

Advancement of Natural Language Processing Technology in Pharma Industry 4.0

Atchuta Kumar Kandukuri^{1*}, Bhagavan Rajesh Babu Koppisetty², Swathi Nageswara³ and Sheik Jakir Hussain Mustaq⁴

^{1,3,4}JNTU-GV College of Pharmaceutical Sciences, JNTU-GV, Vizianagaram, Andhra Pradesh –535003, India

²Department of Pharmaceutical Sciences, Vignan Institute of Pharmaceutical Technology (Autonomous), Visakhapatnam, Andhra Pradesh-530049, India

*Corresponding Author: Professor, Atchuta Kumar Kandukuri
dratchuta99@gmail.com

Received: 28th Feb, 2026; Revised: 6th March 2026; Accepted: 7th April, 2026; Available Online: 20th April, 2026

ABSTRACT

Natural Language Processing includes a number of approaches such as text classification, named entity recognition, sentiment analysis, language generation, machine translation, text summarization, question answering, and information extraction. The pharmaceutical industry has started employing language processing for the improvement of its operations. Generally, NLP is the use of statistical and linguistic methods to enable computers to interpret and generate human language properly. Pharma incorporates language processing changes in drug discovery, clinical trials, pharmacovigilance, and patient engagement. Natural Language Processing technology in conjunction with the pharmaceutical industry anticipates a big step forward in the Industry 4.0 era. Natural Language Processing can fundamentally remake many areas of drug discovery, development and dispensing by enabling faster data analysis, enhanced accuracy in trial execution, improved patient care and better supply chain management. Technological advances in Natural Language Processing are capable of addressing and interpreting vast unstructured data for refreshing insights to boost the operational efficiency of the pharmaceutical industry.

Keywords: Drug discovery; Pharmacovigilance; Clinical trials; Natural Language Processing; Artificial intelligence.

How to cite this article: Kandukuri AK, Koppisetty BRB, Nageswara S, Mustaq SJH. Advancement of Natural Language Processing Technology in Pharma Industry 4.0. Int J Drug Deliv Technol. 2026;16(52s): 257-264. DOI: 10.25258/ijddt.16.52s.30

Source of support: Nil.

Conflict of interest: None

INTRODUCTION

Natural Language Processing (NLP) is a sub-domain of artificial intelligence and computer science concentrating on the interaction between computers and human languages. It involves the creation of algorithms and models to understand, interpret, and generate natural language text or speech¹. NLP covers a large number of tasks and techniques. This involves categorization of text into various predefined types or categories. Some of the applications involve sentiment analysis, topic categorization, spam filtering, and intent identification². Named Entity Recognition is defined as the task of recognizing and extracting specific parts of texts that represent entities such as names of people, organizations, places, dates, and so forth³. Sentiment analysis describes what the text expresses, whether positive, negative, or neutral opinion. To provide insights for social media, customer reviews, and feedback analytics, its utilitarian value is significant⁴. Language Generation is the process which employs algorithms that give rise to human-like text or speech, including that which is used in text generation, chatbots, and virtual assistants. It includes understanding of the input and generating appropriate coherent responses⁵.

METHODS

Machine translation is the automatic conversion of written text from one language to another. Examples of this type of technology include Google Translate and many other language translation systems^{6,7}. Text Summarization is the compendium of a huge document or an article into short texts preserving general points and ideas. Question Answering is the process whereby the computers are able to be programmed to understand and respond to questions put forth in natural language. These could be both simple factoid questions and complicated questions that need a reasoning step or an inference^{8,9}. Information Extraction is the process of obtaining structured data from some unstructured text, including relation extraction of entities, events, or facts^{10,11}. They tend to include algorithms ranging from neural networks via hidden Markov models to probabilistic models and other statistics and machine-learning approaches. They require a large range of annotated datasets upon which they are to be trained and quite significant computational power for processing and analysis¹². NLP, in the context of big data in the pharma industry, is that which has lent the term "Big Pharma" to the large volumes of data involved. In Fig. 1, NLP is seen in rapid advances in the pharmaceutical field. Here are some of the NLP applications in big data in the

*Author for Correspondence: dratchuta99@gmail.com

pharmaceutical domain, summarized in **Table I**. The enhanced pharmaceutical operations with NLP is represented in **Fig. 2**.



Fig.1 Natural Language Processing in Pharmaceutical Industry 4.0

Table I: Diverse application of NLP in field of pharmaceuticals.

PHARMA DOMAIN	APPLICATION
<i>Knowledge Discovery and Data Exploration</i>	1) Text Mining and Information Retrieval 2) Text Summarization 3) Topic Modeling 4) Trend Analysis and Pattern Detection 5) Sentiment Analysis and Social Media Monitoring 6) Entity Extraction and Linking 7) Question Answering
<i>Pharmacovigilance and Adverse Event Monitoring</i>	1) Adverse Event Detection 2) Signal Detection 3) Case Prioritization and Triage 4) Literature Review 5) Social Media Monitoring 6) Safety Signal Management
<i>Clinical Data Abstraction and Annotation</i>	1) Data Extraction 2) Entity Recognition 3) Relationship Extraction 4) Temporal Analysis 5) Clinical Coding 6) Clinical Decision Support 7) Quality Assurance 8) Clinical Research
<i>Drug Repurposing and Discovery</i>	1) Literature Mining 2) Entity Recognition 3) Relationship Extraction 4) Text Similarity and Clustering 5) Adverse Event Analysis 6) Knowledge Graph Construction 7) Clinical Trial Identification 8) Data Integration and Analytics

Regulatory Compliance and Document Analysis	<ol style="list-style-type: none"> 1) Regulatory Document Classification 2) Information Extraction 3) Pharmacovigilance and Adverse Event Reporting 4) Labeling Compliance 5) Regulatory Intelligence 6) Document Summarization 7) Compliance Monitoring and Auditing 8) Regulatory Text Analytics
--	--

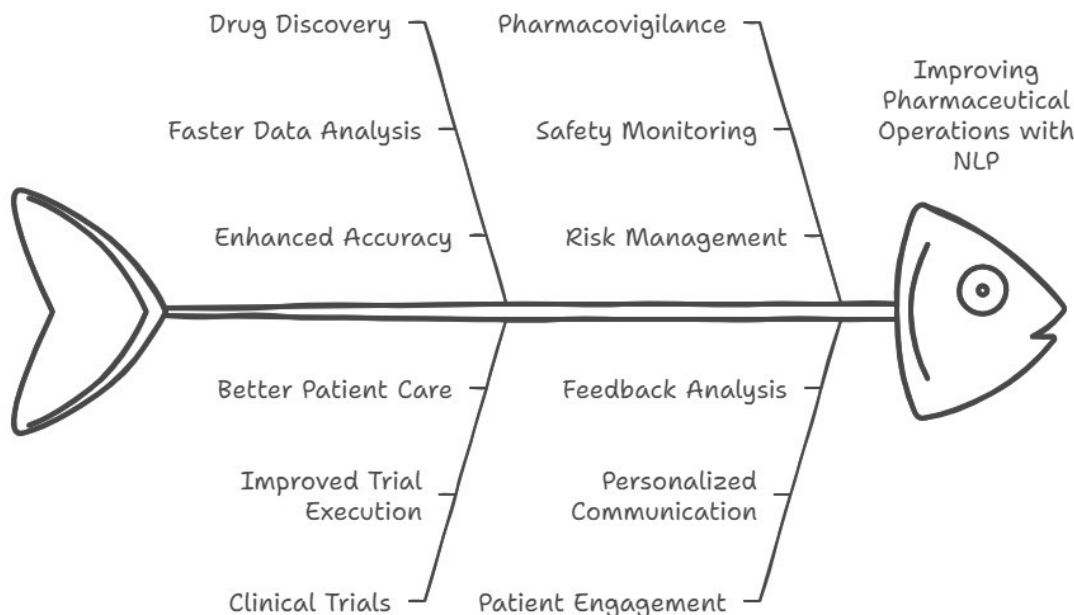


Fig.2 Enhanced pharmaceutical operations with NLP

RESULTS

Knowledge Discovery and Data Exploration

NLP makes big data smart and can lead to knowledge through the acquisition of insights from text documents such as research articles, clinical guidelines, and drug labels. NLP algorithms can delve into the relationships, patterns, and insights hidden in large texts so that researchers and analysts can focus on developing new hypotheses, linkages, and areas for further studies. NLP is important as it opens doors for knowledge discovery and data exploration through the extraction of useful insights and patterns from textual data. Here is how NLP is used in these aspects:

Text Mining and Information Retrieval

NLP techniques can serve with processing and analyzing large volumes of textual data and extracting useful information. Such processes entail, for example, document indexing, keyword extraction, and document retrieval, which allows users to search and fetch relevant documents or pieces of information from large collections¹³.

Text Summarization

Summarization allows a long text to be summarized in an automatic manner so that one can gather the significant points or main information that is contained in a document or a group of documents quickly. Text

summarization provides for the triage of information; thus, users can easily browse through and understand large amounts of text¹⁴.

Topic Modeling

NLP techniques, such as Latent Dirichlet Allocation, are used to identify latent topics in a collection of documents. Topic modeling helps one to understand the main themes and subject areas represented in the data, allowing the user to search and browse through the content based on the identified topics¹⁵.

Trend Analysis and Pattern Detection

NLP enables identifying trends and patterns within textual data. By processing enormous volumes of text, NLP algorithms can detect changing topics, sentiment trends, or linguistic patterns over a specified time. This aids in obtaining insights regarding changing trends, opinion of the public, or a market adjustment¹⁶.

Sentiment Analysis and Social Media Monitoring

The pharmaceutical companies may prepare grand plans utilizing NLP techniques and analyze social media data, patient forums, and online reviews to know public sentiment or opinion and the experiences related to the drugs, treatments, and healthcare services. NLP use cases include sentiment analysis to measure perceptions by patients or consumers and identify trends or measure outcomes of interventions or marketing campaigns¹⁷.

Entity Extraction and Linking

Natural Language Processing is involved with the identification and extraction of an entity (for example, people, organizations, locations, or products) from text data. Additionally, entity linking entails attaching these entities to outside knowledge bases or databases so that they contain contextual information. Extracting and linking these entities facilitates data exploration by providing the user with inquiry and navigation capabilities describing relations between entities¹⁸.

Pharmacovigilance and Adverse Event Monitoring

NLP is important to pharmacovigilance because the large number of adverse event reports become easier to work with due to NLP capabilities. As an example, NLP algorithms can be used to automatically extract and categorize key information from published literature, which could facilitate signal detection, adverse event tracking and enable the real-time monitoring of drug safety. This is how NLP is applied to those domains:

Adverse Event Detection

NLP techniques make it possible to identify and extract adverse events from a wide range of text sources. This involves identifying and categorizing medical terms, symptoms, drug names, and how they're connected within the context. By automating this entire process, NLP significantly speeds up the detection of potential adverse events, making monitoring much more efficient^{19,20}.

Signal Detection

NLP is a important when it comes to detecting signals and uncovering trends hidden within large volumes of adverse event reports. By examining into patterns, analyzing how terms show up together, and understanding the deeper meaning behind the text, NLP algorithms can bring to potential relationship between drugs and adverse effects, events that might need further investigation²¹.

Case Prioritization and Triage

Automatic classification and prioritization of adverse event reports based on their severity, relevance, or novelty can be done using NLP methods. The method makes it possible to use resources economically for further observation and investigation of the most serious cases²².

Literature Review

NLP can retrieve relevant information from clinical trial reports and scientific literature. It can assist in the identification and summarization of adverse event-related key information on some drugs, thereby facilitating the review process for pharmacovigilance professionals²³.

Social Media Monitoring

It is possible to employ NLP algorithms to monitor social media and internet forums for mentions of drug experience, side effect, or adverse effect. Text analysis of websites increases the chances of identifying potential signals for new or previously unidentified adverse effects, providing a secondary source of data for pharmacovigilance²⁴.

Safety Signal Management

NLP techniques can help in handling safety signals through the automation of detection, evaluation, and reporting of potential safety issues related to specific drugs or medical devices. This technique makes the process of signal management simpler and allows quicker decision making on regulatory actions²⁵. Through the use of natural language processing in pharmacovigilance and adverse event monitoring the regulatory agencies, drug companies and healthcare professionals can improve their capability to detect and assess adverse events. This development results in increased patient safety and risk management in the health sector.

Clinical Data Abstraction and Annotation

NLP techniques can automate the abstraction and annotation of clinical data from structured and unstructured sources, such as electronic health records (EHRs) and clinical trial data. The algorithms can pull out key information like diagnoses, treatments, lab results, and patient demographics, which makes large-scale data analysis, cohort identification, and clinical research possible. Here is how NLP is used in these aspects:

Data Extraction

NLP can extract structured data elements from clinical documents such as EHRs or clinical trial reports without any human intervention. It encompasses extracting and bringing out required information such as patient demographics, clinical conditions, treatments, lab results and other clinical data of interest²⁶.

Entity Recognition

NLP models are proficient in identifying and categorizing entities within clinical text, including medical concepts, drug names, procedures, symptoms, and anatomical locations. By systematically recognizing these elements, NLP enables the annotation and extraction of relevant information for downstream analysis²⁷.

Relationship Extraction

Relations among clinical entities can be extracted and detected using NLP methods. They can detect, the relationship between treatment and adverse effect. This assists in building the structured data representations that capture the relation among clinical entities appropriately²⁸.

Temporal Analysis

NLP methods can be employed to uncover and extract relationships between entities in clinical text. For instance, they can be used to identify associations between a treatment and its adverse effects, or between a therapeutic intervention and patient response. This facilitates the generation of structured data representations that effectively capture the relationships among various clinical entities²⁹.

Clinical Coding

NLP is applied to mechanize the task of allocating standardized clinical codes to text information, such as ICD (International Classification of Diseases) codes and

SNOMED-CT (Systematized Nomenclature of Medicine - Clinical Terms) codes. This automation helps with data integration, retrieval, and analysis, promoting consistency in clinical coding practices³⁰.

Clinical Decision Support

NLP methods can be employed to evaluate clinical text and support decision-making through the extraction of relevant information to support clinical decision-making. This includes the identification of contraindications, drug interactions, treatment guidelines, and evidence-based guidelines from clinical literature, research reports, or clinical practice guidelines³¹.

Quality Assurance

NLP methods can be automatically applied to review and validate abstracted clinical information. By comparing extracted data against reference standards or predefined rules, NLP models can detect inconsistencies or errors, thereby enhancing the accuracy and integrity of clinical datasets³².

Clinical Research

NLP plays a vital role in the retrieval of relevant data from clinical trials reports, the scientific literature, and patient medical histories. It enables the clinical trial eligibility criteria, adverse reactions, treatment outcomes, and other key data elements for research [33].

By using NLP for clinical data abstraction and annotation, clinicians, researchers, and healthcare organizations are able to effectively analyze and manage large volumes of unstructured clinical text. This is achieved through the provision of higher-quality data, greater research capability, and more evidence-based clinical decision-making

Drug Repurposing and Discovery

NLP can play a significant role in drug repurposing by analyzing large volumes of scientific literature, patents, and clinical trial data. By uncovering relationships among drugs, diseases, molecular targets, and biological pathways, NLP algorithms can identify potential candidates for novel therapeutic indications. These methods are increasingly adopted in drug discovery and repurposing efforts, as they facilitate the efficient screening of extensive biomedical text sources. Here is how NLP is used in these aspects:

Literature Mining

NLP methods are used to query information from biomedical databases and scientific publications. Through text analysis, NLP is able to recognize potential drug candidates, therapeutic targets, molecular interactions, and disease relationships. This can unveil known drug knowledge and their potential applications other than their intended use³⁴.

Entity Recognition

NLP models are capable of identifying and classifying key entities in text like drug names, target proteins, genes, diseases, and biological processes. By accurately spotting these elements, NLP helps uncover meaningful links between drugs, targets, and diseases. These associations

can offer valuable insights, especially when it comes to drug repurposing opportunities³⁵.

Relationship Extraction

NLP models can identify and classify different entities such as drug names, target proteins, genes, diseases, and biological processes from text. Through the precise identification of these entities, NLP enables the identification of drug, target, and disease relationships and associations, providing useful information for drug repurposing³⁶.

Text Similarity and Clustering

NLP is able to assess the similarity between different drugs, diseases, or targets by analyzing their text descriptions. By clustering similar entities, NLP is able to reveal potential drug repurposing opportunities by finding drugs with similar mechanisms of action or diseases with similar molecular profiles³⁷.

Adverse Event Analysis

NLP is able to recognize the similarity between different drugs, diseases, or targets by comparing their text descriptions. By categorizing similar entities, NLP is able to recognize potential drug repurposing opportunities by recognizing drugs with similar mechanisms of action or diseases with similar molecular profiles³⁸.

Knowledge Graph

Knowledge graphs can be generated using NLP approaches to depict the interactions between targets, drugs, diseases, and other key entities. The knowledge graphs make it easier to navigate and visualize complex biomedical relationships, which aids drug repurposing and discovery efforts³⁹.

Clinical Trial Identification

Using NLP algorithms, one can mine clinical trial databases and literature to discover relevant clinical trials for the specific medicines or diseases. These data are relevant for assessing the plausibility of candidate drug repurposing treatments, and they guide for subsequent trials⁴⁰.

Data Integration and Analytics

NLP can help consolidate millions of data pools from both structured and unstructured data to identify patterns, correlations, and trends. NLP analyses combine textual data with other biomedical data and thus enable a thorough integrated analysis for drug repurposing and discovery⁴¹.

Regulatory Compliance and Document Analysis

In regulatory compliance, complex documents such as the drug label, safety report, and adverse event narratives are great candidates for NLP methods. NLP algorithms aid in automating the process of extracting and standardising key information, thus making it easier to comply with regulatory requirements. Because of this, their utility is escalating quickly in the context of document analysis. NLP is not only enhancing the efficiency and reducing the chances of errors in regulatory procedures from interpreting, classifying content to extracting key

information from text-heavy sources. This is how these are done with the help of NLP :

Regulatory Document Classification

NLP models has the capacity to sort the regulatory documents by type or intended use case such as regulatory submissions, labeling details, adverse event reports, safety updates and guidelines. Through analyzing the document, identifying the language patterns and recognizing high order terms, NLP algorithms are able to effectively organize these text types with minimal human training⁴².

Pharmacovigilance and Adverse Event Reporting

NLP methods help in extracting valuable information from regulatory documents like drug names, active pharmaceutical ingredients, dosage and the adverse effects, indications, contraindications or its warnings as well as safety info. NLP allows for data retrieval, analytics, and regulatory compliance surveillance, by extracting and aggregating that information^{43,44}.

Labeling Compliance

NLP approaches help determine whether drug labels and product information comply with regulatory guidance and requirements. NLP models can highlight departures from regulatory guidelines and warn of potential compliance problems by performing text analysis of labeling, a convenient way for manufacturers to make sure product information is accurate as well as compliant⁴⁵.

Regulatory Intelligence

NLP could be leveraged to assist in the monitoring and analysis of updates to clinical guidelines, regulatory frameworks, and compliance requirements. NLP models enable the extraction of significant details from regulatory documents so that new changes are captured as soon as these happen, allowing organizations to stay compliant and modify their strategies⁴⁶.

Document Summarization

NLP algorithms can summarize long regulatory documents, guidelines or scientific articles when you need only short summaries highlighting essential content. This enables regulatory experts and compliance officers to quickly know what the key aspects are from long documents⁴⁷.

Compliance Monitoring and Auditing

Methods from NLP can automate compliance monitoring as well as auditing through the analysis of huge volumes of regulatory filings, reports and submissions. They can also be used to detect inconsistencies, flag possible compliance risks and help with risk assessment and regulatory compliance assessment⁴⁸.

Regulatory Text Analytics

NLP provides the ability to perform various techniques like sentiment analysis, opinion mining and topic modeling for extracting useful insights from regulatory documents. All together, the seamless integration with compliance workflows means it can help organizations automate document analysis, continuously tune to new

regulations, provide regulatory intelligence and enable more timely and informed decisions in a complex regulatory landscape⁴⁹.

CONCLUSION

NLP has the capabilities to fundamentally change the way in which we are able to extract actionable insights from a high volume of unstructured sources of biomedical and regulatory data that has previously not been amenable to reliable analysis. NLP plays a central and important role in all the aspects of drug lifecycle from a discovery of a drug to clinical research, pharmacovigilance, patient care and also supply chain management during Industry 4.0 era. NLP drives operational efficiency and fuels innovation by enabling data interpretation, automation, as well as decision-making across pharmaceutical organizations. To implement it, firms must leverage sophisticated AI-based platforms, together with cooperation between pharmaceutical heads and NLP experts respectively, reinforcement by high-quality datasets sourced from clinical and scientific points in time and timely responses to changing regulatory standards. Lastly, robust ethical frameworks guiding data privacy, transparency and algorithmic bias are needed as to allow for safe and trustworthy deployment of NLP technology across both health care and pharmaceutical settings.

Competing interests

The authors declare that they have no competing interests

REFERENCES

- [1] Chowdhary K, Chowdhary KR. Natural language processing. Fundamentals of artificial intelligence. 2020:603-49.
- [2] Choi J, Ijaz MF. A Complete Process of Text Classification System Using State-of-the-Art NLP Models.
- [3] Sharma A, Amrita, Chakraborty S, Kumar S. Named entity recognition in natural language processing: A systematic review. In Proceedings of Second Doctoral Symposium on Computational Intelligence: DoSCI 2021 2022 (pp. 817-828). Springer Singapore.
- [4] Nasukawa T, Yi J. Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the 2nd international conference on Knowledge capture 2003; (pp. 70-77). <https://doi.org/10.1145/945645.945658>
- [5] Reiter E, Dale R. Building applied natural language generation systems. Natural Language Engineering. 1997;3(1):57-87.
- [6] Wang H, Wu H, He Z, Huang L, Church KW. Progress in machine translation. Engineering. 2022; 18:143-53.
- [7] Adhikari S. Nlp based machine learning approaches for text summarization. In 2020 Fourth International Conference on Computing

- Methodologies and Communication (ICCMC) 2020; (pp. 535-538). IEEE.
- [8] Gupta P, Gupta V. A survey of text question answering techniques. *International Journal of Computer Applications*. 2012; 53(4).
- [9] Abacha AB, Zweigenbaum P. MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies. *Information processing & management*. 2015; 51(5):570-94.
- [10] Cowie J, Lehnert W. Information extraction. *Communications of the ACM*. 1996; 39(1):80-91.
- [11] Piskorski J, Yangarber R. Information extraction: Past, present and future. In *Multi-source, multilingual information extraction and summarization 2012*; (pp. 23-49).
- [12] Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*. 2011; 18(5):544-51.
- [13] Venugopal V, Sahoo S, Zaki M, Agarwal M, Gosvami NN, Krishnan NA (2021) Looking through glass: Knowledge discovery from materials science literature using natural language processing. *Patterns*.
- [14] Merchant K, Pande Y. Nlp based latent semantic analysis for legal text summarization. In *2018 international conference on advances in computing, communications and informatics (ICACCI) 2018*; (pp. 1803-1807).
- [15] Sarioglu E, Choi HA, Yadav K. Clinical report classification using natural language processing and topic modeling. In *2012 11th international conference on machine learning and applications 2012 (Vol. 2, pp. 204-209)*.
- [16] Hodorog A, Petri I, Rezgui Y. Machine learning and Natural Language Processing of social media data for event detection in smart cities. *Sustainable Cities and Society*. 2022; 85:104026.
- [17] Kanakaraj M, Guddeti RM. NLP based sentiment analysis on Twitter data using ensemble classifiers. In *2015 3Rd international conference on signal processing, communication and networking (ICSCN) 2015*; (pp. 1-5).
- [18] Greenfield K, Caceres RS, Coury M, Geyer K, Gwon Y, Mattereder J, Mensch AC, Sahin CS, Simek O. A Reverse Approach to Named Entity Extraction and Linking in Microposts. In *# Microposts 2016*; (pp. 67-69).
- [19] Yin Z, Zhang C, Goldberg DW, Prasad S. An NLP-based question answering framework for spatio-temporal analysis and visualization. In *Proceedings of the 2019 2nd international conference on geoinformatics and data analysis 2019* (pp. 61-65).
- [20] Dandala B, Joopudi V, Devarakonda M. Adverse drug events detection in clinical notes by jointly modeling entities and relations using neural networks. *Drug safety*. 2019; 42:135-46.
- [21] Nikfarjam A, Ransohoff JD, Callahan A, Jones E, Loew B, Kwong BY, Sarin KY, Shah NH. Early detection of adverse drug reactions in social health networks: a natural language processing pipeline for signal detection. *JMIR public health and surveillance*. 2019; 5(2):e11264.
- [22] Lieber T, Gosselt HR, Kools PC, Kruijssen OC, Van Lierop SN, Härmark L, Van Hunsel FP. Natural language processing for automated triage and prioritization of individual case safety reports for case-by-case assessment. *Frontiers in Drug Safety and Regulation*. 2023; 3:1120135.
- [23] Luo Y, Thompson WK, Herr TM, Zeng Z, Berendsen MA, Jonnalagadda SR, Carson MB, Starren J. Natural language processing for EHR-based pharmacovigilance: a structured review. *Drug safety*. 2017; 40:1075-89.
- [24] Yang CC, Yang H, Jiang L, Zhang M. Social media mining for drug safety signal detection. In *Proceedings of the 2012 international workshop on Smart health and wellbeing 2012* (pp. 33-40).
- [25] Kreimeyer K, Dang O, Spiker J, Gish P, Weintraub J, Wu E, Ball R, Botsis T. Increased confidence in deduplication of drug safety reports with natural language processing of narratives at the us food and drug administration. *Frontiers in Drug Safety and Regulation*. 2022; 2:918897.
- [26] BuHamra SS, Almutairi AN, Buhamrah AK, Almadani SH, Alibrahim YA. An NLP tool for data extraction from electronic health records: COVID-19 mortalities and comorbidities. *Frontiers in Public Health*. 2022;10:1070870.
- [27] Gligic L, Kormilitzin A, Goldberg P, Nevado-Holgado A. Named entity recognition in electronic health records using transfer learning bootstrapped neural networks. *Neural Networks*. 2020; 121:132-9.
- [28] Li F, Yu H. An investigation of single-domain and multidomain medication and adverse drug event relation extraction from electronic health record notes using advanced deep learning models. *Journal of the American Medical Informatics Association*. 2019; 26(7):646-54.
- [29] Kaufman DR, Sheehan B, Stetson P, Bhatt AR, Field AI, Patel C, Maisel JM. Natural language processing-enabled and conventional data capture methods for input to electronic health records: a comparative usability study. *JMIR medical informatics*. 2016; 4(4):e5544.
- [30] Nguyen AN, Truran D, Kemp M, Koopman B, Conlan D, O'Dwyer J, Zhang M, Karimi S,

- Hassanzadeh H, Lawley MJ, Green D. Computer-assisted diagnostic coding: effectiveness of an NLP-based approach using SNOMED CT to ICD-10 mappings. In *AMIA Annual Symposium Proceedings 2018* (Vol. 2018, p. 807).
- [31] Szlosek DA, Ferrett J. Using machine learning and natural language processing algorithms to automate the evaluation of clinical decision support in electronic medical record systems. *eGEMs*. 2016; 4(3):1222.
- [32] Alzu'bi AA, Watzlaf VJ, Sheridan P. Electronic health record (EHR) abstraction. *Perspectives in health information management*. 2021;18(Spring):1g.
- [33] Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *Journal of Allergy and Clinical Immunology*. 2020; 145(2):463-9.
- [34] Agarwal P, Searls DB. Literature mining in support of drug discovery. *Briefings in bioinformatics*. 2008; 9(6):479-92.
- [35] Sun C, Yang Z, Wang L, Zhang Y, Lin H, Wang J. Deep learning with language models improves named entity recognition for PharmaCoNER. *BMC bioinformatics*. 2021; 22(Suppl 1):602.
- [36] Bhatnagar R, Sardar S, Beheshti M, Podichetty JT. How can natural language processing help model informed drug development?: a review. *JAMIA open*. 2022; 5(2):ooac043.
- [37] Öztürk H, Özgür A, Schwaller P, Laino T, Ozkirimli E. Exploring chemical space using natural language processing methodologies for drug discovery. *Drug Discovery Today*. 2020; 25(4):689-705.
- [38] Young IJ, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. *International journal of medical informatics*. 2019;132:103971. <https://doi.org/10.1016/j.ijmedinf.2019.103971>
- [39] Sang S, Yang Z, Wang L, Liu X, Lin H, Wang J. SemaTyP: a knowledge graph based literature mining method for drug discovery. *BMC bioinformatics*. 2018;19:1-1.
- [40] Ni Y, Wright J, Perentesis J, Lingren T, Deleger L, Kaiser M, Kohane I, Solti I. Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients. *BMC medical informatics and decision making*. 2015; 15:1-0.
- [41] Gardner SP. Ontologies and semantic data integration. *Drug discovery today*. 2005; 10(14):1001-7.
- [42] Zhang J, El-Gohary NM. Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking. *Journal of computing in civil engineering*. 2016; 30(2):04015014.
- [43] Bommarito II MJ, Katz DM, Detterman EM. LexNLP: Natural language processing and information extraction for legal and regulatory texts. In *Research handbook on big data law 2021* (pp. 216-227).
- [44] Haerian K, Varn D, Vaidya S, Ena L, Chase HS, Friedman C. Detection of pharmacovigilance-related adverse events using electronic health records and automated methods. *Clinical Pharmacology & Therapeutics*. 2012; 92(2):228-34.
- [45] Cejas OA, Azeem MI, Abualhaija S, Briand LC. Nlp-based automated compliance checking of data processing agreements against gdpr. *IEEE Transactions on Software Engineering*. 2023; 49(9):4282-303.
- [46] Butler T, O'Brien L. Artificial intelligence for regulatory compliance: Are we there yet?. *Journal of Financial Compliance*. 2019;3(1):44-59.
- [47] Wan X, Xiao J. Exploiting neighborhood knowledge for single document summarization and keyphrase extraction. *ACM Transactions on Information Systems (TOIS)*. 2010; 28(2):1-34.
- [48] Guimaraes CV, Grzeszczuk R, Bisset III GS, Donnelly LF. Comparison between manual auditing and a natural language process with machine learning algorithm to evaluate faculty use of standardized reports in radiology. *Journal of the American College of Radiology*. 2018; 15(3):550-3.
- [49] Maitra A, Annervaz KM, Jain TG, Shivaram M, Sengupta S. A novel text analysis platform for pharmacovigilance of clinical drugs. *Procedia Computer Science*. 2014; 36:322-7.