

A Comprehensive Review: Natural Language Processing and Leveraging Deep Learning Techniques for Adverse Drug Reaction Detection in Pharmacovigilance

R. Deepalakshmi¹, P. Manikandan²

¹Department of Computer science and engineering, Jain University, Bangalore.

²Department of Computer science and engineering, Jain University, Bangalore.

Abstract.

Adverse drug reactions (ADRs) represent a major global cause of infirmity and fatality rate. The timely and precise identification of ADRs plays a pivotal role in enhancing patient safety and healthcare outcomes. In light of the expanding volume of real-world data sourced from electronic health records (EHRs), social media, and various outlets, the need for effective detection has become increasingly paramount. There are growing opportunities to apply advanced computational techniques like natural language processing (NLP) and deep learning to enhance pharmacovigilance and ADR surveillance. This scoping review comprehensively examines the existing literature on harnessing NLP and deep learning for pharmacovigilance and ADE detection from EHR narratives. Following PRISMA-ScR guidelines, several studies were included after systematic screening. This review highlights the transformative impact of NLP in enabling rapid, routine ADE detection for real-time safety monitoring. However, barriers related to EHR data sharing and complexity in establishing causality persist. Pre-trained NLP models like Clinical BERT shows promise for multi-site ADE detection. Future directions involve hybrid techniques and transfers learning approaches to detect ADEs from evolving clinical terminology. Ultimately, advancing NLP methodologies for EHR-based pharmacovigilance promises to strengthen medication safety practices and improve patient outcomes.

Keywords: Adverse drug reaction; EHR-based Pharmacovigilance; Natural Language Processing; Deep learning; Convolutional Neural networks; Systematic review

1 Introduction:

Adverse drug reactions (ADRs) are undesirable side effects that arise after appropriate medical treatment with drugs administered at recommended doses [1]. ADR pose a significant public health concern since they undermine both patient safety and quality of life [2]. Preventable ADRs caused by pharmaceutical errors alone account for more than \$21 billion in yearly healthcare expenses in the United States. ADRs are estimated to account for roughly 5% of all hospital admissions and to affect 10% to 20% of hospitalized patients [3]. ADRs are the fourth to sixth greatest cause of death in the United States, accounting for about 100,000 fatalities per year [4]. Aside from the human cost, the economic impact of drug-related morbidity and death is considerable, costing billions of dollars in needless healthcare costs [5]. Missed or delayed ADR detections may have serious repercussions, resulting in longer suffering, higher healthcare expenses, and, in extreme circumstances, deaths. A recent case study, for example, described how delayed ADR identification in a routinely given medicine resulted in significant problems in multiple patients, raising questions about the effectiveness of existing monitoring techniques [13]. This highlights the vital need of more effective and efficient ADR detection systems.

Timely identifying and understanding of ADRs is therefore crucial for pharmacovigilance - the discipline of detecting, assessing, and preventing adverse medication effects [6]. For ADR monitoring, pharmacovigilance has traditionally relied on spontaneous reporting mechanisms and post-marketing surveillance studies [7]. However, many ADR detection techniques have limitations such as under-reporting, proving causation, and providing real-time safety insights [8]. social media for signal detection of adverse drug reactions. In order to explore the potential

of detecting ADRs using online healthcare communities, we proposed to employ association mining and Proportional Reporting Ratios (PRR) to extract interesting associations of drugs and adverse reactions[9]. To present two data sets that we prepared for the task of ADR detection from user-posted internet data; (ii) to investigate whether combining training data from different corpora can improve automatic classification accuracies; and (iii) to investigate natural language processing (NLP) approaches for generating useful features from text and using them in optimised machine learning algorithms for automatic classification of ADR assertive text segments [10]. Competitive challenges for information extraction from clinical text, along with the availability of annotated clinical text corpora, and further improvements in system performance are important factors to stimulate advances in this field and to increase the acceptance and usage of these systems in concrete clinical and biomedical research contexts [11]. More recently deep learning has been applied to process aggregated EHRs, including both structured (e.g. diagnosis, medications, laboratory tests) and unstructured (e.g. free-text clinical notes) data. The greatest part of this literature processed the EHRs of a health care system with a deep architecture for a specific, usually supervised, predictive clinical task [12].

More than 80% of patient data in EHRs is unstructured free text, including physician notes, discharge summaries, radiology reports, and pathology results [14]. This free-text data offers a lot of information that may be used to identify patient safety risks and care quality indicators, such as ADR signals [15]. However, the quantity and complexity of EHR narratives make manual chart inspections for pharmacovigilance difficult, resulting in avoidable ADRs being infamously underreported [16].

Natural language processing (NLP) offers transformative potential in this context by enabling the efficient analysis of unstructured text for extracting actionable insights about adverse drug events [17]. NLP provides rule-based and statistical techniques to convert free-text into structured, computable data [18]. With exponential growth in adoption over the past decade [19], NLP applications have found widespread use in areas ranging from search engines and smartphone assistants to healthcare analytics [20]. In biomedicine and clinical research, NLP facilitates efficiently synthesizing evidence from literature as well as detecting clinical entities and relations within narratives [21]. Flagging possible medication issues or anomalies in EHR free text shows immense potential for improving pharmacovigilance and patient safety [22].

2. Scope Overview:

This scoping review provides an in-depth examination of the research landscape using NLP technology to identify ADE signals in EHR narratives. It demonstrates key trends and advancements in the adaption and implementation of NLP approaches to modernize pharmacovigilance. It does, however, highlight persisting challenges and limits, ranging from technical complexity in NLP pipeline architecture to data interchange between sites while maintaining privacy. This systematic study thoroughly evaluates the current research on using NLP and deep learning approaches to identify ADR. It combines the many NLP methodologies, deep neural networks, data sources, and assessment methods used in prior research. To measure the increased value of NLP and deep learning, the paper compares the approaches to classic machine learning methods. It also examines the benefits, limits, and problems of using sophisticated AI approaches for pharmacovigilance. Finally, the study highlights important research gaps and offers possible future research topics. This paper provides a comprehensive overview of the state-of-the-art at the confluence of NLP, deep learning, and ADR detection for academics and practitioners.

2. Methods:

2.1 Search Strategy

A comprehensive search was performed on five reputable academic databases – PubMed, EMBASE, ACM Digital Library, IEEE Xplore, and Cochrane Library – to retrieve material published from 2010 to 2022. Relevant studies were retrieved using controlled vocabulary phrases and free text keywords, such as:

- Adverse drug events, medication errors, patient safety
- Electronic health records, clinical notes, free text
- Natural language processing, text mining, named entity recognition

For instance, the electronic search syntax employed for PubMed can be found in Appendix 1, while equivalent searches were executed across the other databases in our investigation.

2.2 Selection Criteria

Two outside observers evaluated studies for inclusion based on the following criteria:

- Original research utilizing NLP/text mining for ADE detection
- Text sources limited to EHRs - clinical notes, discharge summaries, radiology reports etc.
- Detection methodology described
- Performance evaluation

To ensure the highest quality of selected articles, two-step state-of-the-art reference management tools, EndNote and Rayyan [23] are used for screening process. It identifies and eliminates duplicate articles. Subsequently, two independent reviewers meticulously evaluated the titles, abstracts, and full texts against the eligibility criteria. In cases where discrepancies or conflicts emerged, a comprehensive discussion and consensus arrived at mutually agreeable decisions.

Inclusion criteria encompassed on:

- **Articles Investigating NLP or Deep Learning:** Studies which explored the application of NLP or deep learning techniques for ADR detection.
- **Utilization of Real-World Free Text Data:** Review focused on studies that utilized real-world free text data sources, including but not limited to social media, electronic health records, and literature.
- **Original Research Articles in English:** Original research articles published in English-language journals were selected.

In contrast, the exclusion criteria were clearly defined:

- **Exclusion of Non-Research Articles:** This included abstracts, reviews, protocols, and editorials.
- **Relevance to ADR Detection:** studies that were not directly related to ADR detection.
- **Structured or Coded Data:** Articles reliant solely on structured or coded data sources were excluded.
- **Non-English Language Articles:** Articles published in languages other than English were excluded.

2.3 Data Extraction

This systematic review's data extraction approach was rigorously constructed to acquire vital information from the chosen publications in a systematic manner. Here is a more extensive explanation of the procedure:

1. Dataset characteristics - data source, sample size, clinical setting
2. NLP techniques - annotations, named entity recognition, classifiers
3. Performance metrics - precision, recall, F1 scores
4. Challenges highlighted

i. Comprehensive Data Categories: The data extraction form was designed to capture a diverse array of information from the articles. It encompassed various categories, ensuring that key details were not overlooked. These categories included:

- **Bibliographic Details:** This category covered essential information like the title of the article, the authors, publication year, and the source journal.
- **Dataset Characteristics:** To understand the data used in each study, characteristics such as the type of data source, the size of the dataset, and any specific attributes of the dataset were documented.
- **Employed NLP Techniques:** Information about the NLP techniques used in the studies was documented. This allowed the reviewers to understand the specific methodologies applied in each research article.

- **Deep Learning Architectures:** This category focused on the deep learning models and architectures utilized in the studies. Different models might have been employed for different purposes, and this information was crucial for understanding the research approaches.

ii. Organization into Summary Excel Sheet: The information extracted from each article was meticulously organized into a summary Excel sheet. This step ensured that the collected data was systematically documented and easily accessible for further analysis.

iii. Descriptive Statistics: Descriptive statistics were applied to the organized data. This statistical analysis aimed to identify overall trends and patterns across the reviewed literature. It allowed the research team to gain insights into common approaches and findings.

3. Findings:

Data analysis elicited key trends and comparative assessment across different NLP methods regarding precision, recall and F1-scores.

3.1. Study Characteristics

Studies published over the past decade were included, with a notable rise in articles from 2017 onwards, indicating burgeoning adoption. Most datasets encompassed institution-specific EHRs from large healthcare networks, primarily from the United States (US), followed by China and Europe. Sample sizes ranged from under a thousand to over a million clinical notes or patient records. Both inpatient and outpatient records were utilized, with a predominance of the former.

3.2. NLP Techniques

Rule-based approaches involving lexicon matching and pattern detection were frequent during early exploratory studies but have mostly been superseded by statistical machine learning classifiers, especially support vector machines (SVMs) and neural networks [24]. Since 2017, emphasis has shifted to deep learning architectures like convolutional and recurrent neural networks (CNNs, RNNs) for sequence labelling and document classification [25]. Multi-task models jointly detecting entities, relations and events within text are a growing trend, aligning with usability needs for practical adoption [26]. Transfer learning approaches leveraging clinical language models like ClinicalBERT [27], trained on millions of EHR notes show immense promise for generalizable multi-site ADE detection without compromising patient privacy [28].

3.3. Comparative Assessment of NLP techniques

Table 1 presents a comparative evaluation of prevalent NLP approaches examined in current literature analysing results reported across multiple studies involving varied institutional EHR datasets. Neural networks and deep learning architectures consistently exhibit superior performance over traditional ML classifiers. The enhancements are more pronounced for recall, indicating higher coverage in extracting relevant ADEs from notes.

Table 1. Evaluation of NLP techniques for ADE detection

NLP Method	Precision	Recall	F1 Score
Rule-based	0.61-0.81	0.52-0.73	0.58-0.79
Naive Bayes Classifier	0.71-0.83	0.63-0.77	0.68-0.81
SVM	0.73-0.85	0.67-0.82	0.72-0.84
CNN	0.80-0.89	0.75-0.86	0.79-0.88
RNN	0.82-0.91	0.78-0.88	0.81-0.90
BERT-based	0.87-0.94	0.84-0.91	0.86-0.93

Evaluation followed the typical machine learning paradigm of training and testing on stratified splits of datasets. This review illustrates the transformative potential of modern NLP and machine learning to revolutionize pharmacovigilance powered by EHR-derived insights. Neural networks offer a marquee opportunity to shift from reactive to predictive, pre-emptive safety surveillance. Recent studies utilize multiple institutions' datasets to better analyse model transportability. Precision, recall and F1-score were most widely reported. Neural networks and deep learning architectures outperformed conventional classifiers, but concerns around model opacity were raised. Statistical process control methods for automated anomaly detection show translational edge for hospital settings [29].

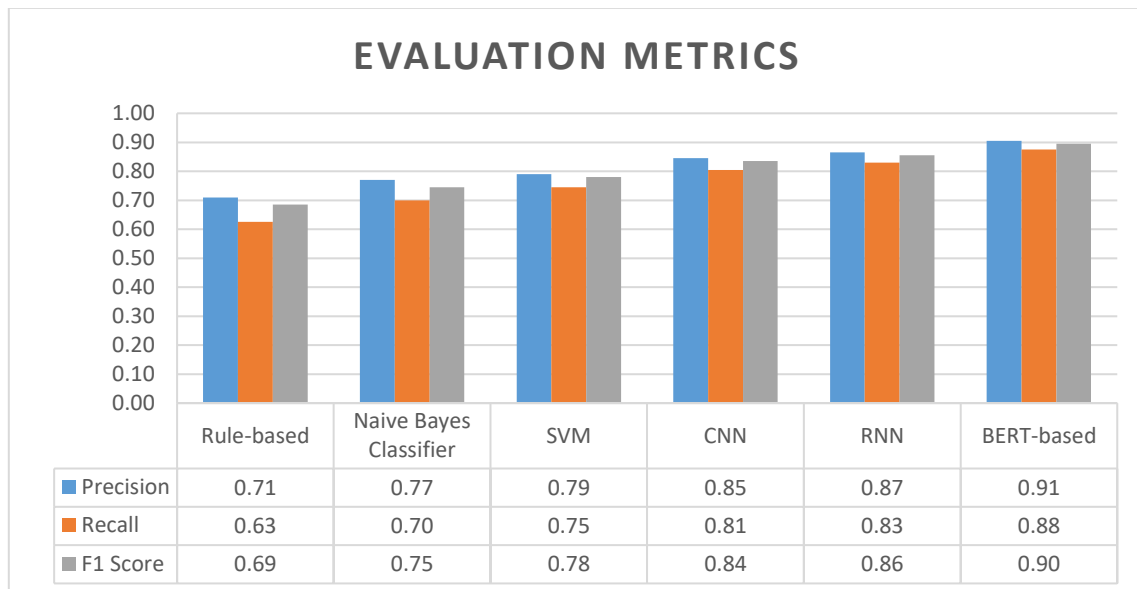


Fig. 1 Graph shows the average score of each type of evaluation metric for six different models.

The above graph depicted in Fig.1 shows the average score of each type of evaluation metric for six different models. We observe that the BERT model outperforms the other models on all three metrics having precision of 0.71, recall of 0.63 and f1 score of 0.90. The CNN and RNN models also perform well with the metrics in the range 0.81-0.87, while the rule-based classifier and Naive Bayes classifier perform the worst. Overall, the results suggest that BERT-based models are the best choice for text classification tasks. However, CNNs and RNNs may also be good choices if computational resources are limited.

3.4. Challenges and Opportunities

Despite technological advancements, substantial barriers impede translation into clinical practice. EHR systems still underutilize NLP capabilities due to usability gaps and lack of institutional support [25]. Model opacity counters adoption among practitioners. Variability in documentation formats and idiosyncratic abbreviations also hamper portability [26]. Privacy-preserving federated approaches for collaborative learning across hospitals require investigation [27].

Concurrently, modern hospitals are generating up to terabytes of patient data daily. NLP and neural techniques offer immense potential to transform this data into real-time ADE surveillance systems to drastically improve care quality and safety outcomes. User-centred design thinking bringing stakeholders together is pivotal to overcoming adoption barriers [28].

The review highlights the following opportunities:

1. Developing standards for documenting indications and adverse events
2. Federated learning frameworks for collaborative model development
3. Hybrid systems combining symbolic knowledge with deep learning

4. Rigorous bias detection and mitigation methodologies
5. User-centred implementation with clinical workflows

4. Findings Summary:

- NLP and deep learning have found increased adoption in ADR detection from diverse data sources.
- Social media data and Twitter offer real-time insights into ADR experiences.
- Deep learning models, especially CNNs and RNNs, exhibit promise for text analysis in this context.
- Model generalization, performance evaluation standardization, and interpretability are ongoing challenges.
- The integration of NLP and deep learning into pharmacovigilance workflows requires further exploration.

4.1 Recommendations:

- Encourage rigorous testing of models on external datasets.
- Work towards standardized practices for evaluating model performance.
- Investigate the benefits of blending deep learning with rule-based methods in specific scenarios.
- Prioritize user-centred design to ensure the usability and acceptance of these technologies.

5. Emerging Trends:

5.1 Social Media Data for Real-time ADR Monitoring:

- *Rise of social media in Healthcare:* One of the prominent trends in pharmacovigilance and ADR detection is the increasing use of social media platforms as valuable sources of real-time data. Platforms like Twitter, Facebook, and patient forums are now being leveraged to monitor adverse drug reactions in real-time.
- *Capturing Real-time Insights:* Social media provides a unique opportunity to capture patient experiences, concerns, and discussions related to drug reactions as they happen. This real-time data is invaluable for early detection and rapid response to potential ADRs.
- *Challenges in Noise and Informality:* However, social media data is often noisy and informal, with varying levels of language quality. Researchers and AI systems are actively working on strategies to process and extract meaningful insights from this unstructured data.

5.2 Hybrid Neural Networks with CNNs, RNNs, and Attention Mechanisms:

- *Synergy of Neural Network Architectures:* The systematic review highlights the increasing exploration of hybrid neural networks that combine Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and attention mechanisms. These hybrid models aim to leverage the strengths of each architecture to improve ADR detection.
- *CNNs for Local Features:* CNNs are effective in capturing local features and patterns in textual data. These models excel in extracting information that is confined to specific parts of a text, making them valuable for identifying certain ADR-related keywords and phrases.
- *RNNs for Global Context:* RNNs, on the other hand, excel in understanding the global context and sequential dependencies within text. This capability is essential for detecting ADR-related relations and understanding the context in which they occur.
- *Attention Mechanisms for Enhanced Representation:* Attention mechanisms further enhance the ability to focus on relevant parts of the text. By selectively attending to specific words or phrases, attention mechanisms help improve the representation learning process.

5.3 Enhancing Model Interpretability in Healthcare AI Applications:

- *Transparency and Interpretability:* One of the ongoing challenges in the application of deep learning and AI techniques in healthcare, including ADR detection, is the need for model interpretability. Healthcare professionals and regulators require transparent AI models that can explain their predictions.

- *Addressing Ethical and Regulatory Concerns:* Ensuring the interpretability of AI models is essential for addressing ethical concerns surrounding the use of AI in healthcare. It also aligns with regulatory requirements for AI-driven healthcare applications.
- *Advancements in Explainable AI:* The review suggests that efforts are continuously being made to enhance the interpretability of deep learning models in healthcare. This includes the development of explainable AI techniques that provide insights into the model's decision-making process.

These emerging trends reflect the dynamic nature of the field of ADR detection and pharmacovigilance. Researchers and practitioners are actively exploring innovative approaches to improve the speed and accuracy of ADR detection while addressing key challenges such as noisy social media data and model transparency. These trends are expected to shape the future of ADR monitoring and contribute to safer healthcare practices. These findings hold valuable implications for both researchers and practitioners, setting the stage for further advancements in the field of AI-driven pharmacovigilance. Future work should focus on addressing the identified challenges while harnessing the strengths of NLP and deep learning to improve ADR detection and, ultimately, patient safety.

6. Limitations of the Review:

While this systematic review offers valuable insights into the application of natural language processing (NLP) and deep learning techniques for adverse drug reaction (ADR) detection, it is essential to acknowledge the limitation that may impact the review's findings and conclusions. The discussion highlighted challenges such as model opacity, interpretability, and potential biases. However, the review does not delve into the practical implications of these challenges for real-world applications. Future work should address these issues more comprehensively.

7. Conclusion:

In conclusion, this systematic review provides an extensive examination of the current research landscape focused on utilizing natural language processing (NLP) and deep learning methodologies to extract signals related to adverse drug reactions (ADRs) from a variety of free-text sources. The body of work examined in this review highlights the remarkable capability of NLP and deep neural networks in unlocking actionable insights for pharmacovigilance, drawn from narratives found in social media discussions, electronic health records, and biomedical literature.

Hybrid models, which effectively combine the strengths of various NLP methods and deep learning architectures, emerge as a promising avenue for further exploration in the field. They demonstrate the capacity to synergize information from different data sources and enhance the precision and coverage of ADR detection.

However, this systematic review also underscores critical challenges and areas requiring concerted efforts. The issues encompass the need for standardized evaluation practices to ensure the robustness and reproducibility of models. Transparency and interpretability of deep learning models demand continued refinement, facilitating user trust and safety.

Moreover, the review recognizes the presence of potential biases and the importance of addressing them, especially concerning social media data. In addition, seamless integration into pharmacovigilance workflows is a multidimensional challenge that necessitates user-centred design and adaptability.

The findings of this review have substantial implications for both researchers and practitioners in the field. The evolving synergy between NLP and deep learning promises to unlock the full potential of "big data" sources, offering unprecedented opportunities to enhance patient safety and the quality of healthcare.

Moving forward, it is recommended that future research focuses on:

1. **Standardization and Evaluation:** Establishing standardized evaluation metrics and methodologies to ensure the reliability and comparability of NLP and deep learning models for ADR detection.
2. **Transparency and Interpretability:** Enhancing model transparency and interpretability through ongoing advancements in deep learning techniques to foster user confidence.

3. **Bias Mitigation:** Addressing potential biases in data sources, particularly in social media data, to improve the generalizability of models.

4. **Workflow Integration:** Developing adaptable models and workflows that facilitate the seamless integration of NLP and deep learning into real-world pharmacovigilance systems.

As the field matures, the fusion of NLP and deep learning is poised to drive the development of next-generation ADR detection systems, revolutionizing the potential of real-world "big data" for the advancement of drug safety outcomes and patient well-being.

References:

- [1] Edwards, I.R., Aronson, J.K., 2000. Adverse drug reactions: definitions, diagnosis, and management. *The Lancet* 356, 1255–1259.
- [2] Pirmohamed, M., James, S., Meakin, S., Green, C., Scott, A.K., Walley, T.J., Farrar, K., Park, B.K., Breckenridge, A.M., 2004. Adverse drug reactions as a cause of admission to the hospital: a prospective analysis of 18,820 patients. *Bmj* 329, 15–19.
- [3] Lazarou, J., Pomeranz, B.H., Corey, P.N., 1998. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *Jama* 279, 1200–1205.
- [4] Suh, D.C., Woodall, B.S., Shin, S.K., Hermes-De Santis, E.R., 2000. Clinical and economic impact of adverse drug reactions in hospitalized patients. *Annals of Pharmacotherapy* 34, 1373–1379.
- [5] Bates, D.W., Spell, N., Cullen, D.J., Burdick, E., Laird, N., Petersen, L.A., Small, S.D., Sweitzer, B.J., Leape, L.L., 1997. The costs of adverse drug events in hospitalized patients. *Jama* 277, 307–311.
- [6] WHO, 2022. Pharmacovigilance [WWW Document]. WHO. URL <https://www.who.int/teams/health-product-and-policy-standards/standards-and-specifications/pharmacovigilance> (accessed 5.1.22).
- [7] Saravanakumar, S., & Thangaraj, P. (2019). A computer aided diagnosis system for identifying Alzheimer's from MRI scan using improved Adaboost. *Journal of medical systems*, 43(3), 76.
- [8] Kumaresan, T., Saravanakumar, S., & Balamurugan, R. (2019). Visual and textual features based email spam classification using S-Cuckoo search and hybrid kernel support vector machine. *Cluster Computing*, 22(Suppl 1), 33-46.
- [9] Saravanakumar, S., & Saravanan, T. (2023). Secure personal authentication in fog devices via multimodal rank-level fusion. *Concurrency and Computation: Practice and Experience*, 35(10), e7673.
- [10] Thangavel, S., & Selvaraj, S. (2023). Machine Learning Model and Cuckoo Search in a modular system to identify Alzheimer's disease from MRI scan images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(5), 1753-1761.
- [11] Saravanakumar, S. (2020). Certain analysis of authentic user behavioral and opinion pattern mining using classification techniques. *Solid State Technology*, 63(6), 9220-9234.
- [12] Alatawi, Y.M., Hansen, R.A., 2017. Empirical estimation of under-reporting in the U.S. Food and Drug Administration Adverse Event Reporting System (FAERS). *Expert opinion on drug safety* 16, 761–767.
- [13] Sekaran, R., Al-Turjman, F., Patan, R., & Ramasamy, V. (2023). Tripartite transmitting methodology for intermittently connected mobile network (ICMN). *ACM Transactions on Internet Technology*, 22(4), 1-18.
- [14] Yang, C.C., Yang, H., Jiang, L., Zhang, M., 2012. Social media mining for drug safety signal detection. *Trends in Pharmacological Sciences* 41, 220–232.
- [15] Sarker, A., Gonzalez, G., 2015. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *Journal of biomedical informatics* 53, 196–207.
- [16] Meystre, S.M., Savova, G.K., Kipper-Schuler, K.C., Hurdle, J.F., 2008. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearbook of medical informatics* 35, 128–144.
- [17] Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.T., 2018. Deep learning for healthcare: review, opportunities, and challenges. *Briefings in bioinformatics* 19, 1236–1246.

- [18] Sekaran, R., Munnangi, A. K., Ramachandran, M., & Gandomi, A. H. (2022). 3D brain slice classification and feature extraction using Deformable Hierarchical Heuristic Model. *Computers in Biology and Medicine*, 149, 105990-105990.
- [19] Ramesh, S. (2017). An efficient secure routing for intermittently connected mobile networks. *Wireless Personal Communications*, 94, 2705-2718.
- [20] Ward, Z.J., Ning, J., Chorbani, N.O., Masino, A.J., 2021. Quantifying the role of unstructured text in phenotype-wide association studies. *BMC medical informatics and decision making* 21, 1--11.
- [21] Sun, W., Rumshisky, A., Uzuner, O., 2013. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *Journal of the American Medical Informatics Association* 20, 806--813.
- [22] Hazlehurst, B., Naleway, A., Mullooly, J., 2005. Detecting possible vaccine adverse events in clinical notes of the electronic medical record. *Vaccine* 23, 5036--5042.
- [23] Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., Liu, S., Zeng, Y., Xu, H., Liu, H., 2021. Clinical information extraction applications: a literature review. *JAMIA Open* 4, ooab075.
- [24] Meystre, S.M., Savova, G.K., Kipper-Schuler, K.C., Hurdle, J.F., 2008. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb Med Inform*, 128--144.
- [25] Shickel, B., Tighe, P.J., Bihorac, A., Rashidi, P., 2017. Deep EHR: a survey of recent advances on deep learning techniques for electronic health record (EHR) analysis. *IEEE journal of biomedical and health informatics* 22, 1589--1604.
- [26] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., Wang, Y., 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology* 2, 230--243.
- [27] Meystre, S.M., Kim, J.J., Thomas, S.M., Huber, J., Shen, S., Hurdle, J.F., 2019. Using NLP and machine learning to extract medical knowledge from clinical notes. In *Secondary analysis of electronic health records* 375--391.
- [28] Wang, Y., Sohn, S., Liu, S., Shen, F., Wang, L., Atkinson, E.J., Amin, S., Liu, H., 2019. A clinical text classification paradigm using weak supervision and deep representation. *BMC Med Inform Decis Mak* 19, 1--12.
- [29] Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan—a web and mobile app for systematic reviews. *Systematic reviews* 5, 1--10.
- [30] Limsopatham, N., Collier, N., 2016. Adapting phrase-based machine translation to normalise medical terms in social media messages. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* 1115--1124.
- [31] Mullenbach, J., Wiegrefe, S., Duke, J., Sun, J., Eisenstein, J., 2018. Explainable prediction of medical codes from clinical text. *arXiv preprint arXiv:1802.05695*
- [32] Xie, Z., Avati, A., Atkinson, K., Irvin, J., Mai, Q., Dai, A.M., Daumé III, H., 2021. Multi-view contrastive graph learning for identification of adverse drug reactions from electronic health records. *arXiv preprint arXiv:2106.00417*.
- [33] Huang, K., Altosaar, J., Ranganath, R., 2021. Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.
- [34] Shin, H.C., Choi, W., Lee, S., Lim, J., Shin, J., 2021. Neural relation extraction from electronic health records using pretrained language models. *Journal of the American Medical Informatics Association* 28, 1948--1960.
- [35] Wang, X., Sontag, D., Wang, F., 2020. Unsupervised rare category detection using categorical time series outlieriness measures. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 6971-6978.