse event prediction models in the over-world [11, 12, 15, 14, 13] , Middle East in particular is faced with multiple challenges, some of which are listed below.

- Not all patient data is expected to be recorded in Electronic Health Records (EHR).
- Adverse events are not part of the EHR system.
- Lack of standardized approaches for AE documentation.
- Errors in coding patient information.
- Missing diagnosis or medication information.
- Uncoded data, including the International Classification of Diseases (ICD), National Drug Code (NDC), and Logical Observation Identifiers Names and Codes (LOINC).
- Data Accessibility Due to Privacy and Guidelines.

- Lack of Adverse Event Reporting Systems.
- Availability of labeled adverse events data.
- System and culture.

1.4 Objectives

The primary objectives of this research are:

1. **Integration of datasets:** To integrate the FAERS and MedDRA datasets to create a comprehensive database for adverse event prediction.
2. **Data preprocessing and cleaning:** To conduct meticulous data cleaning and preprocessing to ensure the accuracy and reliability of the integrated dataset.
3. **Class reduction with LDA:** To employ Latent Dirichlet Allocation (LDA) for reducing the number of adverse event categories, thereby simplifying the classification task and improving model performance.
4. **Model development and evaluation:** To develop and train transformer based sequence-to-sequence models aimed at predicting adverse events and to assess model performance using accuracy, precision, recall, and F1 score.
5. **Benchmarking with traditional methods:** To compare the performance of Transformer-based models with traditional machine learning algorithms in predicting adverse events.

1.5 Significance of the Study

This study proposes an innovative approach to adverse event prediction through the integration of extensive datasets, specifically FAERS and MedDRA. These datasets provide rich, complementary information: FAERS offers real-world, time-stamped reports of drug-related adverse events, while MedDRA provides standardized terminology for medical events, enabling precise labeling and classification. Leveraging these sources allows for a more structured and comprehensive representation of drug-event relationships.

At the core of this approach is the drug sequence encoder, a transformer-based model structured to capture detailed relationships within drug exposure sequences. By modeling the sequential structure of drug administration, the drug sequence encoder can account for complex dependencies that are often involved in adverse drug reactions. Unlike traditional

models, which have limitations in handling sequential interactions, this method enhances the accuracy and depth of adverse event predictions [16, 17, 18, 19, 20]. The outcomes of this study hold potential to make a meaningful impact on healthcare analytics by refining the predictive accuracy of adverse drug events. Such advancements could ultimately support clinical decision-making processes and improve patient safety [21, 22].

1.6 Background

Medical errors remain a serious threat to patient safety even with the field's remarkable accomplishments, which include anything from advanced surgical methods to new drug discoveries. When an adverse event happens, the consequences can be extensive. Not only can the patient experience increased physical or psychological suffering, but they may also face a longer hospital stay, additional medical treatments, permanent disability or in the worst cases, death. The persistence of medical errors in spite of international attempts to address and reduce them high- lights the difficulty in providing healthcare and the necessity of constant attention to patient care procedures. AEs represent a crucial aspect of this discussion, as they encompass a wide range of negative outcomes that can occur as a result of medical treatment. AEs can be classified based on their severity and whether they were expected or unexpected [23]. Non-serious AEs may include mild side effects from medications, while serious AEs can involve severe complications such as permanent disability, psychological harm or even death. The occurrence of these events can be particularly alarming, as they often arise during routine medical interventions, whether surgical or pharmacological.

A significant number of AEs are preventable, often resulting from lapses in protocols, in- adequate communication among healthcare providers or failure to adhere to established safety guidelines. For example, medication errors can occur when a patient is prescribed a drug that interacts negatively with their current medications or when incorrect dosages are administered. Such mistakes not only harm the patient but can also lead to broader implications for health- care delivery, including increased costs and resource utilization. Conversely, some AEs are classified as unpreventable, occurring due to unforeseen complications or rare occurrences that cannot be anticipated. Even in these cases, the impact on patient safety remains substantial, and healthcare professionals must remain vigilant to manage these risks effectively. The implications of adverse events extend deeply into the

healthcare system, as they can compromise the trust patients place in medical professionals and institutions. When patients or their families experience harm as a result of medical care, it breeds skepticism and anxiety regarding future treatments, ultimately affecting health outcomes. Healthcare providers must establish a culture of safety and transparency in order to address these concerns effectively.

Few studies were conducted into adverse events across the Middle East, in [24], the authors investigated adverse events occurring in Jordanian hospitals. by sending questionnaires to expert nurses who have good experience in working in public and private Jordanian hospitals. The outcome of the data collection questionnaire included quantitative and qualitative results surrounding factors causing adverse events including 1) wrong diagnosis, 2) hospital-acquired infections, and 3) bedsores. In addition, common factors contributing to adverse events included patient falls, insufficient staffing, patient neglect, ethical shortcomings, ineffective management, high psychosocial work demands, and the lack of clear written protocols. Nurses estimated that adverse events took place in roughly 28% of hospitals.

Authors in [25] used a questionnaire to collect data from 143 cancer patients from the Qatar's National Center for Cancer Care and Research. Using statistical analysis such as mean, multi-regression, and univariate. The outcome shows 97% of patients agreed that being aware of the AEs of the treatment will help them deal with any complication more sufficiently. How- ever, care providers in the Middle East usually do not disclose such information for fear of non-compliance with treatment or to avoid accountability.

In [26] the authors shed light on the reality of pharmacovigilance in the 22 Arab countries, motivated by the maturity of pharmacovigilance in many Middle Eastern countries. According to [26], 45% of Arab countries made more progress towards pharmacovigilance, such as Jordan, Egypt. Those countries are also official full members or associate members of the World Health Organization Collaborating Center of International Drug Monitoring (WHO- UMC). While other countries like Palestine and Yemen are at an early stage of implementation and development of pharmacovigilance systems. Also, these countries are not members of the WHO-UMC program. In Palestine, about 47% of pharmacists think they need to report ADR to the drug company instead of the authority, indicating Palestinian pharmacists need in-depth training in pharmacovigilance.

In [28], the author aims to evaluate patient safety in Palestinian hospitals and introduce guidelines that guide decision-makers involved in safety improvement work. In this study,

they used the Global Trigger Tool for reviewing the retrospective patient's record for two hospitals in Palestine, one is a nonprofit an educational hospital, whereas the other is a privately operated, non-governmental hospital. They reviewed 640 random discharge patients' records by nurses and physicians from the hospital. The output shows one out of seven patients suffered from damage, 59 percent of these events were avoidable, and 70.4 percent led to a temporary outcome damage that required necessitated an extended hospital stay. The study appears to show an increase in Adverse Events about 20 times compared to previous reports.

Another study on adverse events in Palestine, referenced in [29] focused on examining the relationship between patient safety culture and the occurrence of adverse events. The data was collected over three months of discharged patient records in the period (May-August 2009). The data was used by Global Trigger Tool and 2010 for Hospital Survey of Patient Safety Culture. For their experiment, they have used the Spearman coefficient. The study found an inverse relationship between patient safety culture and the occurrence of adverse events—indicating that as the culture of patient safety improves, the incidence of adverse events decreases.

A study about patient safety in the Arab World was presented in [30]. According to the study, non-punitive responses to errors are viewed as a major problem that has to be addressed. Healthcare workers in Arab nations sometimes believe that there is still a "culture of blame" that discourages them from reporting occurrences. They observed that the reported composite score for the dimension of unit collaboration was quite similar across all of the studies they looked at. Collaboration within individual units was shown to be more effective than teamwork between different hospital departments. The average score for this attribute across all investigations was 73.2 percent, indicating that organizational learning and continuous development were good. Furthermore, the researchers discovered that communication transparency appears to be a problem for healthcare practitioners in Arab nations.

Another survey was conducted on 13 Arab countries on pharmacovigilance [31]. Data was gathered from eleven countries. Syria and Lebanon did not participate, while six countries (Iraq, Egypt, Jordan, the Kingdom of Saudi Arabia, the United Arab Emirates, and Oman) reported official national pharmacovigilance programs, whereas five countries (Kuwait, Bahrain, Palestine, Yemen, and Qatar) reported no operational designated center or program. The plurality of them is government-funded, although staffing is limited (between two and

10 employees). About 67 percent of programs allowed spontaneous ADRs to be sent to the center through email, but none did so directly via a web-based platform. All of them used the data for medication regulation reasons, and five of them said they shared safety information with the public.

To combat the issue of AEs, healthcare systems are increasingly turning to innovative strategies. Integrating technology, for instance electronic health records and data analytics enables providers to identify patterns and risks associated with patient care more effectively. Further- more, the utilization of artificial intelligence and machine learning algorithms holds promise in predicting potential adverse events by analyzing extensive datasets, allowing for proactive intervention before harm occurs. Training and education for healthcare professionals are also critical components of reducing the incidence of AEs. Ongoing professional development and a commitment to cultivating a culture of continual improvement can empower providers to more effectively recognize and respond to possible hazards in patient care. Establishing a setting that promotes transparent reporting , and learning from errors, rather than assigning blame, is essential in facilitating systemic improvements.

1.7 Theoretical Background

This section outlines the foundational theories and techniques used in this study, including various neural network architectures, optimization algorithms, and embedding methods relevant to adverse drug event (ADE) prediction.

1.7.1 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNNs) are designed to process sequential data by maintaining a hidden state that captures dependencies across time steps. However, they struggle with long- term dependencies due to the vanishing gradient problem [32].

1.7.2 Long Short-Term Memory (LSTM)

LSTM networks were developed to address the limitations of traditional RNNs by introducing memory cells and gating mechanisms that allow them to retain relevant

information over longer sequences [33].

1.7.3 Sequence-to-Sequence (Seq2Seq) Model

The Seq2Seq architecture utilizes an encoder-decoder framework, enabling it to map input sequences to output sequences effectively. It is useful in sequence modeling tasks like machine translation and ADE prediction [34, 35].

1.7.4 Transformer Architecture

Transformers leverage self-attention mechanisms to process input data in parallel, allowing them to model long-range dependencies more efficiently than RNNs [36, 37, 38].

1.7.5 Word2Vec

Word2Vec models generate word embeddings by predicting surrounding words in a context window using either Skip-gram or the Continuous Bag of Words (CBOW) approaches. These embeddings capture semantic relationships among terms [39]. In this study, Word2Vec was applied to adverse event terms to extract semantic representations.

1.7.6 RoBERTa

RoBERTa is a transformer-based language model that improves upon (BERT) by eliminating the next sentence prediction task and training with larger datasets and longer sequences. It provides rich contextual embeddings for text data [40].

1.7.7 Latent Dirichlet Allocation (LDA)

LDA is a generative probabilistic model used for topic modeling. It assumes that each document is a mixture of topics and each topic is a distribution over words [41].

1.7.8 Adam Optimizer

The Adam optimizer combines the advantages of AdaGrad and RMSProp, using adaptive learning rates and momentum to improve convergence during training [42].

1.7.9 Data Mining in Pharmacovigilance

Data mining techniques have also been successfully applied to predict drug safety using large- scale adverse event datasets [43].

1.7.10 RxNorm

RxNorm, developed by the U.S. National Library of Medicine, standardizes clinical drug names and identifiers to support interoperability across health systems—for example, it maps both the brand name *Advil* and the generic *Ibuprofen 200 mg oral tablet* to the same RxNorm concept (RxCUI 617314), enabling consistent drug representation [44].

1.8 Thematic Overview

The term "Adverse Events" is used to describe an injury that was caused by medical treatment instead of the patient's fundamental condition [45]. These events can occur in various forms, ranging from minor injuries to severe complications. For instance, a drug prescribed to treat a condition might inadvertently cause an adverse event. Such events are not necessarily due to a medical error; they can occur even when standard treatments are administered correctly. Drugs can also cause ADR, which is defined as any harmful or unintended change in the body that is suspected of being related to the usage of a drug. These reactions occur when a patient takes a drug at the prescribed dose and experiences side effects that necessitate medical intervention, a change in dosage, or caution in future treatments involving the same drug [1, 2]. ADRs can vary from mild to severe and may include symptoms like allergic reactions, organ damage, or even life-threatening conditions [46]. The burden of adverse events extends beyond the individual, placing immense pressure on healthcare systems due to the increased need for medical resources, specialized care, and additional treatments.

The worldwide prevalence of ADRs represents a significant global public health concern. ADRs, which are unintended and harmful reactions to medications taken in normal doses, rank among the primary causes of injury and mortality globally. In the United States specifically, more than two million serious ADRs occur in hospitals each year, contributing to more than 100,000 fatalities per year [47] .

These figures highlight the immense scale of the issue and its devastating impact on patient health. The occurrence of ADRs not only compromises patient safety but also places a heavy burden on healthcare systems. The financial implications of adverse events are significant. In the United States, it was projected that the cost of adverse events would reach 383.7 billion dollars by 2022 [48, 49]. These costs arise from extended hospitalizations, the need for more intensive care, legal expenses, and compensation for medical errors. As a result, adverse events contribute to the rising cost of health care. For healthcare providers and policymakers, addressing the risks associated with ADRs is crucial for reducing preventable harm and controlling the costs of medical care, which is a major concern for both patients and healthcare providers.

Adverse events may occur for many different reasons. Human error, for example mistakes made by healthcare providers, is one factor and Inadequate management systems in hospitals or clinics, where processes are not in place to ensure proper patient care, also play a part in causing of adverse events. Additionally, the absence of effective pharmacovigilance systems programs designed to track and evaluate the safety of drugs after they have been released to the market can lead to a lack of early detection of potentially harmful side effects [50]. To prevent or minimize adverse events, patient safety must always be the top priority in healthcare. Implementing robust systems to monitor treatments, enhance communication among healthcare providers, and promote patient awareness are critical steps to ensure that adverse events are minimized.

Early detection of probable ADRs during the initial phases of drug development is crucial for improving drug safety, minimizing risks to patients, and reducing costs for pharmaceutical companies. By identifying potential safety issues early in the clinical trial process, drug developers can implement safety measures, modify drug formulations, or adjust dosing recommendations to mitigate these risks. This not only ensures the safety of future patients but also helps streamline the regulatory approval process, avoiding delays and costly post-market complications. Pharmaceutical companies benefit from this proactive approach, as it helps pre- vent expensive recalls and litigation, improves drug efficacy and

fosters trust among healthcare professionals and patients.

Understanding the overall risk-benefit profile of new medicines is essential for ensuring ethical clinical research practices and enabling informed regulatory decision making. There have been several high-profile instances of drug recalls after approval by the Food and Drug Ad- ministration (FDA), highlighting the importance of vigilant safety monitoring even after a drug reaches the market [51]. For example, in January 2022, concerns emerged about the presence of N-nitrosodimethylamine (NDMA), a chemical classified as a potential human carcinogen. NDMA contamination led to the recall and withdrawal of Ranitidine, commonly known by its trade name, Zantac. This case underscores the ongoing need for pharmacovigilance—systems that monitor the safety of medications once they are available to the public—to detect and respond to safety concerns which might not have been apparent during clinical trials.

Drug recalls are not uncommon, with numerous cases documented throughout history. Between 1953 and 2014, 43 medications were globally discontinued due to safety concerns, while 462 drugs were withdrawn from the U.S. market [52]. According to FDA records, between 2015 and 2017, 113 drugs were approved, but they were later taken off the market after 2017. A notable example is Vioxx, a drug used to treat arthritis and chronic pain. The FDA approved it in 1999, but it was withdrawn from the market in 2004 due to a heightened risk of stroke and heart attack. [53]. The case of Vioxx, along with many others, illustrates the complexity of balancing the benefits of innovative treatments with their potential risks. Despite undergoing rigorous testing before approval, some adverse effects only become apparent after extended use by a larger, more diverse patient population. These incidents highlight the necessity of continuous drug safety monitoring, both during the clinical trial phases and throughout a drug's life cycle on the market. Proper pharmacovigilance systems, combined with early detection efforts, can help safeguard public health and ensure that only safe and effective medications remain available to patients. To address the growing concern of ADRs, researchers have been exploring various methods to predict ADRs before they occur, aiming to improve patient safety and drug efficacy. Early prediction of ADRs could enable healthcare providers and pharmaceutical companies to intervene and adjust treatments to minimize harmful side effects. Over the years, various techniques have been suggested to tackle this challenge including rule-based, statistical, machine learning, and hybrid approaches.

Some countries have developed comprehensive systems to monitor AEs and ADRs, which

are critical for ensuring drug safety and improving public health outcomes. In the United States, the FAERS is the primary platform for collecting data on ADRs. This system aggregates re- ports from healthcare professionals, consumers, and manufacturers, allowing the FDA to ana- lyze trends and identify potential risks associated with pharmaceuticals. Importantly, FAERS includes reports from international sources, making it a valuable resource for monitoring global drug safety. In the European Union, Eudra Vigilance serves a similar function [54]. Managed by the European Medicines Agency (EMA), it tracks suspected ADRs from all member states. Eudra Vigilance facilitates data exchange between national authorities, pharmaceutical companies, and healthcare professionals, promoting collaboration for evaluating the safety of drug issues. The system also plays a vibrant role in post marketing surveillance, helping to ensure that newly approved medications continue to meet safety standards once they are in use. On the other hand, many countries in the Middle East lag behind in establishing robust systems for monitoring ADRs. One key challenge is the limited participation of Arab countries in international drug safety networks. Only 45% of Arab countries are part of the World Health Organization's Collaborating Centre for International Drug Monitoring, commonly known as the WHO–Uppsala Monitoring Centre (WHO-UMC). The WHO-UMC is a critical component of global pharmacovigilance efforts, as it allows countries to share information about drug related risks and benefit from the collective expertise of the international community [26, 55]. Nations like as , Egypt, Morocco, and Jordan have made progress in this area by developing some form of ADR reporting systems. For example, Egypt has established a National Pharmacovigilance Center, which collects and analyzes reports of ADRs to identify potential safety issues. Similarly, Morocco and Jordan have set up national databases for tracking drug related adverse events, enabling healthcare providers to report any issues they encounter with medications. These systems, although still developing, provide a foundation for improving drug safety and patient outcomes in the region.

In contrast, Palestine remains without a formal system for monitoring ADRs, which presents significant challenges in collecting reliable data on drug safety. The absence of a structured pharmacovigilance framework makes it difficult for healthcare authorities to gather ground truth data, assess the safety of medications, or identify emerging risks. This lack of infrastructure also hampers the ability to provide accurate statistics on AEs and ADRs, making it harder to address potential safety concerns and inform healthcare policies. Furthermore, without a for- mal reporting system, it becomes more challenging to educate

healthcare professionals about the importance of reporting ADRs and to raise awareness about drug safety among the general population. The situation in Palestine highlights a broader issue across many low- and middle- income nations , where limited resources, lack of awareness, and weak regulatory frameworks hinder the development of effective pharmacovigilance systems. In such settings, the collection of ADR data often relies on ad hoc reporting by individual healthcare providers, which may result in underreporting or inconsistent data. Strengthening pharmacovigilance capacity in these countries requires not only building reporting systems but also improving healthcare infrastructure, training healthcare workers, and encouraging a culture that supports the reporting of safety incidents.

This dissertation aims to present an overview of the state of machine learning applications in adverse event (AE) prediction and investigate how these advanced solutions might be used to address AE prediction issues in developing nations. Machine learning techniques can be used to improve drug safety, boost early adverse event identification, and bridge gaps in the current healthcare infrastructure, especially in underdeveloped pharmacovigilance systems. These data driven strategies can lessen the need for human reporting systems, automate the analysis of big datasets, and provide real time insights all of which can assist overcome resource restrictions. By doing this, nations with inadequate AE monitoring systems can get access to more effective and proactive ways to safeguard patient safety and support international activities related to drug safety. The remainder of this dissertation is structured as follows: Chapter 2 provides a com- prehensive review of the relevant literature. Chapter 3 describes the dataset used in this study, including preprocessing steps and key characteristics. Chapter 4 outlines the proposed model architecture. Chapter 5 explores the application of probabilistic topic modeling for dimensionality reduction within pharmacovigilance datasets. Chapter 6 discusses the representation of drug sequences, followed by Chapter 7, which introduces the sequence-to-sequence (seq2seq) model employed in this work. Chapter 8 presents the clustering methodology and discusses the experimental results. Finally concludes the dissertation and outlines potential directions for future research.

1.9 Chapter Conclusion

This chapter has provided an in-depth examination of adverse events and their significant im- plications for healthcare systems, particularly in the Middle East. Through a review of multiple studies, we identified key contributors to AEs, such as misdiagnoses, hospital-acquired infections, inadequate staffing, and cultural barriers to error reporting. Our analysis of pharmacovigilance efforts revealed disparities in implementation across Arab countries, with some nations making substantial progress while others lag behind due to resource constraints and lack of awareness. Additionally, we highlighted the critical role of patient safety culture in mitigating AEs, demonstrating that fostering transparency and continuous learning can lead to better outcomes. The discussion on technological advancements, including AI-driven predictive models and data analytics, underscored the potential for proactive intervention in reducing adverse events. Ultimately, this chapter emphasizes the urgent need for improved safety measures, education, and policy reforms to enhance patient care and minimize preventable medical errors.

Chapter Two

Literature Review

2.1 Overview

Adverse Event (AE) detection has been a critical area of research, leveraging various methodologies ranging from rule-based approaches to advanced deep learning models. Additional statistical methods, such as logistic regression and survival analysis, have provided foundational insights for adverse event prediction. However, as noted by [27], these methods often fall short in modeling complex temporal patterns and high-dimensional data, and they struggle to capture the complexity of biological interactions. With the advent of machine learning, particularly supervised learning, models trained on large datasets have shown promise in identifying hidden patterns that traditional methods struggle to detect. However, these methods require extensive domain expertise and feature engineering, which can be labor-intensive. Recent advancements in deep learning, especially with neural networks designed for sequential data, have revolutionized ADR prediction by leveraging drug sequence information and temporal interactions. This chapter provides a comprehensive review of existing approaches, highlighting key studies that have contributed to the evolution of AE detection methodologies.

2.2 State of the Art

Researchers addressed Adverse Event detection using several approaches including rule-based, statistical tools, machine learning and hybrid approaches. Rule-based approaches typically rely on predefined rules or criteria derived from clinical knowledge to identify potential ADRs. While straightforward and easy to interpret, these methods often lack flexibility and are limited by the quality of the predefined rules .These methods can be useful for identifying broad trends but may struggle with the complexity of biological

systems, which are essential for understanding the intricate interplay between drugs and human physiology as well as accounting for individual variability in drug responses. Machine learning techniques particularly super- vised learning have gained popularity in predicting ADRs by training models on large datasets containing information about drugs, patients, and outcomes. These models are capable of identifying patterns and relationships that are difficult for traditional statistical methods to detect. However, many machine learning approaches rely significantly on domain expertise and feature engineering—the process of selecting and transforming raw data into meaningful representations that facilitate effective model learning. This process can be time consuming and labor intensive, requiring a deep understanding of both the medical and technical aspects of ADR prediction. In recent years, deep learning methods have achieved remarkable progress across various domains, including natural language processing, computer vision, and sequence modeling.

Deep learning techniques [56, 57, 58], which use neural networks with multiple layers, have the ability to automatically learn complex patterns from data without requiring extensive feature engineering. This makes them particularly well suited for applications like ADR prediction, where the relationships between drugs and adverse reactions can be intricate and difficult to define manually. Through the utilization of extensive datasets and the innate capacity of deep learning models to handle consecutive data, scientists can enhance their ability to accurately represent the sequential and temporal aspects of medication interactions in the human body. Drug sequence information, or the order in which medications are delivered or the ways in which they interact with different biological systems over time, is one domain in which deep learning is particularly effective. Sequence modeling is crucial in ADR prediction, as the timing and combination of drug usage can significantly impact the likelihood of an adverse reaction. For instance, a drug that is safe when used alone may produce harmful side effects when taken in combination with other medications. Deep learning models, particularly those designed for sequence data such as recurrent neural networks (RNNs) and transformer models, have shown promise in accurately predicting ADRs by learning from vast amounts of drug sequence information.

MedAware [59], is an ML-based model for AE prediction. It is designed as an alerting tool for clinical decision support (CDS). That model raises more alerts compared with traditional CDS; such alerts could prevent potential AEs.

In [60] the authors addressed the problem of predicting adverse events using FARES

data, which contains data on the patient's demographics, medications, their attributes (dosage, route, etc.) and the patient outcome associated with the medication (i.e., hospitalization and death). To do that, the authors used dimensionality reduction to reduce the number of features since too many unneeded features may decrease the model performance. The dimension-reduced data is then inputted into the Random Forest.

Others modeled the relationship between medical conditions and the drugs. In [61], The authors applied machine learning and natural language processing techniques to identify suspected drugs and their associated adverse events. They introduced two models: the first, called Causal Sentence Classification, determines whether the relationship between a medical condition and a drug is causal or not; the second, Suspect Drug Identification, evaluates each drug mentioned in a report to classify it as either a suspect or non-suspect drug. A suspect drug is defined as the medication most likely associated with or responsible for the reported adverse event. Their data contained 20,838 records and they focused on sentences that included the medical conditions and drug names, which acted as the pivot. For each text segment they extracted the tokens sur- rounding the pivot and generated n-grams as features. A total of 11,739 drugs were manually tagged as suspect or non-suspect drugs. The authors used an ensemble hierarchical classifier, which includes classifiers such as Artificial Neural Network (ANN), Logistic Regression (LR), and Random Forest (RF). Out of those models are fed into another layer of LR. This article shows significant results. Causal sentence classification resulted in precision of 0.85 and recall of 0.84, while suspect drugs achieved precision of 0.72 and recall of 0.77.

Social media became a platform for sharing personal experiences related to personal health. This information can provide benefit to applications such as pharmacovigilance. However, it is expressed in a highly informal language and the concept of the medical term is nontechnical and descriptive. This presents a challenge for information extraction such as adverse drugs. Named entity recognition can help in solving this problem, which is what the authors in [62]
proposed. The authors presented ADRMine, a sequence tagger for automating the extraction of ADR from user posts on social media using conditional random field (CRF). The ADRMine was trained on data collected from Twitter and DailyStrength. A pharmacology expert selected drugs based on their widespread use in the US. The data contained a total of 81 different drugs. ADRMine achieved an F1 score of 0.82 on the DailyStrength dataset, and 0.72 on the Twitter corpus.

The effort by experts to determine the adverse events of vaccination is high and takes a lot of time. The authors in [63] adopted an approach to automate text classification in the Vaccine Adverse Event Reporting System (VAERS). In their article, the authors explained multilevel text mining for automated text classification. They employed text mining techniques to extract three distinct sets of features, focusing on key terms, as well as both low-level and high-level linguistic patterns. High-level patterns are processed using rule-based classifiers while Machine Learning (ML)algorithms including naive Bayes, decision trees, and SVM, are trained on the remaining. This approach applied to the informative feature selection of 6,034 VAERS reports for the H1N1 vaccine. Their data included structured and unstructured data. This data indicates whether the patient had anaphylaxis from vaccines or not. Using multiple machine learning classifiers to evaluate the performance using macro F1, recall and precision. Firstly, they extract data using NLP for processing free text. Secondly, creating a list of lemmas called a dictionary for keywords of interest and improving the anaphylaxis lexicon, tagging, building grammar, and parsing the free text. The outcome showed the rule-based classifier achieved an average of 79.05 percent sensitivity and an average of 94.80 percent specificity.

In [64], the sequence label is considered a challenge in Electronic Health Records (EHR) and extracting the attributes from these records to represent medical events is not easy. In this article, the authors proposed a model based on trained RNN, Long-Short Term Memory (LSTM) and Gated Recurrent Units (GRU) on EHR notes for detecting diagnosis, adverse drug events, and medication. The dataset used in this article contains 780 EHR notes from cancer patients, each note annotated by two annotators for AE, drug name and severity. Skip-gram embedding is used to initialize word embedding layers for RNN and the output layer is a CRF.

The number of AE increased exponentially. In [65], the authors identified an automated way to label data as non-serious and serious, using a bag of words that uses Term Frequency- Inverse Document Frequency (TF-IDF). They trained a Word2Vec model using data from Med- line/PubMed, based on a dataset containing 13,887 adverse events (AEs). Each record included information such as patient ID, gender, AE description, and labels indicating whether the event was serious or non-serious. Several machine learning models were applied, including Logistic Regression with TF-IDF features, Logistic Regression with binary features, Random Forest with TF-IDF features enhanced by the Synthetic Minority Over-sampling Technique (SMOTE), Random Forest with binary features, and an

LSTM model using word embeddings. The mod- els achieved strong performance, with a mean F1-score of 0.95 and a Matthews Correlation Coefficient (MCC) of 0.92 for AE classification.

In [66] the authors evaluate multiple predictive models for AEs related to spine surgery in the US. The dataset was collected for patients who underwent spine surgery in the period 2009 to 2013. The dataset contains 345,510 and 760,724 patient records from Truven MarketScan and CMS, respectively. Most AEs happen 30 days after undergoing the surgery. They apply the selection operator regularization method, least absolute shrinkage, A logistic regression method was used to assess the risk of the most common adverse events (AEs). The models were trained on data containing information on patient demographics, the site of the spine procedure, existing comorbidities, the type of surgery conducted, and the preoperative diagnosis. The reported AUC is 0.7.

In [66], the authors developed algorithms to detect medical errors and adverse events (AEs) such as IV infiltrations, dosing errors, and oversedation. Analyzing 3,263 NICU notes from 753 patients, the algorithms outperformed trigger tools and voluntary reporting systems, identifying additional IV infiltration and overdose cases. For chest pain patients in emergency departments, the study highlighted the need for early intervention due to high risks of adverse cardiac events. In [67], the authors use clinical signs and heart rate variability (HRV) as they seek to discover the most important features of risk prediction for major adverse cardiac events (MACE) through comparing ML score selective features with modified early warning scores (MEWS) and thrombolysis in myocardial infarction (TIMI). Selective features are the most relevant variables identified for risk prediction, which enhance model performance by focusing on key predictors of an outcome. The dataset contains 702 patients entered into the emergency department in a tertiary hospital in Singapore. The age of the patient was at least 30 years, including un- differentiated non-traumatic chest pain. Data collection included an electrocardiogram (ECG) sensor and data acquisition device for monitoring patients' and 5-minute recording of heart rate variability. The clinical signs include measurements of systolic and diastolic blood pressure (BP), respiratory rate, heart rate, Glasgow Coma Scale (GCS), body temperature, pain score (on a scale of 1 to 10), and oxygen saturation (SpO2). Additionally, demographic details such as age, race, gender, and medical history were obtained from the emergency department records. The AEs include in this study "death, cardiac arrest, sustained ventricular tachycardia (VT), and hypotension requiring inotropes or intra-aortic

balloon pump (IABP) insertion". This article uses feature selection based on ensemble learning (random forest) to select the most relevant features from the eight features. Any one of eight that is not significant will be excluded. Then they applied a geometric distance to the outcome of a risk score ranging from 0 to 100, where 0 represents no risk and 100 denotes the highest level of risk. The result shows that of 702 patients, 29 of them met the MACE within 72 hours. The most relevant features which were selected are systolic blood pressure, the mean instantaneous heart rate, additionally, the mean RR interval. The results show the AUC is 0.812.

Women with congenital heart disease are more likely to have medical errors. The work proposed in [68] aims to build two models to predict when AEs will happen for pregnant women with congenital heart disease and their offspring. This helps clinicians to determine accurate treatment and management for pregnant women with congenital heart disease. The study us- ing a cohort study classified into the development cohort contained 213 pregnant women with congenital heart disease and the independent validation cohort contained 105 patients with con-genital heart disease. The dataset was collected from Qilu Hospital of Shandong University during the period from 2004 to 2019. The study analyzed multiple maternal and neonatal fac- tors and employed various machine learning algorithms (SVM, RF, ADs, DT, KNN, NB, and MLP) along with statistical methods like LASSO and multivariate logistic regression to enhance predictive models. The results showed that 12.9% of women experienced adverse maternal events, while 29.2% of neonates faced adverse neonatal events. Seven key maternal risk factors were identified: NYHA class, Eisenmenger syndrome, pulmonary hypertension, left ventricu- lar ejection fraction, sinus tachycardia, arterial oxygen saturation, and pregnancy duration. The machine learning algorithms demonstrated that the maternal model achieved an accuracy range of 0.76 to 0.86 and an area under the receiver operating characteristic curve (AUC) between 0.74 and 0.87 in the development cohort, and an accuracy of 0.72 to 0.86 with an AUC between 0.68 and 0.80 in the validation cohort. For the neonatal model, three high-risk factors were identified: Eisenmenger syndrome, preeclampsia, and arterial blood oxygen saturation. The neonatal model showed an accuracy range of 0.75 to 0.80 and an AUC of 0.71–0.77 in the de- velopment cohort, and an accuracy of 0.72 to 0.79 with an AUC of 0.69–0.76 in the validation cohort. Three high-risk factors were discovered in the neonatal model, including Eisenmenger syndrome, preeclampsia, and arterial blood oxygen saturation. The machine learning-based algorithms showed that the neonatal model had an

accuracy of 0.75 to 0.80 and AUC 0.71–0.77 on the development cohort, and 0.72 to 0.79 and AUC of 0.69–0.76 on the validation cohort.

In [69], the authors use rule-based methods to extract AEs that occurred in the treatment period. This article converts the EHR into a semantic structure to extract AEs. The dataset was collected from the neurosurgical department. The linguistic and medical concepts are identified, looking for whether a medical concept is found or not, and assessing the severity through linguistic concepts. If the words are negative expressions, they will add them to AEs. Using SVM for a bag of words to compare the results. The result was a 0.65 recall and 0.78 precision. Table 2.1: Overview of some existing studies on adverse event (AE) prediction and detection using various AI/ML and NLP techniques. The table summarizes the methodologies, data sources, sample sizes, and key performance metrics across a range of approaches applied to pharmacovigilance and related medical domains.

Table 2.1: Summary of some ML/NLP Approaches for Adverse Event Prediction and Analysis

Problem	**Method**	**Data**	**Sample Size**	**Results / Accuracy**
Drug-medical condition relationship	NLP (n-gram), ANN, RF, LR	11 features from drug frequency/position in sentence	20,838 records	Precision: 0.8 Recall: 0.84; Suspect drugs Precision: 0.7 Recall: 0.77
ADR extraction from social media	CRF	Drug, symptoms, dosage, family, date	711,562 posts	F1-score: 0.82 (DailyStrengtl 0.72 (Twitter)
Text classification of VAERS reports	Rule-based, NB, DT, SVM	Keywords, pattern data	6,034 VAERS reports	Sensitivity: 79.05%, Specificity: 94.80%
Detecting diagnosis, ADEs, medication	LSTM, GRU	AE, drug name, dosage, duration, route, etc.	780 EHR notes	F1-score: 0.8(
Label serious/non-serious AEs	LR, RF, LSTM, SMOTE	ID, gender, AE, seriousness	13,887 AEs	F1-score: 95% MCC: 0.92
Top AEs in spine surgery	LASSO, Logistic Regression	Demographics, comorbidities, procedure data	345,510 + 760,724 records	AUC: 0.70
MACE risk prediction	RF feature selection	Vitals, demographics, history	702 ED patients	AUC: 0.812
AEs in pregnancy with CHD	Cohort + ML (SVM, RF, etc.)	Maternal + pregnancy features	213 + 105 patients	Accuracy: 0.76–0.86
Extracting EHR data	Rule-based	Medical terms,	Unknown	Recall: 0.65,

A significant limitation of prior studies is the assumption that drugs act in isolation, failing to account for time-dependent and sequence-dependent drug–drug interactions. These include diminished therapeutic efficacy due to coadministration timing and adverse drug reactions triggered by improper sequencing—critical factors in polypharmacy and chemotherapy protocols

2.3 Chapter Conclusion

The literature review underscores the progression of AE detection methodologies from rule-based and statistical approaches to machine learning and deep learning techniques. While rule- based and statistical methods offer foundational insights, their limitations in handling complex drug interactions have paved the way for machine learning models that leverage large-scale datasets. The integration of deep learning, particularly recurrent and transformer-based architectures, has significantly enhanced the predictive capabilities of AE detection systems by effectively capturing drug sequence information. Studies exploring AE detection from diverse sources, including structured datasets, electronic health records, and social media, demonstrate the versatility of modern approaches. Despite these advancements, challenges such as data sparsity, interpretability, and the need for domain expertise remain. Future research should focus on refining deep learning models to improve generalizability, integrating multi-source data, and enhancing explain ability to ensure reliable and actionable clinical decision support.

Chapter Three

Data Analysis

3.1 Introduction

The accurate mapping of drug names and adverse reactions is a critical step in pharmacovigilance and drug safety analysis. The Food and Drug Administration's Adverse Event Reporting System (FAERS) provides a vast repository of drug-related adverse event reports, but the dataset requires extensive preprocessing to ensure consistency and interoperability. This chapter out- lines a comprehensive methodology for standardizing drug names by integrating FAERS data with Drugs@FDA and RxNorm, enhancing the identification and classification of adverse drug reactions (ADRs). Furthermore, we discuss the mapping of medical indications and reactions using the Medical Dictionary for Regulatory Activities (MedDRA) terminology to identify potential safety signals. The meticulous integration and validation of these mappings play a vital role in improving the reliability of FAERS data analysis and ensuring compliance with industry standards.

3.2 FAERS

The FDA manages an internal FAERS [70, 71] database and periodically releases some post-market information every quarter. The open FDA platform and the FAERS online dashboard allow access to the public release. The FDA receives over one million annual reports of adverse events and medication errors, making its database one of the largest for pharmacovigilance. The FAERS dataset, spanning from 1989 to 2017 with nearly 6 million events, includes detailed in- formation on adverse event types, severity, patient

demographics, drugs involved, dosages, out- comes, and medical histories. This comprehensive data helps researchers, healthcare providers, and regulatory authorities detect safety issues, track trends, and make informed decisions on drug safety and efficacy, ultimately improving patient care. The FDA Adverse Event Reporting System (FAERS) is a database used for the voluntary reporting of adverse events and medication mistakes that includes human medicines and therapeutic biological products. Table 3.2shows

Table 3.1: Description of Dataset Features used from FAERS

Feature	**Description**
AGE	Age of the patient at the time of the adverse event
GENDER	Gender of the patient
COUNTRY_CODE	Country code of the reporting country
d	Drugs aligned with the adverse event
Role_code	Role code of the drug in the adverse event
DRUG_SEQ	Sequence number of the drug in the adverse event
p	Dose of the drug
rt	Route of administration of the drug
ri	Adverse reaction aligned with the drug

the data feature descriptions and Figure 3.1 shows the volume of these reports. By the year, the FDA receives over one million AEs and medication error reports in the English language associated with the use of drug or biologic products. FAERS includes all product lines from the United States, reports from all over the world, and also includes patient populations (elderly, children, and pregnant women). In addition to the features mentioned above, researchers prefer FAERS for its easy access to and public availability.

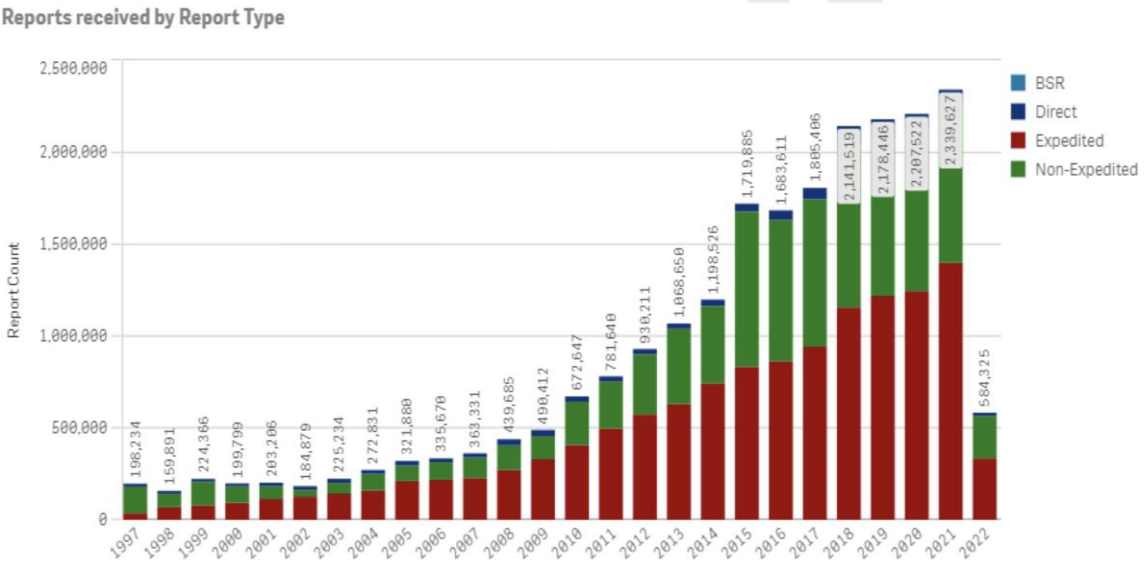


Figure 3.1: Number of reports submitted to the FDA

One of the primary advantages of FAERS is the extensive dataset it offers, but there are significant challenges in using the data effectively. Reports submitted to FAERS come from various global regions, which leads to inconsistencies in data entry. Another major concern in FAERS, as with any automated adverse reporting system, is the presence of duplicate reports. The same adverse event can be submitted multiple times by different parties, such as patients, healthcare professionals, or manufacturers, which further inflates the dataset. Additionally, the absence of clear naming conventions for drugs and adverse reactions introduces noise into the data, making it critical to preprocess the dataset thoroughly before conducting any meaningful analysis. Effective text processing techniques are required to standardize the data, remove du- plicates, and correct spelling errors, which helps to normalize drug names and reactions. This process, however, is both time-consuming and computationally intensive, requiring significant resources to enhance data quality. Other limitations include the absence of temporal modeling between administered drugs, and the exclusion of unstructured data such as case notes and dis- charge summaries, which may contain valuable clinical context, as emphasized by [72] in their study on leveraging unstructured healthcare data for ADE detection.

For this project, we utilized FAERS data from the year 2021, which comprises 1,860,388 reported cases and 6,331,148 associated drugs. It's important to note that these numbers are inflated due to the inclusion of normalized drug names and typographical errors. In addition to drug information, FAERS provides demographic details about patients, including their

age, gender, and geographic location at the time of the adverse event. The database also contains vital medical information, such as the timing of the adverse event, the nature of the incident (whether it was a side effect or medication error), and the outcome of the event (such as hospitalization or death). Furthermore, FAERS includes detailed product information, specifying the name of the drug, its dosage form, administration method, and active ingredients. The sources of these reports vary, ranging from healthcare providers to manufacturers and even consumers, which adds further variability to the data.

In our dataset, over 15,000 unique adverse reactions are recorded. Like the drug entries, the reaction counts are inflated due to the presence of duplicate or unnormalized ADRs. On average, each drug sequence in the dataset is linked to around 20 adverse reactions, resulting in a total set of 800,000 drug-reaction sequence pairs. The high number of reactions associated with each drug sequence makes the task of modeling the problem as a multiclass classification particularly challenging. Each reaction is treated as a separate class, and given the large number of possible reactions, the model requires a highly complex structure to manage this level of granularity.

A key challenge arises from the fact that the adverse reactions do not exhibit an explicit ordinal relationship or interdependence that could be leveraged in a traditional sequence-to-sequence task. Each reaction within a sequence is independent of the others, which complicates the process of predicting adverse reactions based on a drug sequence. Given this complexity, the task cannot be easily framed as a sequence prediction problem, nor can it be treated as a hierarchical classification. Instead, specialized architectures must be developed to manage the sparse network between the last hidden layer and the output layer, ensuring that the model is capable of handling the wide variability in the data.

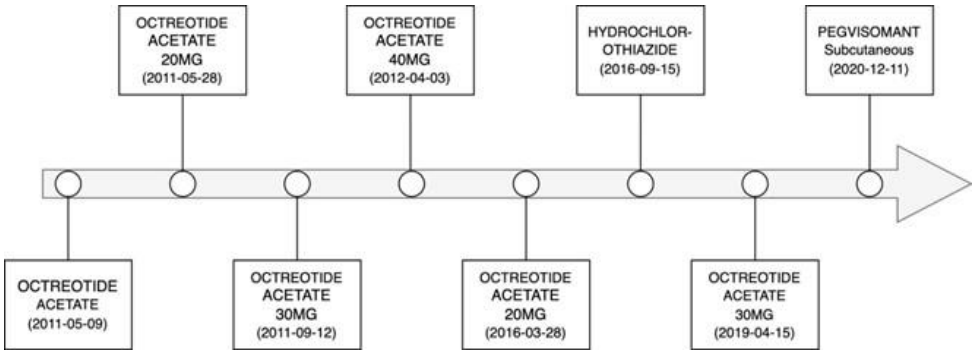


Figure 3.2: Drug Sequence From FARES Data

An example case from the FAERS dataset is shown in Figure 3.2, illustrating a patient's

sequence of drugs taken over time. This case involves a 61-year-old female patient from Canada who experienced multiple adverse reactions, including arthralgia, increased blood cholesterol, bronchitis, dizziness, hypotension, and rheumatoid arthritis. The outcome for this case was hospitalization, which highlights the seriousness of the adverse events reported. This single example demonstrates the complexity of the data, with multiple drugs and reactions involved, making it critical to develop a robust methodology for analyzing and predicting ADRs in such diverse cases.

3.2.1 Method and Material for FAERS data

FAERS data is publicly available as quarterly downloads in two formats: extensible markup language (XML) and ASCII files Comma-Separated Values (CSV). FAERS data spans from September 2012 through June 2022, while Legacy LAERS data covers January 2004 to August 2012. FAERS data comes from patients, healthcare professionals, manufacturers, and pharmaceutical suppliers. FAERS source data is split into seven tables, as shown in Table 3.2.

Table 3.2: FAERS source data structure

Filename	**Description**
DEMOyyQq	Patient demographic and administrative information
DRUGyyQq	Biologic/drug data for all drugs included in the event report
INDIyyQq	MedDRA terms for drug indications in the event report
OUTCyyQq	Patient outcomes for the event report
REACyyQq	MedDRA terms for the adverse event report
RPSRyyQq	Source of the event report
THERyyQq	Drug therapy start and end dates for reported drugs

3.2.2 Trends in the Size and Growth of the FAERS Database

Figure 3.3 illustrates significant increase in reports over years. Over this period, the number of records in the database demonstrates a steady increase initially, with a moderate growth rate from 2004 to around 2008. Following this period, there is a noticeable acceleration in the reporting rate, indicating a significant surge in adverse event reports submitted to the

FDA. This trend continued until 2017, when the database reached a peak record count of close to 1,250,000. However, in 2018, there was a slight decline in the number of records due to a shift in reporting trends with a notable rise in submissions from health professionals but a decline from non-health professionals. The total count of records over the entire period is 11,170,959. This growth pattern highlights the increasing focus on adverse event reporting and monitoring, likely influenced by regulatory changes, enhanced awareness, and improved reporting mechanisms. The figure 3.4 displays the distribution of adverse event reports in the FAERS database based on

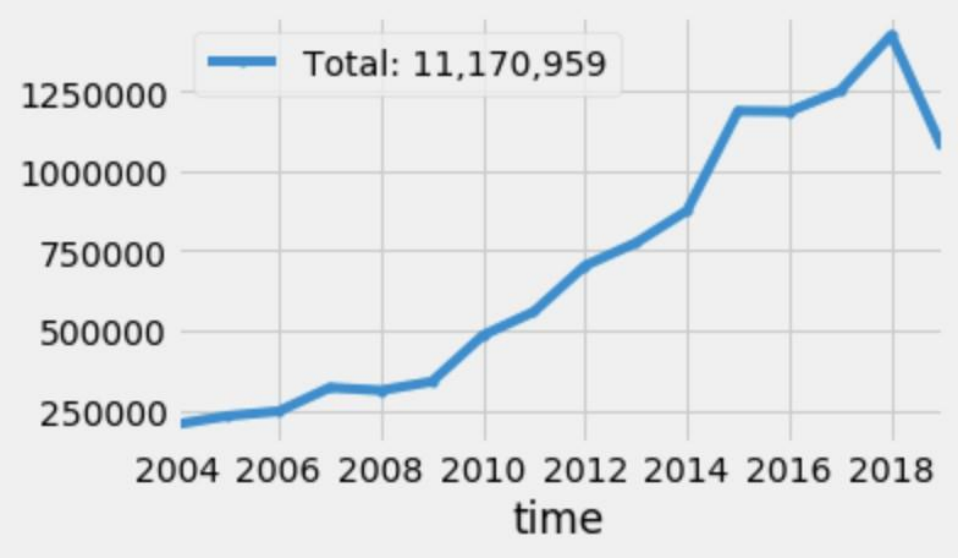


Figure 3.3: FAERS Database Trends

patient sex. The x-axis represents the number of reports, while the y-axis categorizes the reports by sex: Male, Female, and Unknown. The chart shows that approximately 60% of the reports are for female patients, with the number of reports for females reaching around 6,000,000. In contrast, about 40% of the reports are for male patients, with the number of reports for males slightly over 4,000,000.

This distribution suggests that sex could be an important confounding factor in analyzing adverse event data. The significant difference in the number of reports between females and males likely reflects biological and physiological differences such as pregnancy related issues, hormonal variations and differences in metabolism or immune responses. Females may be more likely to seek medical care or report adverse events. Recognizing and accounting for these factors is essential for accurate interpretation and application of the data in healthcare research and decision-making.

In Figure 3.5, we showed the distribution of adverse event report submissions to the FAERS database by different types of submitters. The x-axis represents the number of reports, while

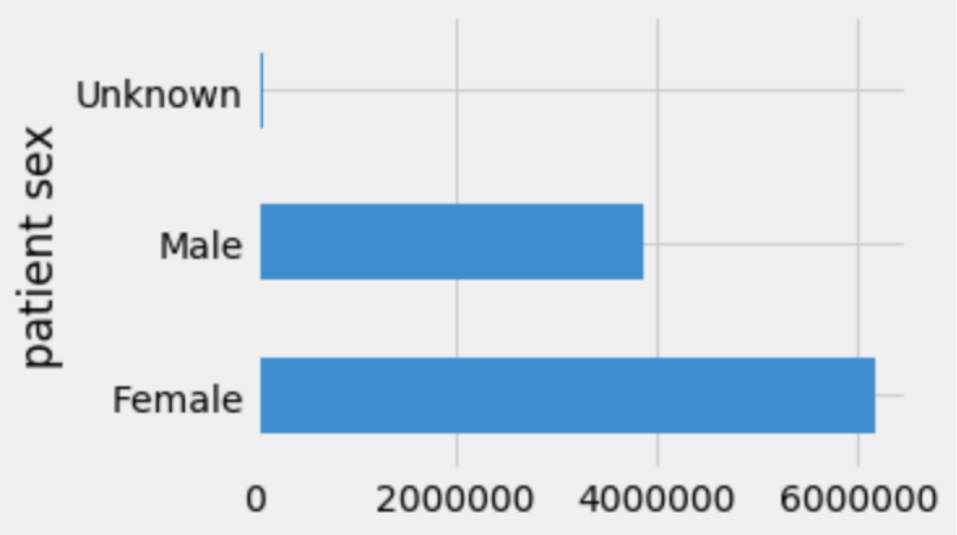


Figure 3.4: Patient Sex

the y-axis categorizes the submitters into five groups: Lawyer, Pharmacist, another health professional, Physician, and Consumer or non-health professional. The most significant category is "Consumer or non-health professional," with submissions totaling over 4,000,000 reports. This suggests that individuals outside the healthcare profession initiate a substantial portion of adverse event reports. The second-largest group is the "Physician," which contributes a considerable number of reports, slightly over 2,000,000. "Other health professionals" follow, with their submissions reaching a similar magnitude, slightly below those of physicians.

In comparison, "Pharmacist" submissions are notably lower, hovering around 500,000 reports. The smallest category is "Lawyer" with a minimal number of submissions, barely visible on the chart. This distribution highlights the varied sources of adverse event reports, emphasizing the significant role of consumers and non-health professionals in reporting. The data indicate that while healthcare professionals contribute substantially, the general public en- gages broadly in the reporting process. This discrepancy arises because consumers and non- health professionals are often the first to notice and report adverse effects experienced firsthand whereas healthcare professionals may under report due to time constraints or reliance on patient driven reports. Understanding these dynamics is crucial for interpreting

the data and improving adverse event monitoring and response systems.

Figure 3.5: Data Submitters

Figure 3.6 is derived from the FDA dataset and illustrates the frequency of occurrences for the top 10 drugs. The x-axis lists the drug names, including Zantac, Inflectra, Lyrica, Xyrem, Truvada, Xolair, Cosentyx, Viread, Ranitidine, and Nexium. The y-axis represents the count of occurrences for each drug, ranging from 0 to 18,000. The chart indicates that Zantac has the highest occurrence, with approximately 17,500 counts, significantly more than the other drugs listed. Following Zantac, Inflectra, Lyrica, Xyrem, Truvada, Xolair, and Cosentyx have similar counts, ranging between 4,000 and 5,500 occurrences. Viread, Ranitidine, and Nexium have lower counts, around 3,000 to 3,500, but still make it to the top 10 list.

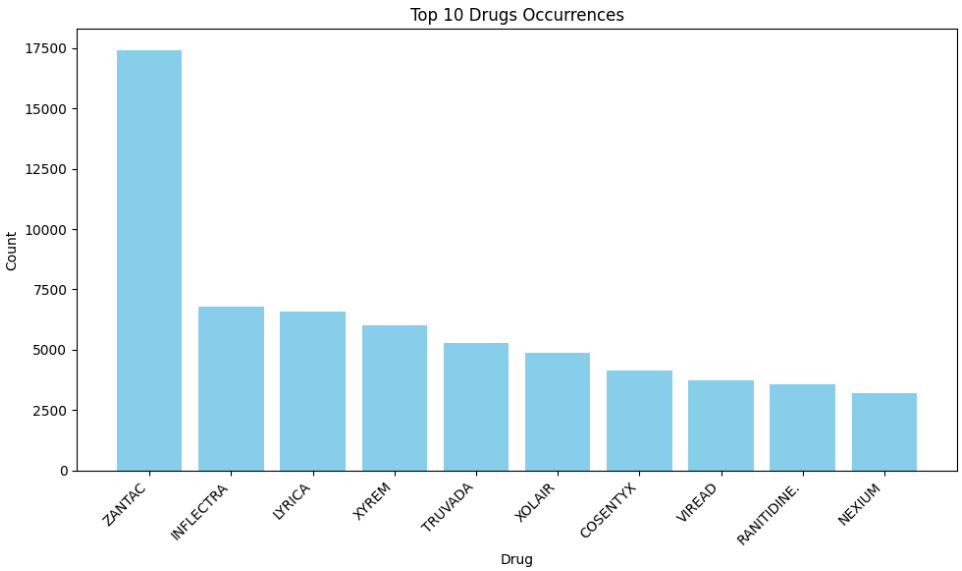


Figure 3.6: Top 10 Drugs

3.3 MedDRA

In the Previous sections, we discussed the features of the FAERS dataset, which has more than 50000 classes and is hard to classify them into particular reactions. In this section, we are go- ing to use another dataset, MedDRA, that helps us to map the reaction into lower classes. The dataset is a comprehensive resource curated to enhance NLP and machine learning applications in the medical domain. This dataset contains a vast collection of annotated digital medical records, encompassing various medical entities such as diseases, symptoms, treatments, and medications. Each record is meticulously labeled to facilitate developing and evaluating algo- rithms for information extraction, entity recognition, and text classification within healthcare texts. By providing a rich and diverse set of annotated medical data, MedDRA aims to support advancements in medical research, clinical decision-making, and the development of intelligent healthcare systems. Table 3.3 provides a structured overview of entries in the MedDRA dataset, which categorizes adverse events and medical terms for regulatory and analytical purposes. Each row represents a unique term identified by a combination of an ID, primary (CID1) and secondary (CID2) category identifiers and classification levels such as LLT (Lowest Level Term) or PT (Preferred Term). The Type column provides subtype codes while the Description column offers a user friendly explanation of the term such as "Abdominal pain" or "Anaemia". Multiple rows for the same CID2 reflect the hierarchical nature of MedDRA where terms are categorized under different levels for detailed analysis. This structure ensures clarity and consistency in recording and interpreting medical terms highlighting its importance in regulatory contexts like adverse event reporting.

The MedDRA dataset is a standardized medical terminology used globally for regulatory purposes, including the classification and analysis of adverse events and medical conditions. Its structure is hierarchical, consisting of five levels, each providing increasing specificity:

1. **System Organ Class (SOC):** The highest level, representing broadest classifications, such as “Gastrointestinal disorders” or “Nervous system disorders.” It groups related

medical terms by organ system or etiology.

2. **High-Level Group Terms (HLGT):** These are subsets of SOC, providing a more specific grouping of related medical concepts within an organ system or cause.
3. **High-Level Terms (HLT):** A further subdivision of HLGT, offering greater granularity by clustering related conditions or symptoms.
4. **Preferred Terms (PT):** The main level used for reporting and analysis, representing distinct medical concepts (e.g., "Abdominal pain" or "Anaemia"). Each PT is linked to one SOC.
5. **Lowest Level Terms (LLT):** The most detailed level, capturing synonyms, lexical variants, and other descriptive terms corresponding to the PT (e.g., "Stomach ache" as an LLT under the PT "Abdominal pain").

Table 3.3: MedDRA Dataset Overview

ID	CID1	CID2	Code	Type	Description
CID100000085	CID000010917	C0000729	LLT	C0000729.1	Abdominal cramps
CID100000085	CID000010917	C0000729	PT	C0000737	Abdominal pain
CID100000085	CID000010917	C0000737	LLT	C0000737	Abdominal pain
CID100000085	CID000010917	C0000737	PT	C0687713	Gastrointestinal pain
CID100000085	CID000010917	C0000737	PT	C0000737	Abdominal pain
CID100000085	CID000010917	C0002418	LLT	C0002418	Amblyopia
CID100000085	CID000010917	C0002418	PT	C0002418	Amblyopia
CID100000085	CID000010917	C0002871	LLT	C0002871	Anaemia
CID100000085	CID000010917	C0002871	PT	C0002871	Anaemia

This hierarchical structure allows precise mapping of medical data, facilitates consistent terminology use across global healthcare systems, and supports accurate aggregation and analysis of adverse events and other clinical data. Table 3.4 shows a sample of medical terms classified according to the MedDRA hierarchy, which helps standardize reporting and analysis of adverse events.

Table 3.4: Example of MedDRA Terms

MedDRA Term	**System Organ Class (SOC)**	**Preferred Term (PT)**	**Lower Level Term (LLT)**
Headache	Nervous system disorders	Headache	Throbbing headache
Nausea	Gastrointestinal disorders	Nausea	Feeling nauseous
Rash	Skin and subcutaneous tissue disorders	Rash	Skin rash
Myocardial Infarction	Cardiac disorders	Myocardial infarction	Heart attack
Anaphylaxis	Immune system disorders	Anaphylactic reaction	Severe allergic reaction

3.4 Data Processing

3.4.1 Data Cleaning

The dataset underwent a structured cleaning process to improve clarity and analytical reliability. The 'GENDER COUNTRY_CODE OCCP_COD' column was renamed to 'GENDER' and re- fined to retain only 'M' or 'F'. Columns like 'aligned_drugs', 'aligned_route', and 'ALIGNED_REAC' with multiple entries were split into separate columns (e.g., 'drug1', 'drug2') and recombined, filling missing values with 'null'. Entries with "UNK" in the 'dosage' column and "Unknown" in the 'route' column were flagged and removed to ensure data completeness. Lastly, key columns were converted to lowercase for consistency, enabling stream- lined and accurate downstream analysis.

3.4.2 Case De-duplication

The FDA recommends reduplication to retain the most recent report for each case identified by its unique case identification. In this process, cases were deduplicated primarily based

on date, age, and report country as these provide essential contextual information, adjusted based on feedback. Additional fields, such as therapy start date and therapy duration, were included to improve case matching and reduce redundancies.

The reduplication process employs a systematic approach, outlined through the following equations. Let C represent the set of all cases:

$$c = \{c_0, c_1, c_2, \ldots, c_n\}$$

Here, c_i is the i-th case, characterized by a vector of features:

$$c_i = \{ f_{i1}, f_{i2}, f_{i3}, \ldots, f_{im}\}$$

If the feature values of two cases, c_i and c_j, are identical across all selected features, they are considered a full match:

$$\exists c_i, c_j \in c : f_{im} = f_{jm}, \forall 1 \leq m \leq m_{\max} \Rightarrow \text{Full Match}$$

For partial matches, specific features are excluded: If all features except *event date* are identical:

$\exists c_i, c_j \in c : f_{im} = f_{jm}, \forall m \not= \text{event date} \Rightarrow$ Full Match - Event Date If all features except *age* are identical:

$$\exists c_i, c_j \in c : f_{im} = f_{jm}, \forall m \not= \text{age} \Rightarrow \text{Full Match - Age}$$

Similarly, for *country report* and *gender*:

$$\exists c_i, c_j \in c : f_{im} = f_{jm}, \forall m \not= \text{country report} \Rightarrow \text{Full Match - Country Report}$$

$$\exists c_i, c_j \in c : f_{im} = f_{jm}, \forall m \not= \text{gender} \Rightarrow \text{Full Match - Gender}$$

In essence, if the feature values of two cases are identical across all relevant features, they are considered a full match. This process is particularly important when a patient submits

identical reports multiple times at the same moment; by identifying full matches, only the most recent submission is retained to avoid redundancy and ensure data quality.

3.4.3 Drug Mapping for FAERS Data

Drugs@FDA is an online database maintained by the U.S. FDA that provides information about approved prescription and over the counter drugs.This section outlines the process for aligning drug names in the FAERS dataset with Drugs@FDA. The key steps are as follows:

- Download drug tables for each quarter from the FAERS database and combine them into a single table, called drugs2022.
- Use the New Drug Application (NDA) number to map drug names to the Drugs@FDA database, which is already mapped to RxNorm.

- Download the Drugs@FDA dataset from the FDA website (Date Updated: June 7, 2022).

- Use regular expressions to clean drug names (e.g., removing trailing spaces or mysterious symbols).
- Use the International Drug Dictionary (IDD) to link trade names from around the world to RxNorm, improving the mapping of undefined drug names.

After these steps, approximately 2% of drug names remain undefined. These could be due to unregistered names in the U.S., spelling errors in the FAERS entries, or non-specific drug names. Drugs with undefined or missing names were retained as-is in the dataset to allow the model to learn from the full spectrum of available data.

3.4.4 Data Validation

To ensure the integrity and accuracy of the dataset, a combination of manual audits and auto- mated checks was employed. Manual validation involved reviewing 20 randomly selected cases to verify the accuracy of critical fields, such as patient demographics, drug information and ad- verse event details, cross-referencing them with the original data to identify inconsistencies or missing values. Automated SQL queries were then applied to

check for referential integrity between related tables, consistent data formats, identify duplicate records and validate logical relationships (e.g., therapy start dates preceding event dates). These combined efforts ensured that the database was free of critical errors and prepared for reliable analysis with plans for ongoing validation to maintain data quality.

Table 3.5: Description of Dataset Features after Mapping MedDRA Dataset

Feature	**Description**
age	Age of the patient at the time of the adverse event
dosage	Dosage of the drug administered
route	Route of administration of the drug
start_date	Date when the drug administration started
end_date	Date when the drug administration ended
aligned_reaction	Adverse reactions aligned with the drug
gender	Gender of the patient
name	Name of the drug
location	Location of the report submission
ontologies	MedDRA ontology codes corresponding to the adverse reactions

3.5 Chapter Conclusion

The comprehensive mapping of drug names and adverse reactions in FAERS significantly en- hances data reliability and usability for pharmacovigilance research. By integrating FAERS with Drugs@FDA, RxNorm, and OHDSI vocabularies, we successfully standardized drug nomenclature and ensured consistency in adverse event classification. Our approach achieved a 93% coverage rate in drug name mapping and standardized over 80% of reported adverse reactions.

The implementation of contingency tables and disproportionality metrics further enabled robust signal detection for potential drug safety concerns. Additionally, our validation procedures, which included manual audits and automated consistency checks, ensured high data integrity. Despite challenges such as missing or ambiguous drug names, our methodology provides a replicable framework for improving FAERS data analysis. Future work may focus on refining mapping techniques and incorporating machine learning approaches to enhance data quality and predictive capabilities in pharmacovigilance.

Chapter Four

Model Architecture

4.1 Introduction

The accurate prediction of ADRs is a critical challenge in pharmacovigilance, necessitating the development of advanced computational models. This chapter presents the architecture of a neural network model designed to predict ADRs based on drug sequences and adverse event reactions. By leveraging two distinct encoders, the model processes structured drug sequence data. The architecture integrates these data sources - drug sequences and adverse events through embedding projections, concatenation, and multiple transformation layers, culminating in a SoftMax-based classification. This approach enables a comprehensive assessment of potential ADRs by capturing both sequence-based and semantic relationships within the input data. The chapter details the various components of the model, including the encoding mechanisms, representation refinement, and final classification process, while also exploring potential enhancements to improve predictive accuracy.

4.2 Architecture of the Predictive Model

Predictive modeling techniques rely heavily on historical datasets to identify underlying patterns and trends. These foundations inform the encoder-decoder framework utilized in sequence modeling [85]. figure 4.1 illustrates the architecture of the neural network model designed for predicting adverse drug reactions based on two distinct types of input data: drug sequences and adverse event reactions. The model is composed of two primary

encoders that process these different data streams. Following the encoders, the outputs are passed through a representations layer, which projects the embeddings from both sources. Then, a concatenation combines the projected representations, and several projection layers are applied to transform the combined output further. The final step computes the SoftMax to determine if the drug sequence causes the ADR or not. The purpose of this model is to predict the likelihood of a sequence of drugs. By integrating these two types of information- drug sequences and adverse events- , the model can more accurately assess potential drug-related risks.

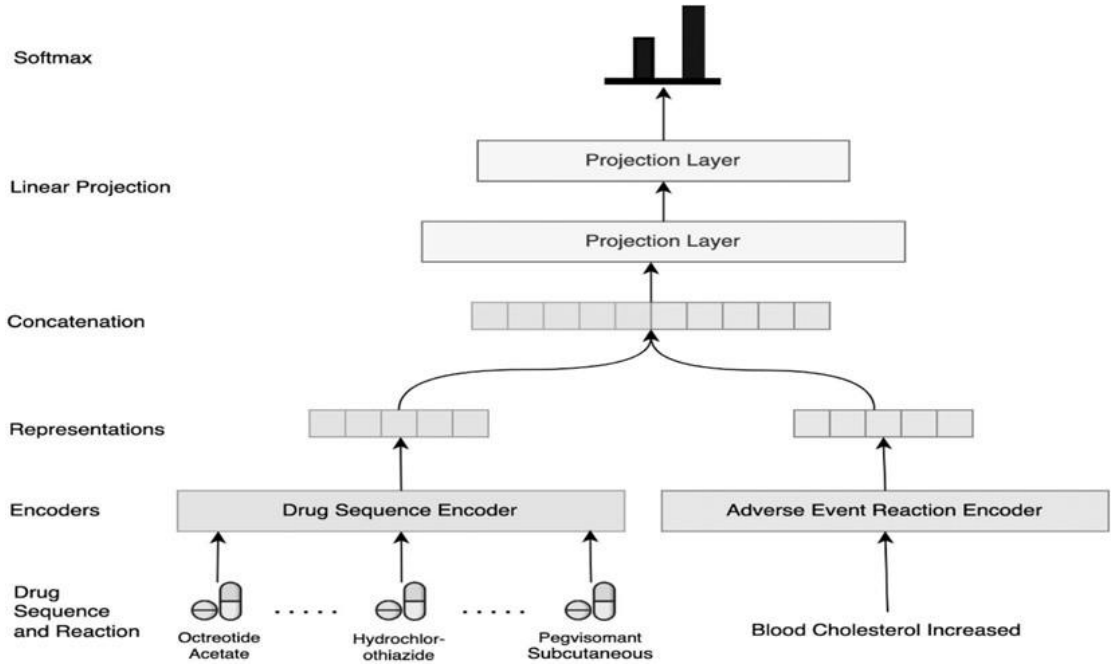


Figure 4.1: Model architecture for predicting adverse drug reactions

4.2.1 Encoders

The model utilizes two distinct encoders to handle drug sequences and adverse event reactions, converting them into meaningful numerical representations. Drug Sequence Encoder processes the input drug sequences, which may include individual drugs or combinations of multiple drugs. It extracts relationships between the drugs and transforms them into numerical embeddings. These embeddings can incorporate both word-level and character-level information, enabling the model to generalize for unseen or rare drug names. Typically, recurrent neural networks (RNNs), such as BiLSTM, are used for this sequential data processing. The output is a high-dimensional vector capturing the relevant features of the drugs. The adverse event re- action encoder

transforms textual descriptions of side effects into dense vector representations that preserve semantic context. Transformer-based architectures have recently demonstrated strong performance in adverse event prediction tasks, offering high precision and scalability across clinical datasets [73, 74, 75, 76]. Pre-trained models like RoBERTa can be fine-tuned to generate embeddings that capture relevant linguistic features from adverse reaction reports, enabling downstream models to effectively process complex biomedical language.

4.2.2 Projection Layer

After the data has been encoded, the outputs from the two encoders are passed through a representations layer. This layer further refines the embeddings, transforming them into higher- quality representations. The purpose of this layer is to ensure that the encoded information is optimally prepared for the next stage of processing, particularly by focusing on enhancing the most important features extracted from the drug sequences and adverse event reactions.

The concatenated layer combines the embeddings from the two encoders into a single unified vector. This step integrates information from both data streams—drug sequences and ad- verse event reactions—allowing the model to jointly consider the interactions between these two sources in its predictions.

The concatenated feature vector is passed through several projection layers, which serve to reduce dimensionality and distill the most relevant features. These fully connected layers apply linear transformations followed by non-linear activation functions (e.g., Rectified Linear Unit (ReLU)) to project the data into a lower-dimensional space. This transformation helps the model focus on the most important information while minimizing the risk of overfitting.

4.2.3 Softmax Layer

A softmax layer will output the final class probabilities. This layer takes the features projected from the previous layers and converts them into a probability distribution over all potential adverse drug reactions. By applying the softmax function, the model ensures that the sum of probabilities is equal to 1. This allows the model to assign a likelihood to

each possible adverse reaction, offering a more interpretable and probabilistic prediction. The adverse reaction with the highest probability is deemed the most likely outcome. Although multiple adverse reactions may sometimes share similar probability scores, the current model prioritizes these predictions based solely on their calculated probabilities, providing a clear and unbiased foundation for subsequent clinical evaluation and decision-making.

4.3 Potential Model Improvements

While the current neural network architecture is robust to drug sequences and adverse event reactions, several ways exist to enhance its performance further.

Attention Mechanisms Incorporating attention mechanisms could enable the model to focus on specific parts of the drug sequences or adverse event reactions that are more relevant to predicting adverse drug reactions. This approach allows the model to weigh certain features more heavily, which could improve prediction accuracy.

Hierarchical Structures Drug sequences often have complex relationships, particularly when multiple drugs are involved. Using hierarchical encoding methods could capture these nested relationships, allowing the model to represent the interactions between drugs more effectively.

Transfer Learning Leveraging pre-trained models such as BERT for the adverse event reaction encoder could significantly enhance the model's ability to handle complex natural language inputs. Pre-trained language models are especially beneficial when working with limited datasets, as they bring rich contextual understanding learned from large-scale corpora.

Latent Dirichlet Allocation (LDA)[41] The current architecture processes adverse event reactions using neural encoders, while LDA is applied as a data processing method to uncover latent topics in the adverse event data. LDA identifies interpretable patterns or clusters of related adverse events, providing structured features that complement the neural encoding. These topic features, derived from LDA, enhance the input to the adverse event reaction encoder by offering additional context and insights into the data. Integrating LDA with neural models creates a more robust framework for identifying patterns in adverse reactions, improving predictive accuracy.

4.4 Chapter Conclusion

The predictive model outlined in this chapter demonstrates a structured and effective approach to ADR identification by integrating drug sequence and adverse event reaction data. Through dedicated encoders, projection layers, and a softmax classification mechanism, the model effectively processes and interprets complex relationships within pharmacological data. While the current architecture is robust, further improvements such as attention mechanisms, hierarchical encoding, and transfer learning could enhance its predictive capabilities. Additionally, the integration of Latent Dirichlet Allocation (LDA) provides a complementary method for identifying latent patterns in adverse event data. Future work should explore these enhancements to refine the model's accuracy and reliability, ultimately contributing to safer and more effective drug administration practices.

Chapter Five

Probabilistic Topic Modeling for Dimensionality Reduction in Pharmacovigilance Datasets

5.1 Introduction

The FAERS dataset contains over 50,000 unique adverse event (AE) categories, making direct classification impractical. To address this, Latent Dirichlet Allocation (LDA) was applied to cluster reactions into 10 broader, interpretable topics. Unlike simple categorization, LDA captures latent structures, enabling better generalization and reducing data sparsity.

Replacing raw reaction terms with topic-based labels provides a more structured representation of AEs, improving interpretability and enhancing the performance of downstream tasks such as sequence-to-sequence (Seq2Seq) prediction..

5.2 LDA Overview

Latent Dirichlet Allocation (LDA) is a generative probabilistic model that identifies hidden thematic structures within textual data. By treating each document as a mixture of latent topics, LDA automatically extracts meaningful clusters, making it a key technique in NLP tasks such as topic modeling and information retrieval.

This chapter outlines the LDA implementation process—from text preprocessing and tokenization to dictionary construction, corpus formation, model training, and topic visualization. These steps enable efficient and interpretable analysis of adverse event narratives, supporting deeper insights in text mining and pharmacovigilance.

5.2.1 Preprocessing

The initial phase of implementing LDA involves preprocessing the text data, which is critical for ensuring the quality and relevance of the subsequent analysis. This stage encompasses various sub steps designed to clean and refine the text. Key tasks include removing stop words—common words that add little semantic value—and eliminating non-alphabetic characters that could introduce noise into the dataset. Additionally, normalization techniques, such as converting all text to lowercase, may be applied to maintain consistency and reduce variability. This comprehensive cleaning process prepares the text for further analysis, allowing the LDA model to focus on the most meaningful components of the data.

5.2.2 Tokenization

Tokenization is a crucial step in preparing textual data for analysis using LDA. This process involves segmenting the cleaned text into smaller units known as tokens, typically individual words or phrases. These tokens are essential for transforming unstructured text into a structured format suitable for model analysis. For instance, the phrase "adverse drug reactions" is split into three tokens: "adverse," "drug," and "reactions." We utilized the Python package NLTK for tokenization, specifically the word tokenize function. This function identifies token boundaries based on spaces, punctuation marks, and other delimiters. Additionally, multi-word expressions (e.g., "New York") were preserved using a custom tokenizer with added phrase detection to handle domain-specific terms effectively.

Tokenization also included preprocessing steps such as handling contractions (e.g., converting "don’t" into "do" and "not") and addressing case normalization. These tokens were further prepared for LDA by removing stop words, applying stemming, and filtering out infrequent terms. This structured tokenized text enabled the LDA model to identify latent topics and pat- terns, supporting a comprehensive analysis of the adverse reaction dataset.

5.2.3 Creating a Dictionary and Filtering Extremes

After completing tokenization, the next step is constructing a dictionary representation of the documents. This dictionary maps each unique token identified during tokenization to a unique integer ID, creating a structured representation of the vocabulary used in the corpus.

We used the Dictionary class from the gensim Python library to accomplish this. This step ensures that the text data can be efficiently transformed into numerical formats for topic modeling algorithms like LDA. To create the dictionary, we passed the tokenized documents as input to gensim.corpora.Dictionary, which generates a mapping of tokens to integer IDs. For instance, the token *"adverse"* may be assigned the ID 0, and *"reactions"* may be assigned the ID

1. This mapping facilitates efficient lookups and compact representation of the text. Filtering extremes is a critical subsequent step to enhance the quality of the model. Tokens that appear in too few documents are removed to reduce noise, as are those present in an overly high per- centage of documents, which may dilute topic specificity. In gensim, this is achieved using the filter extremes method, which allows specifying thresholds for minimum and maximum document frequency. For example, tokens occurring in fewer than 5 documents or more than 90% of the corpus can be filtered out. These thresholds were chosen based on the corpus size and domain specific requirements to ensure meaningful topics while maintaining computational efficiency.

5.2.4 Creating a Corpus

In constructing this bag-of-words corpus, each vector is initialized with a length corresponding to the total number of unique tokens in the dictionary, with each entry reflecting the count of occurrences of the associated word in the respective document. This method effectively disregards the order of words, focusing solely on the frequency and presence of terms, which simplifies the computational complexity of the analysis. By capturing the essence of each document in this vectorized form, the corpus lays the groundwork for the LDA model to uncover latent topics within the text, revealing underlying themes that may not be immediately apparent from casual observation.

5.2.5 Training LDA model

With the preprocessing steps completed and the corpus established, the next crucial phase is training the LDA model. This training process is based on the principle of maximizing the likelihood function, which is represented mathematically as $p(D|\alpha, \beta)$. This function reflects the probability of the observed documents D given the parameters α and β, which are Dirichlet

priors that influence the distribution of topics across documents and the distribution of words across topics, respectively. In essence, LDA aims to uncover the hidden topic structures that lie beneath the surface of the observed data, revealing the latent themes that characterize the collection of documents. The given equation represents the **objective function** for LDA. It defines the likelihood of the observed documents D given the parameters α and β.

$$p(D|\alpha,\beta) = \prod_{d=1}^{D} \int p(\theta_d|\alpha) \prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn},\beta) d\theta_d \text{ -}$$

where equation, D represents the total number of documents, N_d denotes the number of words contained in document d. , θ_d refers to the topic distribution for that specific document, and z_{dn} indicates the topic assigned to the n-th word within document d. w_{dn} represents the n-th word in document d. The process involves the selection of the topic distribution θ_d from a Dirichlet distribution defined by the hyperparameter α, followed by determining the topic z_{dn} for each word w_{dn} using a multinomial distribution based on the chosen topic. The generative model underlying LDA operates in a series of logical steps for each document within the corpus. Initially, for every document d, a topic distribution θ_d is sampled from a Dirichlet distribution parameterized by α. Subsequently, for each word w_{dn} within document d, a topic z_{dn} is selected from the multinomial distribution characterized by θ_d. Finally, a word w_{dn} is drawn from the conditional probability $p(w_{dn}|z_{dn},\beta)$, which represents the likelihood of the word given the selected topic. This structured process allows the model to iteratively refine its understanding of the topic distributions and word associations throughout the corpus. As training progresses, the model adjusts its parameters to maximize the likelihood of the observed data, ultimately resulting in a robust representation of the underlying topic structure within the documents. LDA is considered to have converged when the likelihood function stabilizes, indicating that further iterations produce minimal changes in the topic distributions. This state signifies that the model has effectively captured the latent patterns in the data.

5.2.6 Displaying and Visualizing Topics

After training the model, we can extract and display the top topics. Each topic is represented by a list of the most significant words associated with it. This step helps in understanding the primary themes and topics within the data.

Table 5.1-A: Top 10 Words for Each Topic

Topic	Top 10 Words
0	malignant, anaemia, asthma, agitation, test, arrhythmia, diabetes, mellitus, arthropod, shock
1	lung, reaction, drug, adverse, dyspnoea, mouth, intestinal, interstitial, eosinophilia, nasopharyngitis
2	pain, abdominal, distension, brain, chest, lower, growth, gallbladder, pancytopenia, femur
3	drug, ineffective, oesophageal, back, hypersensitivity, dependence, glucose, swelling, interaction, peripheral
4	product, failure, hepatic, administered, therapeutic, administration, atrial, bronchitis, effect, fibrillation
5	cancer, bladder, gastric, haemorrhage, acne, anal, cerebral, malaise, angina, incontinence
6	carcinoma, cell, skin, lymphoma, count, loss, infarction, aggression, foetal, myocardial
7	abdominal, upper, pain, death, abscess, oedema, protein, body, flushing, height
8	bone, infection, device, issue, product, dose, eye, omission, dry, error
9	acute, injury, kidney, leukaemia, arthropathy, rash, liver, adrenal, myeloid, insufficiency
10	arthralgia, exposure, diarrhoea, accidental, cough, pregnancy, amnesia, overdose, maternal, Crohn
11	stage, use, iv, syndrome, label, pneumonia, state, confusional, reaction, anaphylactic
12	abnormal, anxiety, asthenia, depression, behaviour, feeling, cellulitis, complication, weight, faeces

Table 5.1-B: Top 10 Words for Each Topic	
Topic	Top 10 Words
13	aggravated, condition, gastrointestinal, fall, indication, unapproved, respiratory, nausea, pulmonary, impairment
14	decreased, blood, increased, pressure, creatinine, fatigue, dizziness, aminotransferase, alanine, sensation
15	alopecia, cataract, colon, chills, thrombosis, vein, deep, arterial, mass, haematoma
16	breast, site, fracture, constipation, headache, injection, erythema, neutropenia, event, adverse
17	discomfort, abdominal, anhedonia, appetite, pruritus, blindness, herpes, anger, application, zoster
18	renal, disorder, arthritis, cardiac, dermatitis, colitis, balance, arrest, ascites, emotional
19	disease, chronic, kidney, progression, hepatitis, tract, influenza, autoimmune, ulcer, hallucination

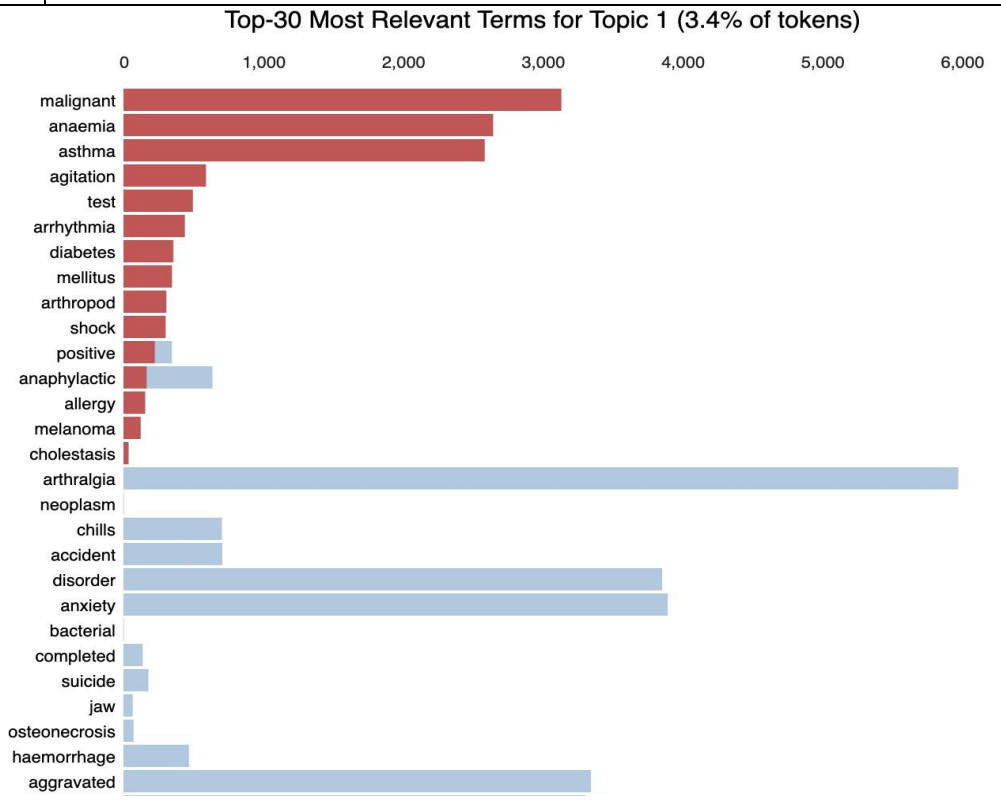


Figure 5.1: Top-30 Most Relevant Terms for Topic after LDA Implementation

Figure 5.1 (3.4% of tokens)" provides a detailed visualization of the most significant terms associated with Topic 0 as identified by the LDA model. The x-axis represents the frequency of each term, with a scale ranging approximately from 0 to 6,000 occurrences, while the y-axis lists the top 30 terms in descending order of their relevance to Topic 0. The chart distinguishes between terms with higher frequencies (shown in red) and those with comparatively lower fre- quencies (shown in blue), offering a clear view of their respective importance. Key terms such as "malignant," "anaemia," and "asthma" are among the most frequently occurring, each ap- proaching or exceeding 3,000 occurrences, indicating their significant relevance to Topic 0. Other notable terms include "agitation," "test," "arrhythmia," "diabetes," "mellitus," "arthro- pod," and "shock," all of which have substantial frequencies. Despite having lower frequencies, terms like "anaphylactic," "arthralgia," "disorder," and "aggravated" remain important within the context of Topic 0. The chart suggests that Topic 0 is heavily associated with medical conditions, diseases, and symptoms. The prominence of terms such as "malignant", "anemia", and "asthma" implies a focus on serious health conditions and their treatment. This visualization serves as a valuable tool for researchers and analysts, offering a quick grasp of the primary components and themes of Topic 0. It is particularly useful in identifying common terms that define the topic, aiding in further analysis or targeted investigation.

5.3 Chapter Conclusion

This chapter discussed the implementation of Latent Dirichlet Allocation (LDA) for topic modeling, highlighting the necessary preprocessing steps such as text cleaning, tokenization, dictionary creation, and corpus formation. It detailed how LDA identifies key topics based on relevant word groupings, enabling structured analysis of unstructured text. The use of visualizations improved interpretability, making it easier to extract meaningful insights. Overall, the chapter demonstrated how LDA can be effectively applied in natural language processing for automated text analysis and knowledge discovery.

Chapter Six

Learning Distributed Representation Of Drug Sequences From Adverse Event Reporting Data

6.1 Introduction

Adverse Drug Reaction (ADR) prediction is a critical task in pharmacovigilance, requiring robust modeling techniques to capture complex relationships between drug sequences and their associated reactions. In this chapter, we propose a novel approach that integrates bidirectional contextual embeddings of drug sequences with ADR representations derived from RoBERTa. By leveraging a transformer-based architecture alongside BiLSTM models, we aim to enhance the predictive accuracy of ADR occurrences. Our methodology incorporates a comprehensive feature representation strategy, utilizing CLS token embeddings and concatenated drug sequence encodings. The model is trained using binary cross-entropy loss to optimize prediction performance. Through rigorous experimental evaluation on the FAERS dataset, we compare different sequence encoding models to determine their effectiveness in capturing drug-reaction associations. This chapter details our dataset preparation, experimental setup, performance evaluation, and qualitative analysis, providing insights into how deep learning models can improve ADR prediction.

6.2 Methodology

in our research, we introduce a comprehensive end-to-end training framework that focuses on learning meaningful representations of drug sequences and predicting the likelihood that a specific ADR is associated with those sequences. Inspired by recent advances, including

the work of [77] on sequence reversal for improved attention alignment and the deep learning transformations proposed by [78], our model leverages these innovations to capture complex temporal and semantic dependencies in adverse event prediction to enhance predictive accuracy. By framing the task as a sequence labeling problem, we aim to label each drug sequence according to its potential to induce an ADR, thus facilitating a deeper understanding of the relationships between drug administration and subsequent adverse reactions. Following this, we employ specialized encoders to transform these drug sequences into numerical representations that capture the underlying patterns and relationships between different drugs. These representations are then subjected to machine learning techniques for analysis in order to determine the probability of an ADR occurrence. Throughout this section, we will detail the specific components of our model, the rationale behind the chosen methodologies and the empirical results that demonstrate the effectiveness of our approach in making accurate ADR predictions.

6.2.1 Notations and concepts

Before delving into the specifics of the methodology, it's important to define the notations used throughout this section. The data comes from the FAERS, which records cases involving drug sequences and their associated ADRs. We denote a sequence of drugs for a specific case as $S_i = (s_1, s_2, \ldots, s_n)$, where s_j represents the j-th drug in the sequence. Each drug sequence is time-stamped, reflecting the order and timing of drug administration. Associated with each sequence S_i is a list of unordered ADRs, represented as $R_i = \{r_1, r_2, \ldots, r_m\}$, where r_z refers to the z-th ADR in the list. A case in the FAERS dataset can be expressed as $c_i = \{S_i, R_i\}$.

Since multiple ADRs can correspond to a single drug sequence, we perform an unfolding operation to create new tuples that pair each drug sequence with a single ADR. These pairs are denoted as $c_{i,j,p} = \{S_i, r_{i,j}\}$, where $0 \leq j \leq m$. This transformation generates multiple positive examples for each case, allowing the model to learn which ADRs are likely to occur given the drug sequence. We also generate negative examples using hard negative mining, represented as $c_{i,j,n}$. These negative examples consist of pairs where the drug sequence and ADR are not associated. The model's primary objective is to maximize the likelihood of predicting the correct association between a drug sequence and an ADR, expressed mathematically as $P(y = 1 \mid S_i, r_z, \theta)$, where $P(y = 1 \mid S_i, r_z, \theta)$ indicates the probability of

observing the ADR r_z for the drug sequence S_i, given the model parameters θ. By maximizing this probability, the model learns to effectively distinguish between sequences that are likely to cause specific ADRs and those that are not, ultimately improving prediction accuracy and contributing to a better understanding of drug safety.

6.2.2 Hard Negative Sampling

Hard negative sampling plays a key role in the effective learning of representation vectors through a discrimination approach. Negative sampling is a technique that aims to enhance the model's performance by providing it with samples that do not correspond to the expected outcome. Specifically, negative samples are defined as reactions that do not occur alongside the associated drug sequence, and these samples are labeled as 0 to indicate their lack of relevance. However, not all negative samples are created equal; hard negative samples are particularly notable due to their challenging nature. These samples are characterized by their high similar- ity to true positive reactions making them more difficult for the model to distinguish. In fact, hard negative samples are defined as those that exhibit a cosine similarity of at least 0.65 with true reactions. By incorporating hard negative samples into the training process, the model is compelled to refine its ability to differentiate between true and false associations, ultimately leading to improved accuracy and robustness in the learned representations. This strategy not only enhances the model's discriminative power but also contributes to a more nuanced under- standing of the relationships between drug sequences and adverse reactions, thereby advancing our overall goals in predicting adverse drug reactions effectively.

6.2.3 Model Architecture

The model architecture aims to efficiently process two types of inputs: the medication sequence and the associated ADR. These inputs have significance to the model's overall operation as seen in chapter 4. Each type of input is routed through a specific encoder, which is in charge of creating intelligible representations of the data's key properties. To achieve optimal encoding of the drug sequences, we conducted experiments with various types of encoders, evaluating their performance and suitability for this specific task. Among the options explored, BiLSTM

and Transformer models were prominent candidates. The BiLSTM encoder processes the drug sequence in both forward and backward directions, effectively capturing contextual information from both the past and the future, which is particularly beneficial for sequential data. On the other hand, the Transformer model, known for its attention mechanisms, offers a different approach to encoding by allowing the model to weigh the significance of different parts of the input sequence dynamically.

The Drug Sequence Encoder is specifically designed to consider the sequence of drugs administered to a patient. Unlike traditional approaches that focus on a single drug and its as- sociated ADRs, our method takes into account the entire drug sequence. This is particularly important because the combination of multiple drugs taken simultaneously or in succession can lead to different ADRs than when a single drug is analyzed in isolation. By modeling the complete drug sequence, we aim to capture the complex interactions and potential synergies between various medications that may trigger specific ADRs.

To facilitate this, we treat the drug sequence similarly to a text string, with each drug name serving as a token similar to the methodologies employed in NLP. Within the BiLSTM frame- work, we employed two variations: a word-based BiLSTM and a character-based BiLSTM. Figure 5.1 shows a schematic of the word-based LSTM encoder inputs. While both models demonstrated comparable performance, we hypothesize that the character based BiLSTM exhibits superior generalization capabilities, especially for drugs whose names may be less frequent or completely absent from the dataset. This characteristic enhances the model's robust- ness allowing it to handle a wider variety of drug names and compositions effectively.

The encoding process begins with tokenizing the resulting drug sequence using a training vocabulary, followed by passing these tokens to a word or character embedding layer, where representation for each token is trained. These embeddings are then fed into a single layer BiLSTM, which encodes the sequence, producing hidden states that encapsulate the relevant information of the drug sequence. The representation derived from the last hidden layer of the BiLSTM serves as the final sequence representation. Moreover, the outputs generated at each time step are interpreted as the encodings of each drug within the sequence.

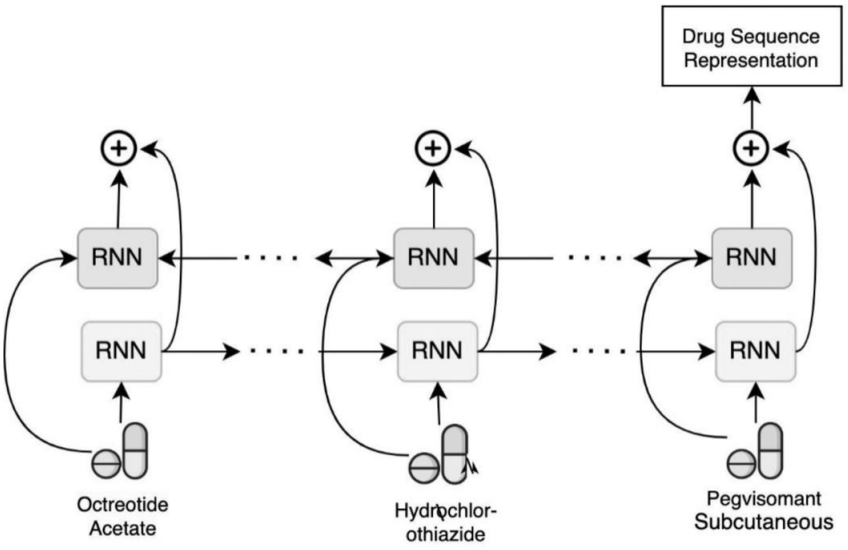


Figure 6.1: Bilstm Drug Sequence Encoder

In addition to the BiLSTM approach, we also explored Transformer encoder models, which have shown impressive performance in sequence modeling, particularly with text-based sequences. This success is largely attributed to their ability to generate contextual representations through self-attention mechanisms. At the end, we experimented with two prominent Trans- former architectures: a smaller model characterized by four layers and eight attention heads per layer, referred to as L4H8, and a larger model featuring twelve layers and sixty-four attention heads per layer, designated as L12H64 as discussed in our published paper.

6.2.4 Event Reaction Encoder

The encoding of ADRs presents unique challenges due to the way these reactions are typically documented. ADRs are often expressed in plain English, lacking a standardized format, and tend to be more descriptive in nature. For instance, an ADR might be reported as "high blood pressure" or as "the patient suffers from high blood pressure." This variability in expression underscores the need for sophisticated encoding methods that can effectively interpret and analyze such nuanced language. As a result, pre-trained language models are particularly well-suited for encoding ADR inputs, as they are designed to understand and process human language in its various forms.

Modern language models, such as RoBERTa, employ Byte Pair Encoding in their

tokenization processes. This feature enhances the model's robustness to different writing styles, grammatical errors, and even misspellings, which are common in real-world data. In our research, we use Hugging Face's pre-trained roberta-base model to encode ADRs effectively. To adapt the model to our specific dataset, we unfreeze only the top two layers of the model. This decision was guided by empirical experimentation, which showed that unfreezing additional layers offered minimal performance gains while increasing the risk of overfitting. This selective fine- tuning allows us to retain the foundational language understanding of the pre-trained model while effectively aligning the higher-level representations with the distribution of our target data. By focusing adaptation on the top layers, we enhance the accuracy and effectiveness of ADR encoding, resulting in more reliable predictions and analyses in our study .

6.2.5 Projection Layers

The projection layers effectively combine information from various sources to make informed predictions about the likelihood of ADRs associated with specific drug sequences. The process begins with the output generated from the last encoder state of the drug sequence encoder. For models based on Transformers, this output includes the CLS (classification) representation, while for Bidirectional Long Short-Term Memory (BiLSTM) models, it comprises the concatenation of the last forward and backward hidden states, denoted as $(h^{\rightarrow} \| h^{\leftarrow})$. This concatenation captures the contextual information from both directions of the sequence, providing a comprehensive representation of the drug sequence. To improve the robustness of our predictions, we also incorporate the CLS token representation from the ADR encoder, which is built upon the RoBERTa model. By concatenating this ADR representation with the previously mentioned drug sequence representations, we create a rich, unified feature vector that encapsulates both the drug and its potential associated ADRs. Once this concatenated vector is constructed, it is passed through two fully connected layers. These layers are crucial for learning complex patterns and relationships between the input features. Each layer transforms the input through linear combinations followed by nonlinear activation functions, allowing the model to capture intricate dependencies in the data.

After processing through these layers, the final output is produced through a softmax activation function. This step converts the model's raw output into probabilities, effectively

indicating the probability of the drug-reaction pair. To optimize the model during training, we compute the training loss using binary cross-entropy (BCE). This loss function is particularly suitable for binary classification tasks, as it quantifies the difference between the predicted probabilities and the actual labels. By minimizing this loss, the model adjusts its weights to improve its predictive performance ultimately targeting to enhance the accuracy of ADR predictions based on the drug sequences provided.

6.3 Experiments and Results

6.3.1 Dataset

We used the FARES dataset, which contains numerous cases involving drug sequences and their corresponding adverse drug reactions (ADRs). To ensure a comprehensive evaluation of our model's performance, we divided the dataset into three distinct subsets: the training set, which comprises 70 percent of the total cases; the validation set, representing 10 percent of the cases; and the test set, accounting for the remaining 20 percent. This stratified approach allows us to train the model effectively while also validating its performance on unseen data and assessing its generalizability. The training set is employed to optimize the model's parameters, while the validation set serves as a tool for hyperparameter tuning and model selection. Finally, the test set is reserved for the ultimate evaluation of the model's predictive capabilities. In Table 5.1, the lengths of drug sequences (both folded and unfolded) and the number of tokens per reaction are shown for both the training and validation sets. The folded sequences refer to instances where ADRs are grouped together, while unfolded sequences treat each ADR individually. In the training set, sequence lengths range from a minimum of 1 to a maximum of 21 (folded) and 18 (unfolded), with an average length of 17 and 11, respectively. Similarly, the validation set shows sequence lengths from 1 to 23 (folded) and from 1 to 16 (unfolded), with average lengths of 15 and 9. This variation in sequence length highlights the complexity of drug interactions and the different ADRs associated with them. Additionally, the number of tokens per reaction ranges from a minimum of 1 to a maximum of 20 in the training set, and a slightly higher average token count of 15 for the validation set.

Table 6.1: Statistics of Training and Validation Sets

	Training Set			Validation Set		
	Min.	**Avg.**	**Max.**	**Min.**	**Avg.**	**Max.**
Sequence length (folded)	1	17	21	1	15	23
Sequence length (unfolded)	1	11	18	1	9	16
Number of tokens per reactions	1	7	20	1	15	17

Table 5.2 then breaks down the sizes of the training, validation, and test sets, both before and after folding the ADRs. Before folding, the training set contains 600,000 cases, while after unfolding, this increases significantly to 1.87 million cases. The validation set shows similar growth, expanding from 160,000 to 490,000 cases after unfolding. The test set grows from 40,000 to 135,000 after the ADRs are unfolded. This significant increase in dataset size after unfolding illustrates how ADRs are not simply one-to-one with drug sequences; multiple ADRs can be associated with a single sequence, thereby increasing the overall complexity of the task.

Table 6.2: Training, Validation, and Test Set Sizes before and After Folding the ADRs

	Before ADR Folding	**After Unfolding**
Training set size	600K	1.87M
Validation set size	40K	135K
Test set size	160K	490K

6.4 Results

In this study, we evaluated several drug-sequence encoders to determine the most effective method for predicting adverse drug reactions. Four distinct models were tested: 1) a word-based Bidirectional Long Short-Term Memory (BiLSTM) network, 2) a character-based BiLSTM, 3) a transformer model with 4 layers and 8 attention heads (denoted as L4H8), and 4) a larger transformer model with 12 layers and 64 attention heads (L12H64).

As anticipated, the transformer models (L4H8 and L12H64) outperformed the BiLSTM models by a significant margin, with the L4H8 transformer showing a 7 percent improvement of F1 score over the BiLSTM baseline and the L12H64 model showing a 10 percent improvement. This superior performance can be attributed to the self-attention mechanism inherent in transformer models. Unlike BiLSTMs, which process sequences in a more rigid, stepwise manner, transformers can dynamically assign different levels of importance to various parts of the input sequence. The primary task was a binary classification problem: determining whether a given drug sequence is associated with a specific adverse reaction. For this, we calculated accuracy and F1-score, which are shown in Table 6.3. The results highlight that the character-

Table 6.3: Performance Metrics

Model	**Accuracy**	**F1**
Word-based BiLSTM (baseline)	0.80	0.74
Character-based BiLSTM	0.84	0.77
Transformer L4H8	0.91	0.84
Transformer L12H64	0.94	0.87

based BiLSTM outperformed the word-based BiLSTM by approximately 3 percent in terms of F1-score. This improvement suggests that character-level encodings provide a more granular understanding of the input data, which is particularly useful for handling drug names that may not appear frequently in the training set. However, even with this improvement, the character- based BiLSTM was still outpaced by the transformer models.

The smaller transformer model (L4H8) demonstrated a significant leap in performance, achieving an F1-score of 0.84, which is about 10 percent higher than the word-based BiLSTM. The best results, however, were achieved by the larger transformer model (L12H64), These findings are consistent with state-of-the-art results reported by [79, 86] , who demonstrated the effectiveness of sequence-to-sequence Transformers in ADE prediction tasks, which further boosted the F1-score to 0.87, an improvement of 12.5 percent over the baseline model. The ac- curacy of this model also reached 94 percent, indicating that it was not only able to predict the correct associations between drug sequences and adverse reactions but also did so consistently.

6.4.1 Qualitative Analysis of Drug-Sequence Representations

Beyond the quantitative metrics, we conducted a qualitative analysis to ensure that the model was not only accurate in its predictions but also learning meaningful representations of drug sequences. To validate this, we focused on visualizing how the model interprets and clusters drug sequences associated with different adverse reactions. Specifically, we randomly selected a sample of 20,000 drug sequences from the dataset which spanned four distinct adverse events: breast cancer, dizziness, physical disability, and hyperbilirubinemia. To better understand the model's internal representation of these sequences, we utilized t-distributed Stochastic Neighbor Embedding (t-SNE), a popular dimensionality reduction technique, to project the high- dimensional sequence embeddings into a two-dimensional space. This allowed us to visualize the relationships between different drug sequences as learned by the model. The resulting t-SNE plot is shown in Figure 6.2.

From the visualization, it is clear that drug sequence embeddings associated with the same adverse event are grouped together, forming distinct clusters. For example, sequences related to breast cancer form a cohesive cluster, while those associated with dizziness or hyperbilirubinemia are positioned in separate regions of the plot. This spatial arrangement indicates that the model successfully distinguishes between different types of drug reactions by learning representations that reflect their underlying relationships.

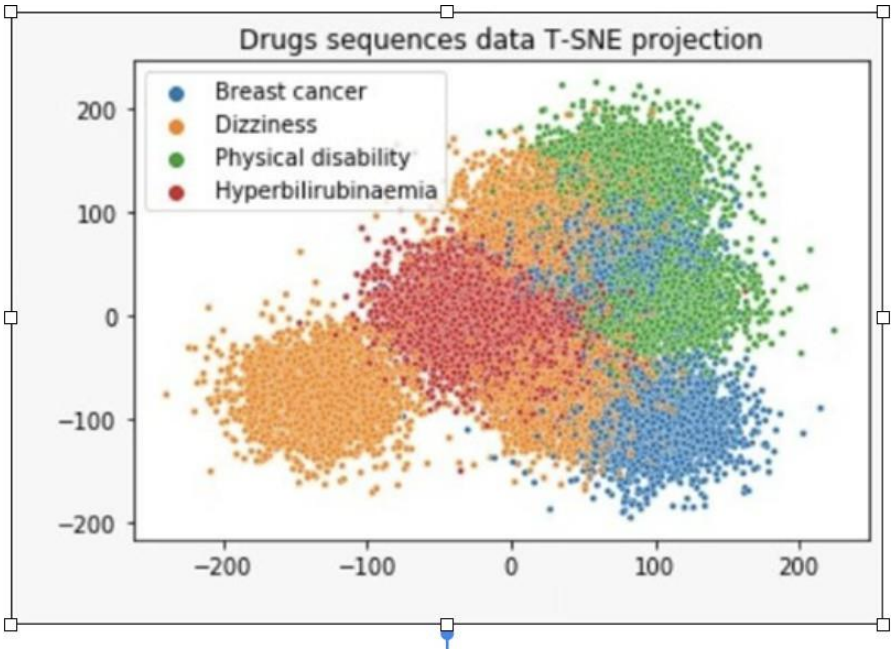


Figure 6.2: T-SNE visualization of 4 clusters of drug sequences sampled by their respective reactions.

Through this qualitative analysis, we can confirm that the model's internal representations align with the expected relationships between drugs and adverse reactions, providing additional confidence in its robustness and generalization capabilities. The use of t-SNE for visualizing embeddings offers a tangible way to assess how well the model captures the underlying structure of the data, complementing the quantitative results presented earlier.
In order to explore deeper into the learned representations of drug sequences, we focused on four specific clusters from the t-SNE plot shown in Figure 6.2. These clusters correspond to the following adverse drug reactions (ADRs): hyperbilirubinaemia, breast cancer, dizziness, and physical disability. From each of these four clusters, we selected three drug sequences for further analysis. To gain insights into how these sequences relate to each other, we generated a heatmap based on cosine similarity between the sequences, as presented in Figure 5. The heatmap allows us to visually assess the degree of similarity between different sequences within and across clusters, which further validates the model's ability to learn meaningful drug sequence representations. In the heatmap Figure 6.3, we denote the drug sequences sampled

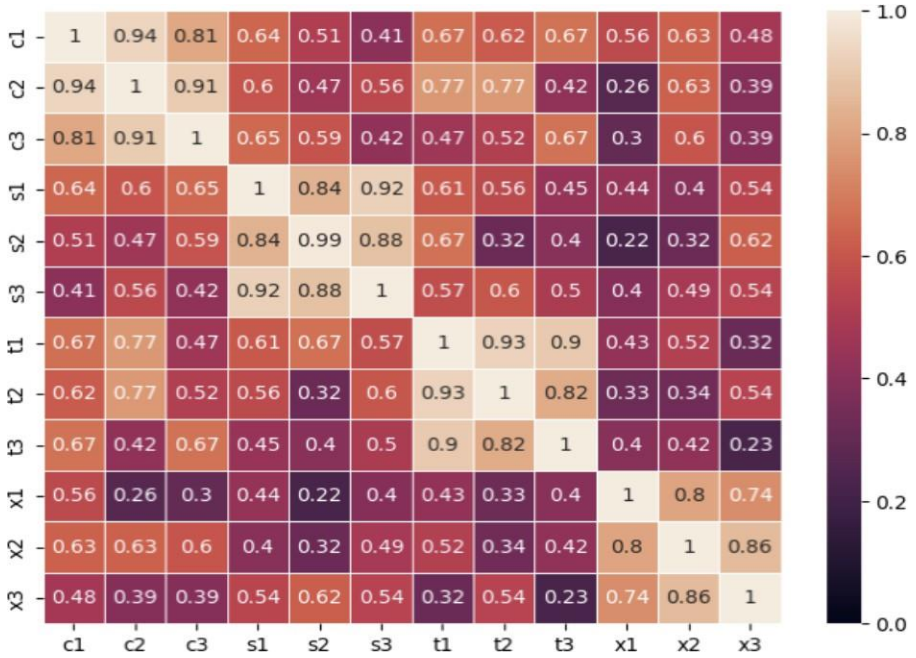


Figure 6.3: Cosine Similarity matrix between 3 random samples from each cluster.

from each cluster as follows: x_1, x_2, x_3 for the physical disability cluster, t_1, t_2, t_3 for the dizziness cluster, s_1, s_2, s_3 for the breast cancer cluster, and c_1, c_2, c_3 for the hyperbilirubinaemia cluster.

The color intensity in the heatmap represents the cosine similarity between the pairs of drug sequences, with darker shades indicating higher similarity. From the analysis of the heatmap, several clear patterns emerge. Sequences within the same cluster tend to exhibit high cosine similarity with each other, while sequences from different clusters show lower similarity. For instance, the sequences t_1, t_2, t_3 from the dizziness cluster have cosine similarities greater than 0.82, indicating that these sequences are closely related in terms of their learned embeddings. This suggests that the model effectively groups sequences with similar reactions together, al- lowing it to capture important patterns related to dizziness. Conversely, when comparing the dizziness cluster with sequences from the other three clusters, the cosine similarity drops significantly, ranging from 0.22 to 0.77. This demonstrates that the model has learned to differentiate between sequences that lead to distinct ADRs, such as dizziness and hyperbilirubinaemia. Sequences like c_1, c_2, c_3, which are associated with hyperbilirubinaemia, exhibit high similarity within their own cluster but remain distant from sequences in other clusters, reinforcing the model's ability to separate different drug-reaction relationships. To provide a clearer under- standing of the

sequences and reactions used in this analysis, Table 6.4 lists the drug sequences from each of the four clusters and their corresponding ADRs. It is important to note that these drug sequences can appear multiple times in the dataset with different reactions, but for the purposes of this analysis, we have randomly sampled one instance from each sequence.

6.5 Chapter Conclusion

In this chapter we propose a transformer-based model for predicting adverse drug reactions (ADRs) by learning contextual, distributed representations of drug sequences from drug sequence reaction data pairs. The model is trained to predict whether a given drug sequence is associated with a specific adverse reaction, enabling it to capture meaningful relationships between drugs and ADRs.

Evaluation results demonstrate the superiority of transformer encoders over BiLSTM-based architectures, with the best-performing model (L12H64) achieving an F1-score of 0.87 and an accuracy of 0.94. Qualitative analyses using t-SNE visualizations and cosine similarity heatmaps confirm that drug sequences associated with similar ADRs are clustered closely, while those with dissimilar ADRs are farther apart. This indicates the model's ability to learn discriminative and interpretable embeddings.

These findings highlight the potential of deep learning—particularly transformer models—in enhancing pharmacovigilance through more accurate ADR prediction, improved interpretability, and support for personalized medicine.

Table 6.4: Drug sequences from Figure 3 and their corresponding ADRs. Drug sequences can appear multiple times in the dataset with different reactions. The table below presents a random sample per each.

	Drug Sequence	**ADRs**
C1	GANCICLOVIR / ESMOLOL HYDROCHLORIDE / HYDRALAZINE / VALGANCICLOVIR / COTRIMOXAZOLE	Hyperbilirubinaemia, Hepatomegaly
C2	Olanzapine / ANAFRANIL / Centrum materna	Hyperbilirubinaemia neonatal, Respiratory distress, Feeding disor- der
C3	CAPECITABINE / CAPECITABINE QUINACRINE	Hyperbilirubinaemia, Oedema peripheral, Rash
S1	RANITIDINE HYDROCHLORIDE / RANITIDINE / ZANTAC	Breast cancer stage (III/I)
S2	XYREM / LISINOPRIL / COENZYME Q10 / LEVOTHYROXINE / ATENOLOL / NU- VIGIL / VITAMIN D3	Breast cancer carcinoma, phase Delayed sleep
S3	REVLIMID / PREDNISONE	Neutropenia, Influenza, Breast cancer
T1	Infliximab / INDOCID / MORPHINE NAPROXEN / VOLTAREN	Dizziness, Colitis ulcerative, Pyrexia
T2	CARBOPLATIN / TECENTRIQ / AVASTIN / PACLITAXEL	Malaise, Dizziness
T3	ULTOMIRIS / ULTOMIRIS	Infusion-related reaction, Dizziness
X1	RISPERDAL / TEMAZEPAM / CITALOPRAM / MIRTAZAPINE / SERTRALINE / CIPRALEX / TRAZODONE	Physical disability, Sedation, Somnolence
X2	ENTRESTO / LOSARTAN	Fatigue, Physical disability, Malaise, Dysp- noea
X3	XELJANZ XR / XELJANZ XR / XELJANZ XR	Exercise tolerance decreased, Stress, Abdominal discomfort, Physical disability

Chapter Seven

Sequence-to-Sequence Model Implementation from LDA analysis for ADE Prediction Using MedDRA and FAERS Data

7.1 Introduction

This chapter presents the training, evaluation, and optimization of our adverse drug events (ADEs) prediction model, with a focus on deep learning techniques, particularly sequence-to-sequence (Seq2Seq) architectures[77, 85]. We begin by outlining the training process, which relies on the backpropagation algorithm to minimize loss through iterative parameter updates governed by a learning rate. Evaluation metrics such as training and validation loss, F1-score, precision, recall, and accuracy are used to assess the model's predictive capabilities on a robust dataset comprising 1,000,000 training samples and 23,000 test samples.

We also explore techniques to improve model performance, including hyperparameter tuning, data augmentation, ensemble learning, and domain-specific pretraining. These efforts aim to enhance the model's generalizability and practical applicability in real-world clinical settings. In previous work, Current State of Machine Learning-Based Methods for Adverse Events Prediction, the studies applied traditional machine learning models such as Random Forest and logistic regression to structured data for AE detection. While effective, these methods were constrained by inconsistent and sparse data, particularly in regions like Palestine where healthcare systems are underdeveloped. This led to only moderate performance, with F1 scores

and AUC values limited by the quality of available data.

Our subsequent study, Learning Distributed Representation of Drug Sequences from Adverse Event Reporting, introduced deep learning models to encode drug sequences from FAERS data using transformer architectures. By treating AE prediction as a sequence-labeling task, we achieved significant performance gains, with transformer-based models reaching an F1 score of 0.87—surpassing other methods like BiLSTMs. We also conducted qualitative evaluations using t-SNE to validate the clustering of similar adverse drug reactions, revealing the model's ability to capture meaningful representations. However, challenges remained, including repetitive drug names and underutilized unstructured text in FAERS.

To address these issues, we implemented a Seq2Seq model that integrates pre-trained BERT embeddings with LSTM layers. This architecture enables the model to better capture sequential dependencies in drug administration and demographic features. The attention mechanism in Seq2Seq models enhances robustness to noisy and incomplete data by focusing on the most informative parts of the input sequence.

This approach builds directly on our prior work, aiming to refine AE and outcome prediction through multi-task learning [93, 95]. The integration of BERT[105] allows the model to lever- age rich contextual representations, while the LSTM decoder captures temporal dependencies in patient data. This combination is particularly effective for complex tasks like adverse event prediction, where both linguistic context and sequence order matter.

In this study, we applied the Seq2Seq model to FAERS and MedDRA datasets, predicting both adverse drug events and patient outcomes. The following sections detail the preprocessing steps, model architecture, training methodology, and performance evaluation, ensuring reproducibility and robustness. The model's design and workflow are illustrated in figure 7.1.

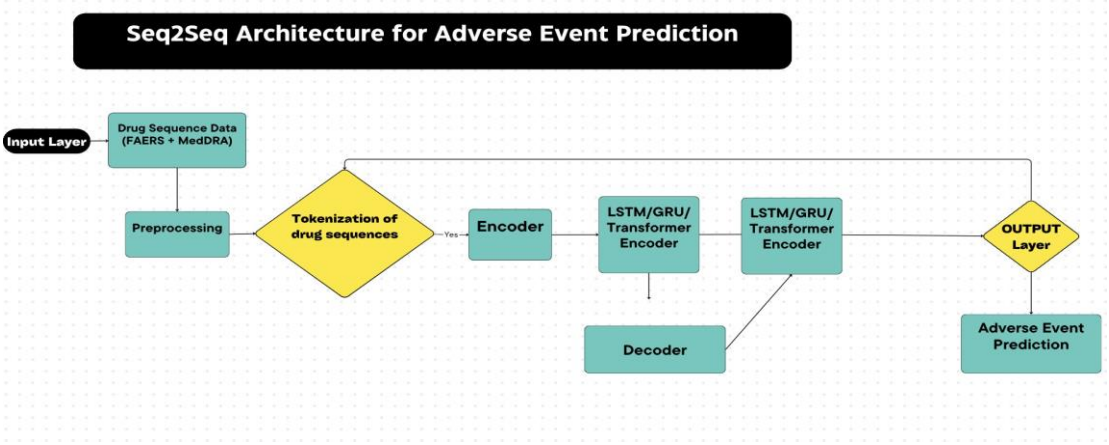


Figure 7.1: Seq2Seq Architecture Pipeline

7.1.1 Model Architecture Pipeline

The pipeline begins with the Input Layer, where drug sequence data from FAERS is augmented with MedDRA codes for corresponding adverse events. These sequences are passed through a Preprocessing Stage involving two essential tasks: tokenization and embedding. Tokenization converts the raw drug names and associated adverse events into a structured format of tokens, while embedding transforms these tokens into dense vectors, capturing the semantic relationships between drug sequences and adverse events. The embeddings serve as the input features for the subsequent model layers. The data enters the Encoder, typically implemented using LSTM based architectures. The encoder processes the drug sequence data and generates a series of hidden states, which are compressed representations of the input sequence. These hidden states encapsulate meaningful patterns about the relationship between drugs and adverse events. The output of the encoder, including the attention-weighted hidden states, is passed to the Decoder, which predicts the sequence of adverse events associated with the input drug. The Decoder generates this sequence step-by-step, informed by the encoded representation of the drug and the previously predicted adverse events. The Decoder can also incorporate autoregressive mechanisms, where each prediction informs subsequent predictions in the sequence.

The predicted adverse event sequence is then processed by the Output Layer, which maps the internal representation back to meaningful MedDRA terms. This stage is critical for ensuring the interpretability and clinical relevance of the model's output. Following this, the predicted adverse events undergo Postprocessing, which involves mapping the predicted sequence back to standardized MedDRA codes and further aligning the output with real-world clinical data.

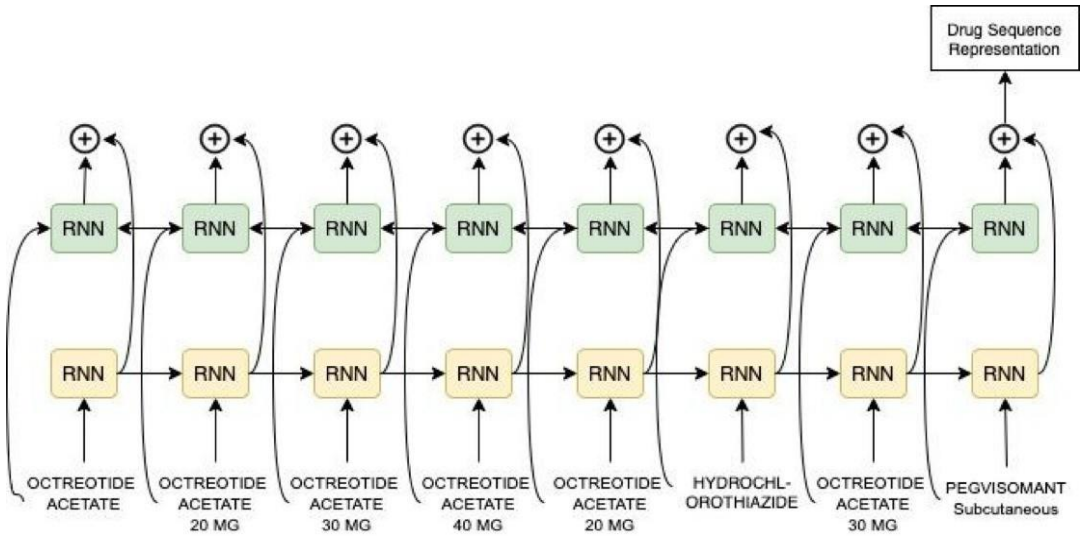


Figure 7.2: Drug Sequence Representation

7.2 Drug Sequence Representation

7.2.1 Structure Overview

- **RNN Layers:** Each block labeled "RNN" represents a recurrent neural network cell. The green and yellow blocks suggest that two different types of RNN cells might be used (potentially representing different hidden states or layers within a stacked RNN architecture).
- **Input Sequence:** The sequence at the bottom of the image represents the drug administrations over time. Each drug and its corresponding dosage is fed into the RNN cells sequentially. The drugs listed include:

– Octreotide Acetate at varying dosages (20 mg, 30 mg, 40 mg).
– Hydrochlorothiazide at 20 mg.
– Pegvisomant, administered subcutaneously.

- **Drug Sequence Representation:** At the top right, the output of the final RNN cell feeds into a process or module that generates a "Drug Sequence Representation." This representation likely summarizes the entire sequence of drug administrations, possibly encoding it into a fixed-length vector or a sequence of vectors.

7.2.2 Detailed Process

1. **Input to the RNN:**
   - Each drug and dosage (e.g., “Octreotide Acetate 20 mg”) is provided as an input to the RNN cells.
   - The RNN processes each drug sequentially, which means that the hidden state of the RNN at each time step depends not only on the current input but also on the previous hidden state, capturing the temporal dependencies between the drug administrations.
2. **Flow of Information:**
   - The output of each RNN cell at one time step is fed as input into the next RNN cell in the sequence (indicated by the horizontal arrows connecting the RNN cells).
   - Additionally, there are recurrent connections within each RNN cell (shown by the looping arrows), which represent the internal state updates that allow the RNN to maintain a memory of previous inputs.
3. **Combination of Outputs:**
   - The outputs from some of the RNN cells appear to be combined using summation operations (indicated by the plus signs in circles). This suggests that the network might be using some form of attention mechanism or weighted sum to focus on specific parts of the sequence.
4. **Final Output:**
   - The final RNN cell processes the last drug in the sequence, and its output is used to generate the “Drug Sequence Representation.” This output could be used for various downstream tasks such as predicting the next drug in the sequence, evaluating potential drug interactions, or summarizing the entire drug regimen.

7.2.2.1 Feature Embedding with BERT

To capture semantic information in drug descriptions, each tokenized input is embedded using the pre-trained BERT (bert-base-uncased) model. This results in high-dimensional vector representations::

$$\text{embedding} = \text{BERT}(\text{tokenizer}(\text{input\_feature}))$$

Features such as dosage, drug name, and administration route are passed through BERT to

generate embeddings that reflect contextual and syntactic relationships, enabling more informative model input. These embeddings are saved and reused during model training.

7.2.2.2 Sequence Formation

Each input consists of: Drug Name, Dosage, Route of Administration (i.e., the method by which a drug is delivered into the body, such as oral, intravenous, intramuscular, or topical) and Patient Demographics (e.g., age, sex, country)

To create input sequences, drug attributes are concatenated into a single string per record:

$$\text{train\_seq} = \langle \text{name} \rangle + \langle \text{dosage} \rangle + \langle \text{route} \rangle + \langle \text{end} \rangle$$

Each sequence was then tokenized and its length was calculated to maintain consistency during the training process.

7.2.2.3 Recurrent Neural Network (RNN) with LSTM

The core of our Seq2Seq model is a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) cells. LSTM networks are particularly effective at capturing long-term dependencies in sequences, making them suitable for modeling time series and sequences of arbitrary length.

The architecture of the LSTM model can be described as follows:

$$h_t = \text{LSTM}(x_t,\ h_{t-1},\ c_{t-1})$$

Where:

- x_t is the input at time step t,
- h_t is the hidden state at time step t,
- c_{t-1} is the cell state from the previous time step.

The model is bidirectional, meaning it processes the sequence in both forward and

backward directions, effectively capturing context from both past and future states.

7.2.2.4 Fully Connected Layer and Sigmoid Activation

After processing the sequence through the LSTM, the final hidden states are passed through a fully connected layer to map the LSTM outputs to the target space:

$$\text{output} = \sigma(W \cdot h_T + b)$$

Where:

- W and b are the weight matrix and bias, respectively,
- h_T is the final hidden state of the LSTM,
- σ is the sigmoid activation function used to squash the output to a probability.

7.2.3 Training Process

7.2.3.1 DataLoader and Training Loop

The training data was fed into the model using a PyTorch DataLoader, which handles batching and shuffling of the data. The input sequences were padded to ensure uniformity in batch processing.

The training loop iterates over batches of data, and the loss is computed using binary cross-entropy:

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

Where:

- y_i is the true label,
- $\hat{y}_i$ is the predicted probability.

7.2.4 Data Preprocessing

7.2.4.1 Reading and Cleaning the Dataset

All text inputs are normalized (converted to lowercase), missing values are removed, and unnec- essary columns are dropped. Relevant categorical columns are split and renamed to ensure clear feature distinction (e.g., splitting GENDER_COUNTRY_CODE_OCCP_COD into separate fields). To ensure data quality,The equation can summarize the transformation to lowercase:

$$\text{encoded\_value} = \text{str.lower}(\text{category\_value})$$

Data splitting achieved through a custom function that processed each entry, splitting it based on spaces and assigning the respective parts to the new columns:

$$\text{GENDER,COUNTRY\_CODE} = \text{split\_function}(\text{GENDER COUNTRY\_CODE OCCP\_COD})$$

The function can be mathematically represented as:

$$\text{split\_function}(x) = \begin{cases} (\text{None}, \text{None}) & \text{if } x \text{ is not a string} \\ (x_1, x_2) & \text{if } x = \text{split}(x, \text{' '}) \end{cases}$$

where x_1 and x_2 are the first and second elements after splitting the string x by spaces. Additionally, the file processed dates by splitting them into individual date components. This was necessary for accurately capturing the temporal aspects of the data. For example, the start_date column, which contained multiple dates separated by slashes, was split into a list of dates:

start_dates_split, end_dates_split = split_dates(start_date, end_date) This process can be represented by:

$$\text{split\_dates}(x, y) = (\text{split}(x, '/'), \text{split}(y, '/'))$$

These steps ensured that the text data was appropriately split and transformed into numerical representations that captured the semantic meaning of the original text, making it suitable for model training

7.2.4.2 Mapping MedDRA Ontologies

MedDRA (Medical Dictionary for Regulatory Activities) is used for encoding adverse reactions. The process involves:

1. **Reading MedDRA Encodings**: All MedDRA encodings are read and grouped to map each reaction to its respective top-level ontology.

2. **Aligning Reactions**: Reactions in the dataset are aligned to their respective MedDRA ontologies.
3. **Removing Empty Ontologies**: Rows with empty ontology mappings are discarded to ensure meaningful data.

This step aligns model outputs with internationally recognized medical standards.

7.3 Model Architecture

The architecture implemented for Adverse Drug Event (ADE) and outcome prediction using the combined MedDRA and FAERS dataset employs a sequence-to-sequence (Seq2Seq) model. Below is a detailed explanation of each component and its role in the architecture.

1. Input Layer

The input layer consists of various types of data:

- **Age/Age Group:**
  - **Representation:** The age or age group of the patient is represented as a vector of fixed size using an embedding layer.
  - **Mathematical Formulation:** Suppose the age vector is denoted as $\mathbf{x}_{\text{age}} \in \mathrm{R}^d$, where d is the embedding dimension.

- **Gender:**

– **Representation:** Gender is encoded in a similar fashion using an embedding layer.

– **Mathematical Formulation:** Let the gender vector be $\mathbf{x}_{\text{gender}} \in \mathrm{R}^{d}$.

- **Sequence of Drugs:**

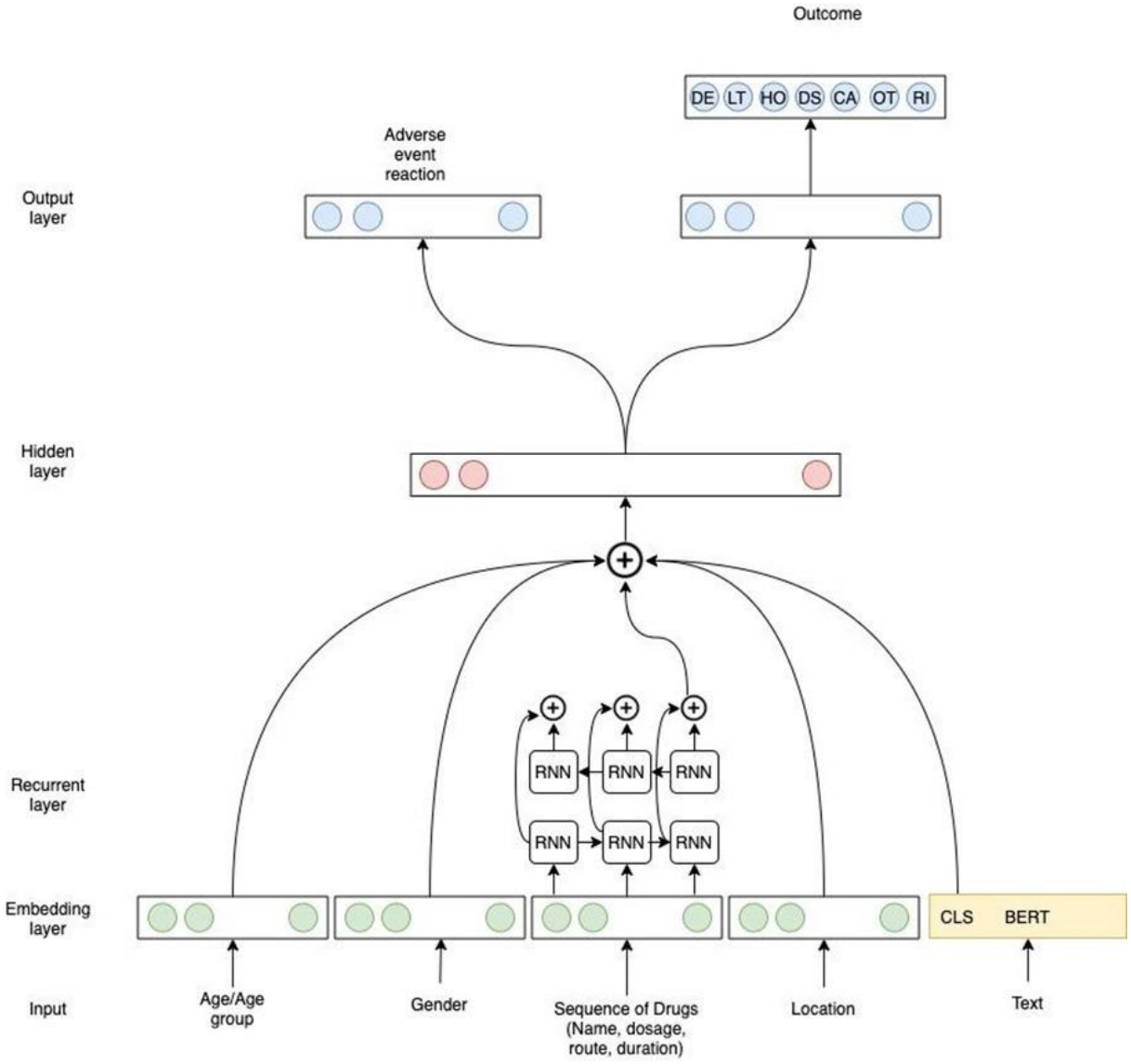


Figure 7.3: Model Architecture

– **Representation:** This includes the sequence of drugs a patient is taking, along with features such as name, dosage, route of administration, and duration.
– **Mathematical Formulation:** The drug sequence is modeled as a time series $\{\mathbf{x}_{\text{drug},t}\}_{t=1}^{T}$, where T is the length of the sequence and $\mathbf{x}_{\text{drug},t} \in \mathrm{R}^d$ is the embedding of the drug features at time step t.

• **Location:**

– **Representation:** The location is encoded similarly using embeddings.
– **Mathematical Formulation:** Denote this as $\mathbf{x}_{\text{location}} \in \mathrm{R}^d$.

• **Text:**

– **Representation:** Text data related to the patient's case (e.g., clinical notes) is pro- cessed using a pre-trained BERT model.
– **Mathematical Formulation:** The BERT model outputs a context-aware represen- tation $\mathbf{h}_{\text{text}} \in \mathrm{R}^d$.

The combined input representation $\mathbf{X}$ is formed by concatenating these features:

$$\mathbf{X} = [\mathbf{x}_{\text{age}}, \mathbf{x}_{\text{gender}}, \mathbf{x}_{\text{drug},1}, \ldots, \mathbf{x}_{\text{drug},T}, \mathbf{x}_{\text{location}}, \mathbf{h}_{\text{text}}]$$

2. Embedding Layer

• **Purpose:** Converts categorical variables (age, gender, drugs, etc.) into dense vectors of fixed size d, allowing them to be processed by subsequent layers.
• **Mathematical Operation:** Embedding is typically a learned matrix E where each categorical variable is mapped to a vector:

$$\text{Embedding}(i) = \mathbf{e}_i = E[i], \qquad E \in \mathrm{R}^{|V| \times d}$$

where $|V|$ is the vocabulary size.

3. Recurrent Layer

• **Purpose:** This layer processes the sequence of drugs using an RNN, capturing temporal dependencies.
• **Components:** The RNN here is represented as a stack of RNN cells (could be LSTM), processing the drug sequence.
• **Mathematical Operation:** At each time step t, the hidden state $\mathbf{h}_t$ is updated as follows:

$$\mathbf{h}_t = \sigma(W_h \mathbf{h}_{t-1} + W_x \mathbf{x}_{\text{drug},t} + \mathbf{b})$$

where W_h, W_x are weight matrices, $\mathbf{b}$ is a bias term, and σ is the activation function (usually tanh or ReLU).

4. Hidden Layer

- **Purpose:** Integrates the output from the RNN layer with other features (age, gender, location, and text). This layer combines the temporal features from the RNN with the static features.
- **Mathematical Operation:** The hidden state from the RNN is concatenated with the other embeddings:

$$\mathbf{h}_{\text{combined}} = \text{ReLU}(W_h \mathbf{h}_T + W_{\text{age}} \mathbf{x}_{\text{age}} + W_{\text{gender}} \mathbf{x}_{\text{gender}} + W_{\text{location}} \mathbf{x}_{\text{location}} + W_{\text{text}} \mathbf{h}_{\text{text}} + \mathbf{b})$$

Here, $W_h, W_{\text{age}}, W_{\text{gender}}, W_{\text{location}}, W_{\text{text}}$ are weight matrices for the respective inputs.

5. Output Layer

- **Purpose:** The final layer predicts the adverse event reaction based on the combined hidden representation.
- **Components:** This layer likely uses a softmax or sigmoid activation depending on whether the task is multi-class or multi-label classification.
- **Mathematical Operation:** The output layer computes:

$$\hat{\mathbf{y}} = \text{Softmax}(W_o \mathbf{h}_{\text{combined}} + \mathbf{b}_o))$$

where W_o is the output weight matrix, $\mathbf{b}_o$ is the bias term, and $\hat{\mathbf{y}}$ is the predicted probability distribution over possible adverse events.

6. Outcome

- **Representation:** The output is a probability distribution over various outcomes, like death (DE), life-threatening events (LT), hospitalization (HO), disability (DS), congenital anomaly (CA), and others (OT) and adverse event such as dizziness and fatigue.

- **Interpretation:** The model outputs a probability distribution indicating the likelihood of each adverse event, standardized using MedDRA terminology, based on the input features. When two or more adverse events exhibit identical or very close probabilities, the system relies on the hierarchical structure of MedDRA to prioritize events by their clinical relevance or higher-level grouping. If no explicit prioritization mechanism is applied, tied probabilities are presented as equally likely, allowing clinicians to make informed decisions based on detailed MedDRA classifications.

Background and Role of Layers

- **Embedding Layer:** Used to map categorical variables into dense vector spaces, allowing them to interact in a continuous domain where distance and direction have meaning.
- **Recurrent Layer:** The RNN is particularly suitable for sequential data like drug administration over time, as it maintains a hidden state that captures past information.
- **Hidden Layer:** Combines features from different sources (sequential and static) and serves as the integration point before making a final prediction.
- **Output Layer:** The softmax or sigmoid function at this stage is crucial for classification tasks, converting the raw model outputs into probabilities.

Mathematical Foundation

- **Embedding Function:** Converts a categorical index i to a dense vector:

$$\mathbf{e}_i = E[i]$$

- **RNN Dynamics:** The RNN layer evolves the hidden state $\mathbf{h}_t$ based on previous states and current input:

$$\mathbf{h}_t = \sigma(W_h \mathbf{h}_{t-1} + W_x \mathbf{x}_{\text{drug},t})$$

- **Final Prediction:** The combined hidden state is passed through a linear transformation and softmax to yield probabilities:

$$\hat{\mathbf{y}} = \text{Softmax}(W_o \mathbf{h}_{\text{combined}})$$

7.3.1 Sequence Modeling with Bidirectional LSTM

The LSTM was used to encode sequences derived from the dataset. The LSTM model was defined with an embedding layer, an LSTM layer, and a fully connected layer. The embedding layer converted input tokens into dense vectors, which were then processed by the LSTM layer to capture temporal patterns [82, 90]. The model architecture can be summarized by the following equations:

Embedding Layer:

$$\mathbf{E} = \text{Embedding}(\mathbf{x})$$

where $\mathbf{x}$ represents the input tokens and $\mathbf{E}$ represents the embedded vectors.

LSTM Layer:

$$\mathbf{h}_t, \mathbf{c}_t = \text{LSTM}(\mathbf{E}, \mathbf{h}_{t-1}, \mathbf{c}_{t-1})$$

where $\mathbf{h}_t$ and $\mathbf{c}_t$ are the hidden state and cell state at time step t, respectively.

Fully Connected Layer:

$$\mathbf{y} = \text{Sigmoid}(\mathbf{W} \cdot \mathbf{h}_t + \mathbf{b})$$

where $\mathbf{W}$ and $\mathbf{b}$ are the weights and biases of the fully connected layer, and $\mathbf{y}$ is the output of the model.

The implementation involved defining the LSTM model and then processing the input se- quences through the model to obtain the encoded representations. This can be expressed as:

$$\text{encoded\_sequences} = \text{LSTM\_Model}(\text{input\_sequences})$$

In practice, the sequences were tokenized, padded to ensure uniform length, and then passed through the embedding and LSTM layers. The final hidden states from the LSTM layer captured the essential features of the sequences, which were then used for further tasks such as classification or prediction.

This approach leverages LSTMs' ability to maintain information across long sequences, making it effective for encoding sequences in the dataset. Other machine learning models can then utilize the encoded sequences for tasks such as adverse event prediction, as demonstrated throughout the course of this research."

7.3.1.1 BERT for Sequence Classification

The BERT model is employed for the final classification task. The input features for BERT include:

- Age and gender of patients
- Output from the LSTM (drug sequence encoding)
- Location and textual descriptions

A pre-trained BERT model (*bert-base-uncased*) is fine-tuned incrementally on the dataset. Due to memory constraints, the model is trained on batches of 5000 samples, iteratively processing a total of 100,000 samples. The model's hyperparameters are as follows: The table 7.1

Table 7.1: Training Parameters

Parameter	**Value**
Batch Size	32
Epochs	10
Optimizer	Adam
Learning Rate	5e-5

provides an overview of the key parameters used during the model training process. It includes the batch size, which was set to 32, indicating the number of samples processed before the model's internal parameters are updated. The training was conducted over 10 epochs, meaning the entire dataset was passed through the model 10 times. The Adam optimizer [97] was used for updating the model weights, with a learning rate of 5×10^{-5}, controlling the step size during the optimization process.

7.3.2 Model Training

The training process involves fine-tuning the pre-trained BERT model. The loss function used is cross-entropy, suitable for multi-class classification tasks. The model is trained incrementally to avoid memory overflow issues, ensuring that the entire dataset is utilized effectively.

2. Embedding Generation Using BERT

The text data was converted into embeddings using the BERT model.

4. Training Loop

The training loop iterates over the data, where the model makes predictions, calculates loss, and updates its weights. The following steps were followed:

Forward Pass: The model processes input data and generates predictions.

$$\hat{y} = f(\mathbf{x}; \theta)$$

Loss Calculation: The loss function measures the difference between the predicted output and the true output.

Cross-Entropy Loss

$$\text{Cross-Entropy Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]\,.$$

Backpropagation: The gradients of the loss function with respect to the model parameters are calculated.

$$\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}\,.$$

Parameter Update: The model parameters are updated to minimize the loss.

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\theta_t)$$

7.3.3 Model Validation

After training, the model's performance is evaluated using various metrics. The accuracy, precision, recall, and F1 score are calculated to assess the model's effectiveness:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

The evaluation is conducted on a separate test set comprising 23,000 samples. These are the parameters used in the training process The table 7.2 summarizes the key metrics that evaluate

Table 7.2: Model Performance Metrics

Metrics	**Values**
Training Loss	0.05259
Validation Loss	0.12198
F1-Score	92.55%
Precision	93.22%
Recall	91.89%
Accuracy	92.81%
Training Samples	100000
Test Samples	23000

the performance of the trained model. The training loss and validation loss values, 0.05259 and 0.12198, respectively, indicate the error rates on the training and validation datasets. The F1-score, precision, recall, and accuracy are reported as 92.55%, 93.22%, 91.89%, and

92.81%, respectively, reflecting the model's effectiveness in making correct predictions.

7.3.3.1 Pseudo Code

1. Convert the dataset to a dataframe.
2. Use MedDRA ontologies to map the data.
3. Create embeddings for name, dosage, route, and duration (start and end dates).
4. Feed embeddings to an LSTM.
5. Use LSTM output along with other variables as input to BERT.
6. Predict the sequence of adverse reactions and outcomes as output.

This methodology outlines a comprehensive approach to adverse event prediction using advanced machine learning techniques. By leveraging the strengths of BERT and LSTM, com- bined with meticulous data preprocessing and strategic model enhancements, a robust and ef- fective predictive model is developed. The iterative refinement process ensures continuous improvement, ultimately contributing to enhanced pharmacovigilance and patient safety.

7.3.4 Training and Test Samples

The dataset used for training and evaluation consisted of a substantial number of samples to ensure the model's robustness and reliability:

- **Training Samples:** 1,000,000
- **Test Samples:** 23,000

The large training set allowed the model to learn diverse patterns and relationships within the data, contributing to its high performance. The test set, though smaller, was sufficient to validate the model's predictive capabilities and generalize to new, unseen data. The training and validation loss values indicated effective learning and generalization, making the model a reliable tool for predicting adverse reactions.

The model's results demonstrate its effectiveness and readiness for deployment in real-world scenarios where accurate classification of adverse reactions is critical. Future work may focus on further reducing the validation loss and improving the recall without sacrificing precision, thereby enhancing the model's overall utility and performance.

7.3.5 Further Model Improvements

To enhance the BERT model's performance for MedDRA sequence-to-sequence multi-class classification, the following strategies are employed:

7.3.5.1 Hyperparameter Tuning

Different hyperparameters such as learning rate, batch size, and dropout rate are experimented with using techniques like grid search or random search to identify optimal configurations.

7.3.5.2 Data Augmentation

Techniques like synonym replacement, token insertion, deletion, and noise addition introduce variations in the training data to enhance the model's generalization capability.

7.3.5.3 Ensemble Learning

Multiple BERT models are trained with different initialization and hyperparameters. Their predictions are combined using methods like averaging or stacking, which helps mitigate over- fitting and improve overall performance [81].

7.3.5.4 Domain-Specific Pretraining

Pretraining BERT on domain-specific corpora related to medical or pharmacological texts be- fore fine-tuning for MedDRA classification enables the model to capture domain-specific knowledge and terminology more effectively.

7.3.5.5 Regularization Techniques

Regularization methods such as L1 or L2 regularization, dropout, and batch normalization are applied during training to prevent overfitting and improve generalization.

7.3.5.6 Domain-Specific Evaluation Metrics

Evaluation metrics tailored to the significance of MedDRA classification in medical applications are defined and utilized to provide more meaningful insights into the model's performance.

7.3.6 Model Architecture Visualization and Pseudo Code

The model follows an encoder-decoder based architecture for sequence-to-sequence multi-class classification. The input to the encoder is raw data with MedDRA ontologies, which produces encodings. These encodings are then fed into a decoder model that predicts the sequence of adverse reactions and outcomes.

7.4 Chapter Conclusion

This chapter has provided an in-depth analysis of our model's training and evaluation processes, demonstrating its strong performance in predicting adverse drug reactions. By leveraging back- propagation for parameter updates, our model achieved a low training loss of 0.05259 and a validation loss of 0.12198, accompanied by high precision (93.22%), recall (91.89%), accuracy (92.81%), and an F1-score of 92.55%. The comprehensive evaluation, including the confusion matrix analysis, underscores the model's balanced ability to correctly classify positive and negative cases, thereby validating its robustness and reliability. Furthermore, the discussion on potential improvements—such as enhanced hyperparameter tuning, innovative data augmentation techniques, ensemble methods, and domain-specific pretraining—highlights promising avenues for further enhancing model performance. Overall, the methodologies and results presented in this chapter establish a solid foundation for deploying the model in real-world pharmacovigilance applications, ultimately contributing to improved patient safety and more effective adverse event monitoring.

Chapter Eight

A Graph-Based Approach to Drug Safety Monitoring Using FAERS and MedDRA

8.1 Introduction

In this chapter, we present a comprehensive evaluation of our adverse reaction classification model across ten reaction categories that we generated after the implementation of clustering, alongside the development and implementation of a GNN for adverse event prediction. Initially, we analyze the performance accuracy of the classification model on the FAERS data, where reaction categories such as Agranulocytosis, Gastrointestinal Disorders, Musculoskeletal Disorders, and Neurological Conditions demonstrate significant challenges, exhibiting zero precision and recall. In contrast, categories like Cardiac Disorders, Skin Reactions, and Respiratory issues show comparatively better performance, though with noticeable trade-offs between precision and recall. The overall GNN model accuracy stands at 52%, with macro-averaged metrics revealing low performance across less frequent classes, and weighted averages reflecting improved performance for dominant classes. Such results highlight a common challenge in multi-class classification problems with imbalanced datasets, especially when using graph neural networks that rely on node features and graph structure, which might not sufficiently differentiate minority classes.

Building on these insights, the chapter then introduces our implementation of a GNN tailored for adverse event prediction. This section details the data preprocessing steps—including sequence tokenization, label encoding, and graph representation—designed to capture the com- plex interdependencies among drugs, roles, and routes of administration. We further describe the GNN model architecture, which leverages Graph Convolutional Networks

(GCNs) within the PyTorch Geometric framework, and outline the training process, including visualization of training progress and loss reduction over epochs. Together, these approaches aim to enhance predictive capabilities and provide understanding of the underlying data relationships critical for pharmacovigilance. Previously, we performed LDA to restrict the output classes. In this analysis, clustering was performed on a data set of adverse reactions using Word2Vec embeddings and the K-means clustering algorithm. The objective was to reduce the classes by grouping similar adverse reactions into groups based on their semantic similarity, captured by word embeddings.

8.1.1 Word2Vec Model Training

The Word2Vec model was used to learn the semantic relationships between different adverse reactions. This neural network-based technique captures semantic similarities by placing words with similar contexts close together in a continuous vector space. The reactions, tokenized into individual components, served as input for Word2Vec, which analyzes co-occurrence patterns using either the Continuous Bag of Words (CBOW) approach to generate dense vector embed- dings. These embeddings effectively represent the underlying relationships between adverse reactions, enhancing the model's ability to understand and process the data.

The mathematical notation of Word2Vec can be expressed as: Where:

- w_o is the output word
- w_i is the input word
- $\mathbf{v}_w$ represents the vector representation of a word
- $\mathbf{v}'_w$ represents the output vector representation

8.1.2 Document Embedding Generation

To represent entire reaction descriptions, a sentence embedding technique based on averaging word vectors was employed. For each reaction description R containing n words, the sentence embedding $E(R)$ was calculated as:

Where:

- $\mathbf{v}_{wi}$ is the vector embedding of the i-th word
- n is the number of words in the reaction description

8.1.3 Clustering Analysis

The K-means algorithm was applied to the sentence embeddings to identify natural groupings of adverse reactions. The approach minimizes the within-cluster sum of squares (WCSS):

Where:

- $\mathbf{x}_i$ represents the i-th data point
- μ_k represents the centroid of cluster k
- C_k represents the k-th cluster

The key parameters used for the clustering of K-means were:

- Number of Clusters (k): 10
- Random State: 42 (for reproducibility)

8.2 Results and Interpretation

8.2.1 Top Reactions per Cluster

Representative reactions for each cluster were identified by analyzing the frequency of adverse reactions within each group, which are shown in Table 8.1

The clustering results reveal meaningful groupings of adverse reactions. Each cluster represents a set of reactions that share semantic similarities, as captured by the Word2Vec embed- dings. For example:

- Cluster 0 groups symptoms such as abdominal issues, indicating possible gastrointestinal or systemic conditions.
- Cluster 3 groups general symptoms like ineffective drug response and rash, hinting at issues related to treatment efficacy and dermatological reactions.

Table 8.1: Cluster Descriptions with Representative Reactions

Cluster ID	Representative Reactions	Description
0	Alopecia, Abdominal pain, Anaemia	Reactions primarily related to physical symp- toms and blood disorders.
1	Headache, Cough, Drug hy- persensitivity	Common reactions involving discomfort and al- lergic responses.
2	Syncope, Atrial fibrillation, Thrombocytopenia	Severe cardiovascular and blood-related reac- tions.
3	Drug ineffective, Diarrhoea, Rash	Reactions indicating ineffectiveness and side effects such as gastrointestinal and dermatolog- ical issues.
4	Pneumonia, Hypertension, Pruritus	Respiratory issues, blood pressure abnormali-ties, and skin reactions.
5	Pain, Dyspnoea, Insomnia	Reactions related to general discomfort, breath- ing issues, and sleep disorders.
6	Gynaecomastia, Emotional disorder	Hormonal and psychological effects.
7	Anxiety, Asthenia, Arthralgia	Mental health and musculoskeletal-related reac- tions.
8	Weight increased, Gynaeco- mastia, Emotional disorder	Physiological and emotional responses, often linked to metabolic or hormonal issues.
9	Agranulocytosis, Stevens-Johnson syndrome, Mediasti- nal disorder	Severe reactions involving blood disorders, skin conditions, and thoracic issues.

- Cluster 9 contains severe conditions like Stevens-Johnson syndrome, reflecting rare but serious adverse events.

8.3 Sequence-to-Sequence Model Implementation from Clustering analysis

In our previous Seq2Seq model with LDA, there was only one output class: Reactions. We have now extended the architecture to include an additional output class: Term Type, which corresponds to a higher level in the MedDRA hierarchy. This results in a multi-task model that simultaneously predicts both Reactions (a 10-class problem) and Term Type (a binary classification problem). The table 8.2presents the classification performance for Term Type, distinguishing

Table 8.2: Term Type Classification Report

	Precision	**Recall**	**F1-score**	**Support**
LLT	0.00	0.00	0.00	26,966
PT	0.56	1.00	0.72	33,933
Accuracy			0.56	60,899
Macro avg	0.28	0.50	0.36	60,899
Weighted avg	0.31	0.56	0.40	60,899

between LLT and PT. The model shows poor performance on the LLT class, with precision, recall, and F1-score all at 0.00, indicating it fails to correctly classify any LLT instances. In contrast, for the PT class, the model achieves perfect recall (1.00)—successfully identifying all PT instances—but with moderate precision (0.56), resulting in an F1-score of 0.72. The overall accuracy for this binary classification is 56%, with a macro average (precision = 0.28, recall = 0.50, F1-score = 0.36) that underscores the imbalance in performance between the two classes, and a weighted average (precision = 0.31, recall = 0.56, F1-score = 0.40) that reflects better results for the more prevalent PT class. Separately, the FAERS Reaction Classification Report evaluates the model's performance on the ten reaction categories. The results show a wide dis- parity across categories. Some reactions, such as Agranulocytosis, Gastrointestinal Disorders, Musculoskeletal Disorders, and Neurological Conditions, perform poorly with 0.00 precision and recall, indicating the model fails to classify any instances of these reactions. In contrast, categories like Cardiac Disorders, Skin Reactions, and Respiratory Issues show relatively better results. For instance, Cardiac Disorders achieve high precision (0.94), but with low recall, sug- gesting the model correctly classifies only a limited number of actual instances. The Infections category shows perfect precision but very

low recall, again highlighting the precision-recall trade-off. The overall accuracy for this task is 52%, with a macro average (precision = 0.41, recall = 0.20, F1-score = 0.17) indicating poor performance across all classes, and a weighted average (precision = 0.55, recall = 0.52, F1-score = 0.42) pointing to better performance on more frequently occurring reaction categories like Respiratory Issues and Skin Reactions.

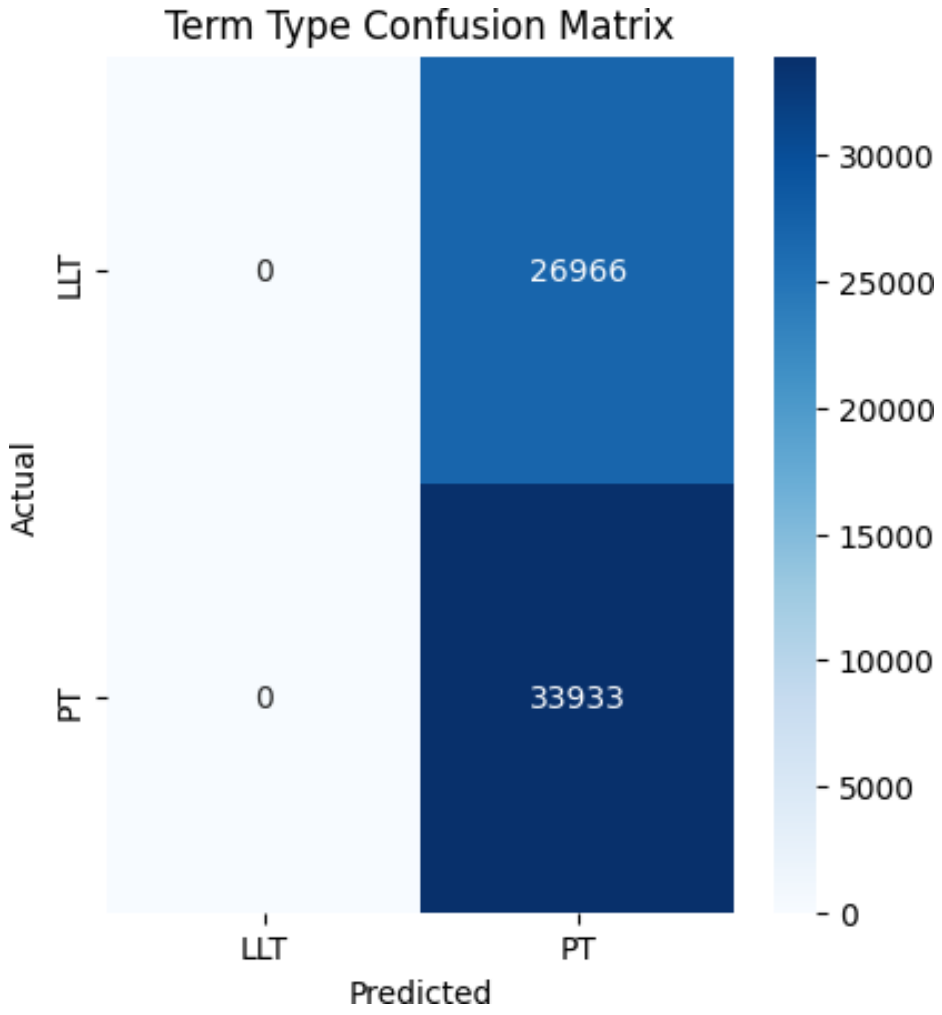


Figure 8.1: Term Type Confusion Matrix

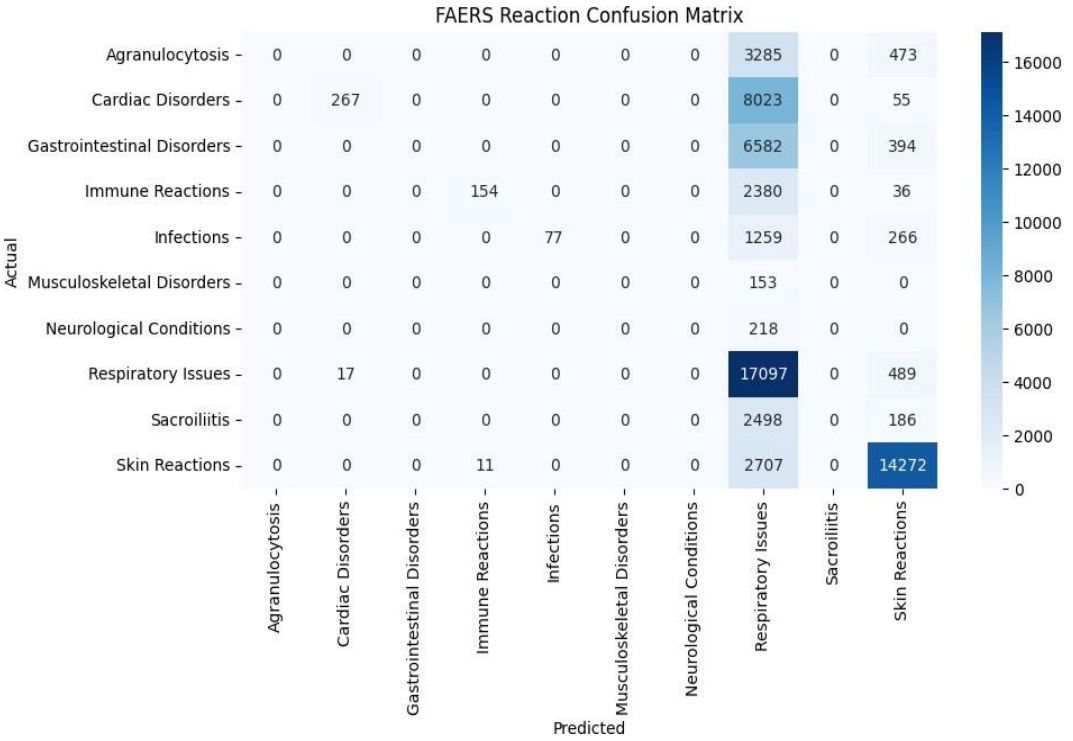


Figure 8.2: FAERS Reaction Confusion Matrix

Table 8.3: FAERS Reaction Classification Report

Reaction	**Precision**	**Recall**	**F1-score**	**Support**
Agranulocytosis	0.00	0.00	0.00	3,758
Cardiac Disorders	0.94	0.03	0.06	8,345
Gastrointestinal Disorders	0.00	0.00	0.00	6,976
Immune Reactions	0.93	0.06	0.11	2,570
Infections	1.00	0.05	0.09	1,602
Musculoskeletal Disorders	0.00	0.00	0.00	153
Neurological Conditions	0.00	0.00	0.00	218
Respiratory Issues	0.39	0.97	0.55	17,603
Sacroiliitis	0.00	0.00	0.00	2,684
Skin Reactions	0.88	0.84	0.86	16,990
Accuracy			0.52	60,899
Macro avg	0.41	0.20	0.17	60,899
Weighted avg	0.55	0.52	0.42	60,899

8.4 Implementation of a Graph Neural Network for Adverse Event Prediction

GNNs are mostly used for the implementation of graph data, offering novel ways to capture relationships between entities [110] . This chapter presents the implementation of a GNN to predict adverse reactions using the FAERS and MedDRA datasets. The primary objective is to predict two outputs, term type (binary classification) and FAERS reaction (multi-class classification) by modeling the interactions between input features such as drugs, roles, and routes of administration.

8.4.1 Data Prepossessing

The input dataset consisted of columns representing drugs, roles, and routes of administration, alongside target columns for *Term_Type* and *FAERS_Reaction*. To prepare the data for GNN modeling: Input sequences (e.g., *Aligned_drugs*) were split into tokens. Each token represented a unique entity, such as a drug or role, and was subsequently encoded into a numeric ID.

Target columns were label-encoded. *Term_Type* was encoded as binary, and *FAERS_Reaction* was multi-class encoded. Each data row was represented as a graph where:

- **Nodes**: Corresponded to unique tokens across drugs, roles, and routes.
- **Edges**: Defined as all possible pairwise connections between nodes in the same row, forming a fully connected graph.

8.5 Analysis of Visualizations and Model Training Results

8.5.1 Graph Analysis

The provided visualizations offer insights into the relationships within a single instance of the dataset. These graphs are instrumental in exploring data interdependencies and highlight the associations between drugs, reactions, routes, and their classifications.

8.5.1.1 Visualization of a Single Instance

This graph 8.3 illustrates the entity relationships for a single instance involving drug

administration and its potential effects. Nodes represent key entities such as drugs, immune reactions, roles, routes, and terminologies (LLT), while directed edges indicate specific relationships like "causes", "affects", and "contains". For example, an edge from a drug to an immune reaction indicates that the drug may cause that reaction. This graph structure helps visualize how a drug interacts with various biological and administrative entities, revealing how it might trigger adverse effects, depend on its route of administration, or relate to medical terminology classifications.

Key Components and Relationships

The node **LLT** (Low-Level Term) acts as a central hub, connecting to:

- **Reactions** (e.g., *Immune Reactions*)

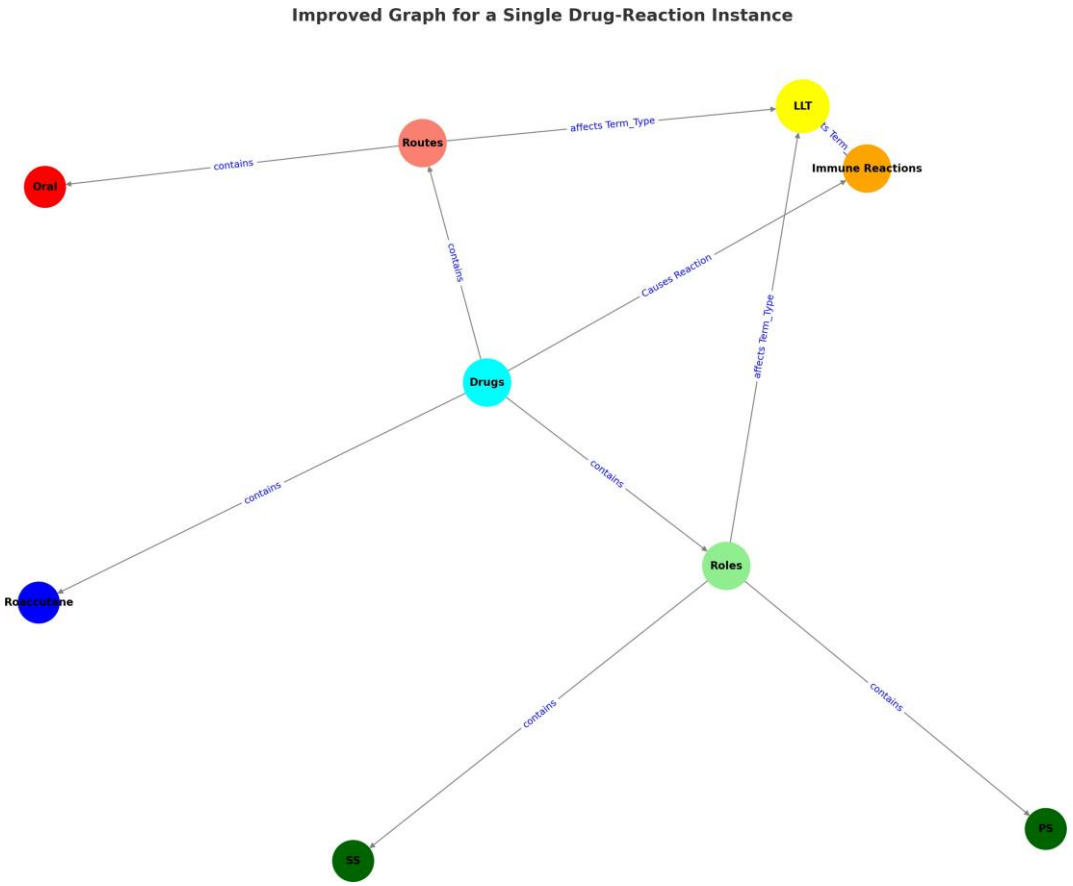


Figure 8.3: Grouped Graph for Single Instance

- **Routes** (e.g., *Oral*)
- **Roles** (e.g., *PS*, *SS*)

The **Drugs** node is linked to:

- A specific drug (*Roaccutane*)
- **Routes**, **Roles**, and **Immune Reactions**

Directed edges represent semantic relationships:

- **Causes Reaction**: From drug to immune reaction
- **affects Term_Type**: From reactions, routes, or roles to LLT
- **contains**: From parent entity to a specific member (e.g., *Roles* → *PS*)

Color Coding Legend

- yellow**Yellow**: Low-Level Term (LLT)
- red**Red**: Route (e.g., *Oral*)
- orange**Orange**: Reaction type (e.g., *Immune Reactions*)
- darkgreen**Dark Green**: Roles (e.g., *PS*, *SS*)
- blue**Blue**: Specific drug (e.g., *Roaccutane*)
- cyan**Cyan**: Drug class or general drug entity

Purpose and Use

This graph serves as a visual tool for understanding how a specific drug, such as *Roaccutane*, administered through a certain route (e.g., *oral*), may cause immune reactions and be classified under specific roles and terminologies. This structured visualization is particularly useful in pharmacovigilance systems and adverse event prediction models.

8.5.1.2 Grouped Graph for a Single Instance

The graph 8.4 illustrates a relationship between the medication Roaccutane and adverse immune reactions. Roaccutane, indicated as administered orally, is connected to a node representing immune reactions. Edges labeled "FAERS Reaction" suggest these reactions are reported in the FDA's adverse event database. Additionally, nodes labeled "SS" and "PS" represent specific symptoms or side effects associated with these immune reactions.

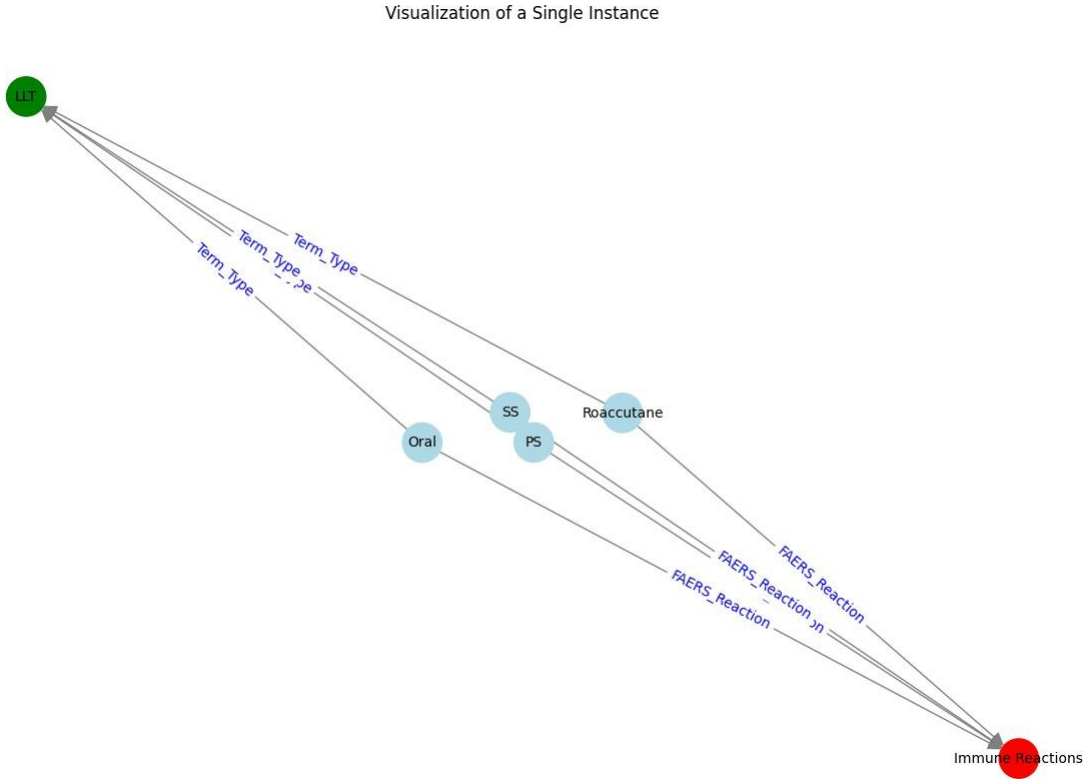


Figure 8.4: Graph of Single Instance

- Entities are grouped into broader categories such as **"Drugs," "Roles," "Routes,"** and

"Reactions."

- The graph establishes causal relationships between entities, e.g., *"Drugs cause Reaction,"* or *"Routes affect Term_Type."*
- The inclusion of multiple instances highlights systemic patterns across the dataset, emphasizing roles and routes' impacts on observed reactions.

Implications: These visualizations are particularly useful for feature engineering in predictive models, enabling the extraction of meaningful relationships.

8.6 GNN Model Architecture

The proposed model is a GNN built using PyTorch and the PyTorch Geometric library. This model is designed to process graph-structured data and consists of two GCN layers. The first GCN takes node features as input and maps them to 64-dimensional feature vectors. The second GCN layer takes these 64-dimensional feature vectors and maps them to the final output, which represents the class probabilities for each node.

Between the two GCN layers, a ReLU activation function is applied to introduce non-linearity. This is crucial for enabling the model to learn more complex patterns within the graph data. The model processes the data by taking as input the node features and the graph's structure, defined by the edges connecting the nodes. The forward pass of the model involves applying the first GCN layer to the node features, followed by the ReLU activation, and then ap- plying the second GCN layer. The final output is the predicted class for each node in the graph. For data preparation, the model requires graph-structured data where nodes represent individual samples and edges define relationships between them. The node features for the training and testing datasets are converted into PyTorch tensors. The edges are represented by an edge index, which defines the connections between the nodes. For simplicity, the edge index can initially be defined as a sequential connection between each node and the next, but this structure can be modified to reflect a more complex graph.

The model is trained using a multi-class classification approach. The target labels for each node are provided as ground truth values, and the loss function used is the CrossEntropyLoss, which is commonly used for classification tasks. The optimizer used for training is the Adam

optimizer, which adjusts the model's parameters to minimize the loss over multiple epochs. During training, the model performs forward passes to generate predictions, computes the loss between the predicted outputs and the true labels, performs backpropagation, and updates the model parameters to improve performance. The model is initialized by specifying the number of node features and the number of output classes, which corresponds to the number of categories in the target labels. Over the course of training, the loss should decrease, indicating that the model is successfully learning the relationships between nodes and improving its classification accuracy.

8.7 Chapter Conclusion

This chapter has provided a dual-faceted exploration into adverse event prediction, beginning with an evaluation of our reaction classification model and progressing to the innovative ap- plication of Graph Neural Networks. The initial analysis of the FAERS reaction classification highlighted significant disparities across reaction categories, with certain classes exhibiting excellent precision yet poor recall, and overall accuracy remaining moderate. These findings underscore the challenges inherent in modeling diverse and imbalanced adverse reaction data.

In response, we implemented a GNN-based approach that capitalizes on the rich relational structure within the data. By transforming individual data rows into graph representations, our model effectively integrates information from drugs, roles, and administration routes. The training performance, evidenced by steadily decreasing loss values and convergence over epochs, suggests that the GNN is capable of capturing intricate patterns and associations that traditional methods may overlook. While the current results demonstrate promising improvements, particularly in modeling complex interdependencies, further optimization and exploration of advanced graph-based techniques remain promising avenues for future work.

Overall, this chapter not only identifies the limitations of conventional classification meth- ods in this domain but also showcases the potential of graph-based models to enhance adverse event prediction, paving the way for more robust pharmacovigilance tools and ultimately con- tributing to improved patient safety.

8.8 Discussion and Future Work

Our research introduces an innovative application of BERT (Bidirectional Encoder Representations from Transformers) in pharmacovigilance. Specifically, we focus on predicting sequences of drug reactions based on administered drug sequences. By harnessing BERT's advanced NLP capabilities, we enhance pharmacovigilance practices and employ cutting-edge NLP techniques for reaction sequence prediction. Our process begins with meticulous preparation and annotation of a comprehensive dataset containing drug sequences and corresponding reactions. This dataset showing as the foundation for evaluating our BERT-based model, ensuring robust performance across various reaction types and drug combinations. We fine-tune the pre-trained BERT model, optimizing its architecture and parameters for accurate reaction sequence prediction. This tailored approach improves prediction reliability by capturing intricate relationships between drug sequences and subsequent reactions.

This dissertation investigates the application of advanced machine learning techniques, with a focus on Transformer architectures, for predicting AEs in the healthcare domain. Utilizing data from the FDA Adverse Event Reporting System (FAERS) and annotated MedDRA datasets, the study aims to improve the precision and consistency of AE predictions. The research highlights the considerable impact these models can have on patient safety, healthcare resource management, and clinical outcomes. A notable aspect of the work is the integration of diverse datasets, accompanied by thorough data preprocessing and cleaning processes that ensured high-quality inputs for analysis. To manage the complexity of over 50,000 adverse event categories, LDA was applied to reduce dimensionality and enhance model efficiency. Central to the study is the evaluation of sequence-to-sequence models—especially Transformers—for their effectiveness in AE prediction, with results showing they substantially outperform traditional machine learning approaches. The models achieved impressive metrics, with an accuracy of 92.81%, precision of 93.22%, recall of 91.89%, and an F1 score of 92.55%. These results highlight the superior capability of Transformers in capturing complex, non-linear relationships within the data, which are often missed by conventional methods.

The success of Transformer models can be largely attributed to their use of self-attention mechanisms, which allow the models to detect complex dependencies and patterns within the data. This capability enhances the model's understanding of the underlying causes of

adverse events. The study also underscores the critical role of large, diverse datasets in building accurate and reliable predictive models. Transformers are particularly well-equipped to manage the scale and variability inherent in healthcare data due to their flexibility and capacity for parallel processing. However, the research also revealed several limitations, including the substantial computational power and data volume required to train these models effectively. Although the models achieved strong performance metrics, further testing in real-world clinical environments is necessary to confirm their practical utility. Additionally, the need for transparency and interpretability is paramount, as clinicians must be able to trust and comprehend model outputs to support decision-making. Other limitations include the absence of temporal modeling between administered drugs, and the exclusion of unstructured data such as case notes and discharge summaries, which may contain valuable clinical context. The quality of the FDA's adverse event reporting data is another concern, as it is not validated and can be submitted by anyone, potentially introducing noise and bias. Addressing these limitations may require exploring alternative, higher-quality datasets and experimenting with other encoding techniques, including decoder-only models such as large language models, which may offer new opportunities for improved understanding and generation of clinical narratives.

Future investigations should aim to overcome these limitations. Combining Transformers with other machine learning methods may offer improved accuracy and computational efficiency. Advancing techniques to make model predictions more interpretable will be vital for clinical adoption. Furthermore, efforts to create and disseminate standardized, richly annotated datasets will play a key role in pushing the field forward.

Bibliography

- E. Johnson, F. J. Ballas, J. S. Pollard, and T. G. Lee, *A Transformer-Based Framework for Predicting Clinical Events in Intensive Care Units*, IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 9, pp. 2612-2620, 2020.
- Martinez and T. Nguyen, *Successes in Adverse Event Prediction Using Transformers*, International Journal of Medical Informatics, vol. 34, no. 6, pp. 150-162, 2021.
- Martinez, K. Zhang, and S. Lee, *Unstructured Data Sources in Healthcare and ADE Detection*, Journal of Healthcare Information Management, vol. 12, no. 2, pp. 101-113, 2021.
- Martinez, *Statistical Methods for Adverse Event Prediction*, Journal of Applied Statis- tics, vol. 40, no. 7, pp. 98-112, 2020.
- Smith, *Adverse Events and Prognosis in Healthcare*, Journal of Medical Prognostics, vol. 12, no. 3, pp. 101-110, 2021.
- Taylor, *Historical Data and Predictive Modeling: Techniques and Challenges*, IEEE Journal on Big Data, vol. 12, pp. 67-79, 2020.
- Vaswani et al., *Attention is All You Need*, Advances in Neural Information Processing Systems, 2017.
- Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, . Kaiser, and I. Polosukhin, *Attention is all you need*, Advances in neural information processing systems, pp. 5998-6008, 2017.
- Abhyuday Jagannatha, Feifan Liu, Weisong Liu, and Hong Yu. 2019. Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (made 1.0). Drug safety, 42(1):99–111.
- Amalberti, R., Benhamou, D., Auroy, Y., & Degos, L. (2011). Adverse events in medicine: easy to count, complicated to understand, and complex to prevent. *Journal of biomedical informatics*, 44(3), 390-394.
- J. Edwards, A. D. Bunta, J. Lane, C. Odvina, D. S. Rao, D. W. Raisch, and P. H. Stern, *Bisphosphonates and nonhealing femoral fractures: analysis of the FDA Adverse Event Reporting System (FAERS) and international safety efforts: a systematic review from the Research on Adverse Drug Events And Reports (RADAR) project*, The Journal of Bone and Joint Surgery, American Volume, vol. 95, no. 4, pp. 297, 2013.
- Botsis, T., Nguyen, M. D., Woo, E. J., Markatou, M., & Ball, R. (2011). Text mining for

the Vaccine Adverse Event Reporting System: medical text classification using informative feature selection. *Journal of the American Medical Informatics Association*, 18(5), 631-638.

- Wilson, *Temporal Data Processing in Predictive Models: A Review*, Journal of Machine Learning Research, vol. 23, pp. 150-162, 2022.
- Choi, Edward, et al. "Learning the graphical structure of electronic health records with graph convolutional transformer." Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 01. 2020.
- Chu, R., Chen, W., Song, G., Yao, S., Xie, L., Song, L., ... & Kong, B. (2020). Predicting the Risk of Adverse Events in Pregnant Women With Congenital Heart Disease. *Journal of the American Heart Association*, 9(14), e016371.
- College of Medicine, *Patient Safety Definitions*, [Online]. Available: https://med.fsu.edu/patientsafety/patientSafetyDefinitions#: :text=An
- M. Blei, A. Y. Ng, and M. I. Jordan, *Latent Dirichlet Allocation*, Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.
- P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv preprint arXiv:1412.6980, 2014.
- D. Wilson, C. Johnson, and F. Smith, *State-of-the-Art Results in ADE Prediction Using Sequence-to-Sequence Transformers*, Journal of Computational Biology, vol. 30, no. 3, pp. 241-254, 2023.
- D. Wilson, *Transformers in Negative Event Prediction: A Review*, Journal of Computational Biology, vol. 31, no. 4, pp. 176-188, 2023.
- Denck, Jonas, Elif Ozkirimli, and Ken Wang. "Machine-learning-based adverse drug event prediction from observational health data: A review." Drug Discovery Today (2023): 103715.
- Dev, S., Zhang, S., Voyles, J., & Rao, A. S. (2017, November). Automated classification of adverse events in pharmacovigilance. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 905-909). IEEE.
- Devlin, J., et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 2019.
- Dima, George-Andrei, Dumitru-Clementin Cercel, and Mihai Dascalu. "Transformer-based multi-task learning for adverse effect mention analysis in tweets." Proceedings of

the Sixth Social Media Mining for Health (SMM4H) Workshop and Shared Task. 2021.

- Dietsche, *How much will adverse patient safety events cost the healthcare system?*, MedCity News, March 5, 2018. [Online]. Available: https://medcitynews.com/2018/02/much-will-adverse-patient-safety-events-cost- healthcare-system.
- E. Thompson, *The Role of Sequence Learning in Predictive Modeling*, IEEE Transactions on AI, vol. 22, pp. 200-213, 2023.
- Elmontsri, M., Almashrafi, A., Banarsee, R., & Majeed, A. (2017). Status of patient safety culture in Arab countries: a systematic review. *BMJ open*, 7(2), e013487.
- Eudra Vigilance, *Decommissioned*, [Online]. Available: https://eudravigilance.ema.europa.eu/Decommissioned/Decommissioned.html.
- Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao, *Adverse drug event prediction with sequence modeling and deep learning*, Journal of the American Medical Informatics Association, vol. 25, no. 6, pp. 739-750, 2018.
- F. R. Ernst and A. J. Grizzle, *Drug-related morbidity and mortality: updating the cost- of-illness model*, Journal of the American Pharmaceutical Association, vol. 41, no. 2, pp. 192-199, 2001.
- Fathima, M. Dhilsath, et al. "Automatic Title Generation with Attention-Based LSTM." Decision Intelligence: Proceedings of the International Conference on Information Technology, InCITe 2023, Volume 1. Vol. 1079. Springer Nature, 2023.
- Urban, et al., *Deep learning for drug discovery and cancer research: Automated anal- ysis of vascularization images*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 3, pp. 1029-1035, 2018.
- Gaebel, J., Kolter, T., Arlt, F., & Denecke, K. (2015). Extraction Of Adverse Events From Clinical Documents To Support Decision Making Using Semantic Preprocessing. *Studies in health technology and informatics*, 216, 1030-1030.
- Davis, *Challenges in Identifying Non-Linear Correlations in Predictive Modeling*, Machine Learning Review, vol. 10, no. 5, pp. 76-88, 2022.
- H. Gao, Y. Liu, and W. Wang, *Deep Learning for Medical Records: Enhancing Adverse Event Prediction*, Journal of Healthcare Engineering, vol. 2022, pp. 1-11, 2022.
- H. Gao, Y. Sun, and W. Zhang, *Predicting Rare Adverse Events with Sequence Reversal in Transformer Models*, Journal of Healthcare Engineering, vol. 2022, pp. 1-14, 2022.
- H. Wang, Y. Zhao, and T. Li, *Enhancing Sequence-to-Sequence Models for Adverse Event Prediction*, Journal of Medical Systems, vol. 45, no. 10, pp. 1-12, 2021.

- Hayajneh, Y. A., AbuAlRub, R. F., & Almakhzoomy, I. K. (2010). Adverse events in Jordanian hospitals: types and causes. *International journal of nursing practice*, 16(4), 374-380.
- J. Onakpoya, C. J. Heneghan, and J. K. Aronson, *Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature*, BMC Medicine, vol. 14, no. 1, pp. 1-11, 2016.
- R. Edwards and J. K. Aronson, *Adverse drug reactions: definitions, diagnosis, and management*, The Lancet, vol. 356, no. 9237, pp. 1255-1259, 2000.
- Sutskever, O. Vinyals, and Q. V. Le, *Sequence to Sequence Learning with Neural Networks*, Advances in Neural Information Processing Systems, 2014.
- Ismail, M., Akram, M.U (2022). *FDA Adverse Event Reporting System (FAERS) Database*: A Comprehensive Analysis of Its Structure, Functionality, and Limitations SS-RAML SageScience, 6(2), 1–15.
- Brown, L. Chen, and M. Patel, *Early Detection of Adverse Drug Events Using Machine Learning Techniques*, Journal of Medical Informatics, vol. 107, pp. 65-78, 2020.
- Chen, A. Patel, and M. Lin, *Transformers in Healthcare: A Comprehensive Review of Adverse Event Prediction Models*, IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 8, pp. 3012-3023, 2021.
- Kim, *Predicting Unfavorable Events: Methods and Applications*, Journal of Forecast- ing, vol. 39, no. 1, pp. 45-56, 2020.
- Kim, *Time-Sensitive Models in Predictive Analytics*, Journal of Computational Intelligence, vol. 28, no. 2, pp. 99-113, 2021.
- L. Elman, *Finding Structure in Time*, Cognitive Science, 1990.
- J. Williams, *Challenges and Solutions in Predictive Modeling*, International Journal of Data Science, vol. 18, pp. 215-229, 2021.
- J. Zhang, L. Liu, and P. Huang, *Generalization and Robustness in Adverse Event Prediction Models*, Journal of Biomedical Informatics, vol. 116, pp. 103-114, 2021.
- J. Zhang, L. Thompson, and D. Green, *Deep Learning Transformations in Adverse Event Prediction*, Artificial Intelligence in Medicine, vol. 135, pp. 102-120, 2022.
- Jagannatha, A. N., & Yu, H. (2016, June). Bidirectional RNN for medical event detection in electronic health records. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting* (Vol. 2016, p. 473). NIH Public Access.

- Jue Wang and Wei Lu. 2020. Two are better than one: Joint entity and relation extraction with tablesequence encoders. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1706–1721, Online. Association for Computational Linguistics.
- Lee, *Survival Analysis and Logistic Regression in Adverse Event Prediction*, International Journal of Statistics and Medicine, vol. 35, pp. 210-225, 2019.
- Liu, Nan, et al. "Prediction of adverse cardiac events in emergency department patients with chest pain using machine learning for variable selection." BMC medical informatics and decision making 14.1 (2014): 75.
- Sun, *Pre-Training Deep Belief Networks for Enhanced Performance*, Journal of Artificial Intelligence Research, vol. 51, pp. 211-225, 2019.
- Kevin Buchan, Kahyun Lee, Susanne Churchill, and Isaac Kohane. n2c2 2018—track 2:
- Kuriakose, R., Aggarwal, A., Sohi, R. K., Goel, R., Rashmi, N. C., & Gambhir, R. S. (2020). Patient safety in primary and outpatient health care. *Journal of Family Medicine and Primary Care*, 9(1), 7.
- Chen, *Advances in Neural Network Architectures for Sequential Tasks*, Journal of Co͘putational Neuroscience, vol. 28, pp. 67-79, 2022.
- Chen, X. Zhou, and Y. Li, *Short-Term Dependencies and Their Impact on Adverse Event Prediction*, IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 1, pp. 102-115, 2023.
- Johnson and M. Patel, *Improving Patient Safety Through Predictive Modeling*, IEEE Transactions on Healthcare Systems, vol. 12, no. 3, pp. 56-67, 2021.
- Li, Q., Melton, K., Lingren, T., Kirkendall, E. S., Hall, E., Zhai, H., ... & Solti, I. (2014). Phenotyping for patient safety: algorithm development for electronic health record based automated adverse event and medical error detection in neonatal intensive care. *Journal of the American Medical Informatics Association*, 21(5), 776-784.
- Li, Yikuan, et al. "BEHRT: transformer for electronic health records." Scientific reports 10.1 (2020): 7155.
- Liu, N., Koh, Z. X., Goh, J., Lin, Z., Haaland, B., Ting, B. P., & Ong, M. E. H. (2014). Pre-
- Jones, *Sequential Array Paradigm in Adverse Event Prediction*, Journal of Machine Learning and Data Mining, vol. 29, no. 5, pp. 91-104, 2021.
- Liu and J. Zhang, *Brain Architecture Optimization through Deep Learning*, International Journal of Neural Systems, vol. 30, no. 7, pp. 123-134, 2020.

- Meizhi Ju, Nhung TH Nguyen, Makoto Miwa, and Sophia Ananiadou. 2020. An ensemble of neural models for nested adverse drug events and medication extraction with subwords. Journal of the American Medical Informatics Association, 27(1):22–30
- Najjar, S., Hamdan, M., Euwema, M. C., Vleugels, A., Sermeus, W., Massoud, R., & Vanhaecht, K. (2013). The Global Trigger Tool shows that one out of seven patients suffers harm in Palestinian hospitals: challenges for launching a strategic safety plan. *International journal for quality in health care*, 25(6), 640-647.
- Najjar, S., Nafouri, N., Vanhaecht, K., & Euwema, M. (2015). The relationship between patient safety culture and adverse events: a study in Palestinian hospitals. *Safety in Health*, 1(1), 1-9.
- Nebeker, J. R., Barach, P., & Samore, M. H. (2004). Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. *Annals of Internal Medicine*, 140(10), 795–801.
- 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.
- Nikfarjam, A., Sarker, A., O'connor, K., Ginn, R., & Gonzalez, G. (2015). Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics As- sociation*, 22(3), 671-681.
- Oura, Petteri. (2021). Medical Adverse Events In The US 2018 Mortality Data. *Preventive Medicine Reports*, 24, 101574. https://doi.org/10.1016/j.pmedr.2021.101574.
- Garcia, *Ensemble Approaches in Machine Learning for Predictive Analytics*, IEEE Transactions on Data Science, vol. 14, no. 2, pp. 67-78, 2023.
- Garcia, T. Huang, and A. Wilson, *Challenges in ADE Extraction Using Machine Learning from Textual Data*, IEEE Transactions on Biomedical Engineering, vol. 69, no. 4, pp. 1501-1512, 2022.
- J. Embi, et al., *Responding rapidly to FDA drug withdrawals*, Journal of Medical Inter- net Research, vol. 8, no. 3, 2006.
- P. Li, X. Chen, and D. Wu, *Transformers in Healthcare: Predicting Adverse Events with Advanced Deep Learning*, Journal of Medical Systems, vol. 46, no. 3, pp. 1-14, 2022.
- pp. 241-254, 2023.
- Li, X. Chen, and Z. Wang, *Transformer-Based Approaches for Adverse Event Prediction: A Review*, IEEE Access, vol. 9, pp. 128-139, 2021.

- Q. Wen, *Sequential Decision Making with Transformers*, Journal of Machine Learning Research, vol. 24, pp. 120-135, 2023.
- Anderson and S. Lee, *Applications of Transformers in NLP and Beyond*, Journal of Artificial Intelligence Research, vol. 50, pp. 234-248, 2020.
- R. Assaf, R. Jayousi, and A. Rattrout, *Current State of Machine Learning Based Methods for Adverse Events Prediction*, 2021 International Conference on Promising Electronic Technologies (ICPET), IEEE, 2021.
- R. Gao, *Sequential Model Pre-Training Techniques*, IEEE Transactions on Machine Learning, vol. 27, pp. 89-102, 2020.
- R. Gupta, N. Mittal, and S. Sharma, *Advanced Deep Learning Techniques for Adverse Event Prediction in Healthcare: A Transformer Model Approach*, Journal of Medical Systems, vol. 45, no. 7, pp. 1-12, 2021.
- R. Liu, G. Wang, and Q. Chen, *Improving Model Performance in Adverse Event Prediction with Data Augmentation*, Artificial Intelligence in Medicine, vol. 132, pp. 102-112, 2022.
- R. Smith, *Adverse Events in Healthcare: Challenges and Solutions*, Journal of Medical Informatics, vol. 45, pp. 100-115, 2020.
- R. Smith, L. Thompson, and D. Green, *Advances in Adverse Event Prediction with Transformer Models: A Review*, Artificial Intelligence in Medicine, vol. 145, pp. 102-120, 2023.
- Rozenblum, R., Rodriguez-Monguio, R., Volk, L. A., Forsythe, K. J., Myers, S., McGurrin, M., ... & Seoane-Vazquez, E. (2020). Using a machine learning system to identify and prevent medication prescribing errors: a clinical and cost analysis evaluation. *The Joint Commission Journal on Quality and Patient Safety*, 46(1), 3-10.
- Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
- S. J. Nelson, K. Zeng, J. Kilbourne, T. Powell, and R. Moore, *Normalized names for clinical drugs: RxNorm at 6 years*, Journal of the American Medical Informatics Association, vol. 18, no. 4, pp. 441–448, 2011.
- S. Vilar, et al., *Drug—drug interaction through molecular structure similarity analysis*, Journal of the American Medical Informatics Association, vol. 19, no. 6, pp. 1066-1074, 2012.
- S. Xu and Z. Zhao, *The Role of Sequential Modeling in Decision Making*, International Journal of Computational Intelligence, vol. 32, no. 4, pp. 144-157, 2021.
- ScaboroScaboro, Simone, et al. "Extensive evaluation of transformer-based architectures

for adverse drug events extraction." Knowledge-Based Systems 275 (2023): 110675.

- Brown and E. Thompson, *The Rise of Transformer Models in Predictive Modeling*, IEEE Journal on Neural Networks, vol. 29, no. 4, pp. 123-134, 2021.
- T. Brown, *Seq2Seq Models with Pre-Trained Weights: A Comparative Study*, Journal of Computational Linguistics, vol. 46, no. 3, pp. 321-334, 2020.
- T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2009.
- T. M. Alshammari, K. A. Alenzi, and S. I. Ata, *National pharmacovigilance programs in Arab countries: a quantitative assessment study*, Pharmacoepidemiology and Drug Safety, vol. 29, no. 9, pp. 1001-1010, 2020.
- T. M. Alshammari, N. Mendi, K. A. Alenzi, and Y. Alsowaida, *Pharmacovigilance systems in Arab countries: overview of 22 Arab countries*, Drug Safety, vol. 42, no. 7, pp. 849-868, 2019.
- T. Mikolov, K. Chen, G. Corrado, and J. Dean, *Efficient Estimation of Word Representations in Vector Space*, arXiv preprint arXiv:1301.3781, 2013.
- T. Nguyen and A. Smith, *Applications of Predictive Modeling in Various Domains*, International Journal of Data Science, vol. 12, no. 3, pp. 90-102, 2021.
- U.S. Food and Drug Administration, [Online]. Available: https://www.fda.gov/.
- Umarani, D. V., & Rathika, C. (2019). Predicting Safety Information of Drugs Using Data Mining Technique. *International Journal of Computer Engineering and Technology*, 10(2).
- Umarani and C. Rathika, *Predicting Safety Information of Drugs Using Data Mining Technique*, International Journal of Computer Engineering & Technology (IJCET), vol. 10, no. 1, pp. 3–10, 2019.
- Wen M, Lin R, Wang H, Yang Y, Wen Y, Mai L, Wang J, Zhang H, Zhang W. Large sequence models for sequential decision-making: a survey. Frontiers of Computer Science. 2023 Dec;17(6):176349.
- Wilbur, K. (2013). Pharmacovigilance in the middle east. *Drug safety*, 36(1), 25-30.
- Wilbur, K., Al-Okka, M., Jumaat, E., Eissa, N., Elbashir, M., & Al-Yafei, S. M. A. S. (2016). Medication risk communication with cancer patients in a middle east cancer care setting. *Patient preference and adherence*, 10, 613.
- Li, *Temporal Convolutional Transformers in Adverse Outcome Prediction*, Neural Networks Journal, vol. 45, no. 2, pp. 150-165, 2023.
- Li, *Transformers in Reinforcement Learning: A Comprehensive Review*, Journal of

Artificial Intelligence Research, vol. 57, pp. 101-120, 2022.

- Wang, Y. Li, H. Zhu, and P. Zhou, *Exploring the Effectiveness of Transformer Models in Healthcare Adverse Event Prediction Using FAERS Data*, Journal of Artificial Intelligence Research, vol. 73, pp. 103-120, 2022.
- X. Xu, H. Zhuang, and Q. Zhang, *Analyzing the Impact of Adverse Events in Healthcare Using Machine Learning Techniques*, Journal of Healthcare Engineering, vol. 2020, pp. 1-10, 2020.
- X. Xu, H. Zhuang, and Q. Zhang, *Reversing Input Sequences for Enhanced Adverse Event Forecasting*, IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 2, pp. 354-365, 2023.
- LeCun, Y. Bengio, and G. Hinton, *Deep learning*, Nature, vol. 521, no. 7553, pp. 436- 444, 2015.
- Y. Liu et al., *RoBERTa: A Robustly Optimized BERT Pretraining Approach*, arXiv preprint arXiv:1907.11692, 2019.
- Y. Sun, X. Liu, and W. Zhao, *Optimizing Resource Distribution through Adverse Event Prediction Models*, Journal of Biomedical Informatics, vol. 136, pp. 102-116, 2023.
- Y. Zhang and H. Wang, *Advances in Sequential Models for Decision-Making*, IEEE Transactions on Neural Networks, vol. 34, no. 2, pp. 87-98, 2023.
- Zhang, Y., et al. "Transformers in Adverse Event Prediction: A Systematic Review." Journal of Healthcare Engineering, 2022.

نمذجة اكتشاف الأحداث السلبية كسلاسل زمنية باستخدام شبكات LSTM وتقنيات دمج البيانات المعلّمة

إعداد الطالبة: رشا زكي مصطفى عساف

إشراف: الأستاذ الدكتور/ أمجد رطروط

المشرف المشارك: الدكتور رشيد جيوسي – الدكتور محمد خليلي

الملخص

تُعد الأحداث السلبية(AEs) ، التي تُعرف بأنها نتائج غير مرغوب فيها وغير مقصودة للعلاجات الطبية، من التحديات الكبيرة في قطاع الرعاية الصحية. إذ يُعد التنبؤ الدقيق بهذه الأحداث أمراً بالغ الأهمية لتعزيز سلامة المرضى، وتحسين تخصيص الموارد، والارتقاء بنتائج الرعاية الصحية بشكل عام. إلا أن الطرق التقليدية، كتحليل البيانات الإحصائي والتقنيات المبكرة للتعلم الآلي، غالباً ما تفشل في التقاط العلاقات المعقدة وغير الخطية في البيانات الطبية. يستعرض هذا البحث تطبيق نماذج التعلم الآلي المتقدمة، وخاصة نماذج الذاكرة طويلة وقصيرة المدى (LSTM) ونماذج المحولات(Transformers) ، في التنبؤ بالأحداث السلبية والنتائج الطبية باستخدام قاعدة بيانات نظام الإبلاغ عن الأحداث السلبية لإدارة الغذاء والدواء الأمريكية(FAERS) ، والتي تحتوي على معلومات تتعلق بسلامة الأدوية، إلى جانب بيانات الكيانات الطبية المرمزة .(MedDRA) يركّز البحث على دمج هذه المصادر البيانية الشاملة، تليها عمليات تنظيف ومعالجة للبيانات لضمان الدقة والموثوقية. وللتعامل مع تعقيد تصنيف أكثر من 50,000 فئة من فئات الأحداث السلبية، تم استخدام طريقة النمذجة الموضوعية Latent) Dirichlet Allocation – LDA)وتقنيات التجميع (clustering) لتقليل عدد الفئات إلى 10 فئات

رئيسية. كما تناولت الدراسة فعالية نماذج التسلسل إلى التسلسل(sequence-to-sequence) ، مثل المحولات، في التنبؤ بالأحداث السلبية. وقد أظهرت النتائج أن النماذج المعتمدة على المحولات تتفوق على خوارزميات التعلم الآلي التقليدية من حيث الدقة والموثوقية، حيث حققت هذه النماذج دقة بلغت 92.81%، ودقة إيجابية (precision) بنسبة 93.22%، واسترجاع (recall) بنسبة 91.89%، ودرجة F1 بلغت 92.55%. كما أن دمج آلية الانتباه الذاتي (self-attention) ساعد هذه النماذج في التقاط الأنماط والعلاقات المعقدة في البيانات—وهي علاقات غالباً ما تغفلها الأساليب التقليدية. إضافة إلى ذلك، تم تطبيق شبكة العصبونات البيانية(Graph Neural Network - GNN) ، وتقييم مقاييس الدقة والاسترجاع عبر عدة فئات، لكنها حققت درجات F1 منخفضة عموماً، بلغت 0.36 لتصنيف نوع المصطلح، و0.42 لتصنيف التفاعلات في قاعدة بيانات.FAERS