International Journal of Pharmacy and Pharmaceutical Research (IJPPR)
Volume 31, Issue 5, May 2025 **ijppr.humanjournals.com** ISSN: 2349-7203

AI- Powered Pharmaceutical Informatics: Natural Language Processing in Drug Development and Drug Delivery

Nandhini. A*[1], Dr. Senthil Rajan[2], Nithya.L[3], Vikhash.T[4], Bhuvaneshwaran R.M[5]

[1,2,3,4,5] – Department of Pharmaceutics, Swamy Vivekanandha college of pharmacy, Tiruchengode. India.

Received: 2025-04-25 Revised: 2025-05-05 Accepted: 2025-05-10

ABSTRACT

The pharmaceutical industry is undergoing a transformative shift with the integration of Artificial Intelligence (AI) and Natural Language Processing (NLP) in drug development and delivery. This review explores the application of AI-powered pharmaceutical informatics to revolutionize the drug discovery process. By leveraging NLP techniques, we aim to unlock the potential of unstructured data in the form of free-text, enabling the development of predictive models that can accelerate drug development and delivery. Our approach harnesses the power of machine learning (ML) and deep learning (DL) to analyze large volumes of text data, including scientific literature, clinical trials, and patient reports. This allows for the identification of patterns, relationships, and insights that can inform critical decisions in drug development, such as target identification, lead optimization, and clinical trial design. The review's outcome is expected to improve the efficiency, accuracy, and cost-effectiveness of the drug development process, ultimately leading to faster access to life-saving medications.

Keywords: AI-powered pharmaceutical informatics, Natural Language Processing (NLP) , Unstructured data analysis, Predictive modeling, Drug development

INTRODUCTION:

AI is the branch of computer science that deals with teaching computers to perform tasks ordinarily performed by humans. Machine learning (ML) is a branch of AI that deals with developing models for automated prediction of a given task. Within ML, modeling can be performed using neural networks, which have the ability to learn from large amounts of data without explicitly defined features, an area known as deep learning (DL) .While several Natural Language processing (NLP) techniques that do not utilize modeling such as ML or DL exist, some NLP can be used to perform modeling with ML or DL, using free-text (raw or processed) as input rather than images or pre-defined features [(1)]. The pharmaceutical industry is a critical field that plays a vital role in saving lives. It operates based on continuous innovation and the adoption of new technologies to address global healthcare challenges and respond to medical emergencies, such as the recent pandemic. In the pharmaceutical industry, innovation is typically predicated on extensive research and development across various domains, including but not limited to manufacturing technology, packaging considerations, and customer-oriented marketing strategies. Novel pharmaceutical innovations are range from small drug molecules to biologics, with a preference for better stability with high potency to fulfill unmet needs to treat diseases [(2)] . AI has significantly impacted the field of drug discovery, particularly in the areas of target identification and validation. This process involves identifying potential biological targets and elucidating their roles in diseases, followed by validating these targets to ensure they are directly involved in a disease mechanism and that the modulation of the target is likely to have a therapeutic effect and plays a crucial role in identifying potential drug gets by analyzing the genomic, proteomic, and metabolomic data [(3)]. The need for a proficient workforce in the healthcare industry is persistent, necessitating the continuous provision of training to healthcare personnel to augment their involvement in routine duties. Identifying skill gaps in the workplace is a crucial undertaking within the pharmaceutical industry. The global outbreak of corona virus disease 2019 (COVID-19) has caused significant disruptions to various operations worldwide, including ongoing clinical trials. The implementation of AI is poised to bring about a significant transformation in the way the pharmaceutical industry handles supply chain operations. It also consolidates numerous AI research endeavors from recent decades to create effective solutions for diverse supply chain issues. Additionally, the study suggests potential research areas that could enhance decision-making tools for supply chain management in the future [(2)]. Artificial Intelligence (AI) is the over arching field that encompasses developing systems capable of performing tasks that traditionally require human intelligence. The set asks include but are not limited to problem-solving , understanding natural language , recognizing patterns, and learning from experience . Machine Learning (ML) , a subfield of AI, involves developing algorithms and statistical models that allow computers to perform tasks without being explicitly programmed to do so. Instead, they learn and improve from the data they process. Deep learning (DL) is a subset of ML inspired by the structure and function of the human brain. It uses artificial neural networks, especially deep neural networks, to learn from vast amounts of data .Within AI, Natural Language Processing (NLP) plays a crucial role in the interaction between computers and human language[4] .

There are three types of artificial intelligence:

1. General artificial intelligence - this type of intelligence, theoretically, can perform the tasks that human intelligence can perform.

2. Narrow general intelligence - this type of intelligence performs tasks such as human intelligence, but in a narrow range.

3. Artificial super intelligence - this intelligence is superior to human thinking in all areas and activities.

In many medical specialties, artificial intelligence is constantly growing. It uses different devices, applications, and algorithms to prevent, diagnose and treat conditions. Two major types of devices are used:

1. Natural language processing methods (NLP) that have the role of extracting information from unstructured data, such as medicated journals, and medical notes. 2. Machine learning techniques (ML) that analyze structured data, such as genetic data and images, represent a secondary screening and diagnosis technique. Machine learning techniques allow computers to make predictions by repeatedly learning existing materials [5] .

AI systems can predict drug toxicity by analyzing the chemical structure and char acteristics of compounds. Machine learning algorithms trained on toxicology databases can anticipate harmful effects or identify hazardous structural properties. This helps re searchers prioritize safer chemicals and mitigate potential adverse responses in clinical trials. Overall, AI-driven approaches in drug research and development offer the potential to streamline and expedite the identification, optimization, and design of novel therapeutic candidates, ultimately leading to more efficient and effective medications [2] .

NATURAL LANGUAGE PROCESSING (NLP)

Natural language processing (NLP) is a technique within the broader sphere of Artificial Intelligence (AI) that has been rapidly expanding over the last 20 years with our increasing use of computer technology. NLP is used daily by every smart phone user, and has contributed to the development of language translation, personal digital assistants, and voice-controlled home automation systems. The technique is now being increasingly used throughout medicine to improve utilization of unstructured electronic health records (EHRs) and to provide a form of communication with patients to answer questions and conduct consultations [6] . We use the term natural language processing (NLP) to refer to the field that aims to enable computers to parse human language as humans do. NLP is not a single technique; rather, it is composed of many techniques grouped together by this common aim [7] . A language is a mode of communication, either verbal or written, that consists of structured words, sets of symbols (such as digits, letters and special characters) as well as sets of rules which govern the composition and manipulation of the words and symbols in a conventional way. Natural Languages (NL) are therefore languages that are either spoken or written by human beings for communication [8] .

NLP use in critical care has been focused on improving EHRs, predicting patient outcomes, and enabling identification of patients suitable for critical care trials . To improve EHRs, NLP has been used to extract information from patient notes to generate a more complete problem list. This is particularly powerful as many free text notes within the HER can be analyzed to suggest diagnoses or events that the clinician can consider including in the problem list. One study showed that the sensitivity of the problem list could be improved from 8.9% to 77.4% with the NLP model, with clear potential to improve patient safety and reduce delays and costs [6] .

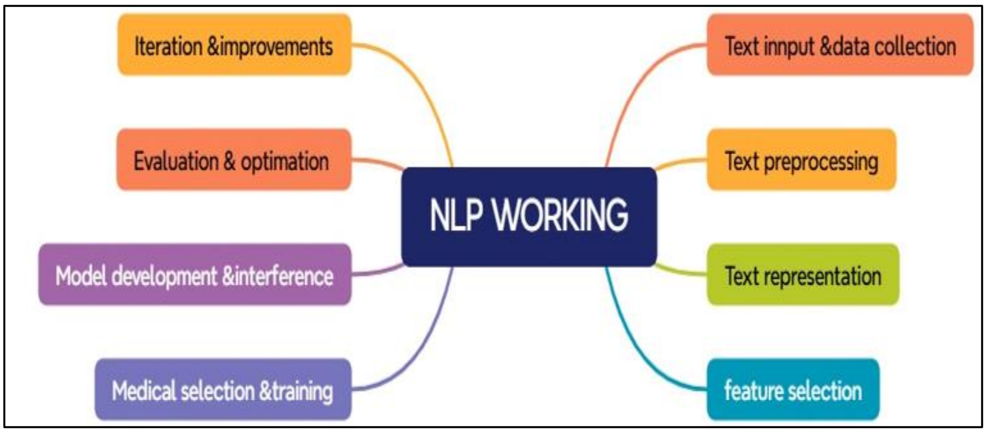

Figure No 1: Nlp Working Process

APPLICATIONS OF NATURAL LANGUAGE PROCESSING IN MEDICINE:

1. Research uses of NLP have included tools to search for relevant clinical trials in large databases and applications to streamline drug discovery such as predicting targets and identifying adverse events.
2. For direct patient care, NLP has been shown to be successful in predicting hospital admission of patients from the Emergency Department, allowing augmentation of the existing triage process and consequent improvement of clinical outcomes.
3. The technology is also successful in a diagnostic setting where it has been used to classify radiology reports to identify the appropriate clinical response, reducing human input.
4. Patient-facing NLP applications are now being developed. Natural language understanding (NLU) and Natural language generating (NLG) chat bots are becoming a large part of the development of NLP applications in healthcare, with the most widely used function being through mobile phone applications such as abylon Health and Health Tap, providing a more efficient and user friendly interface than a search engine [6].

NATURAL LANGUAGE PROCESSING USED IN DIFFERENT DRUG DELIVERY SYSTEMS:

1. OCULAR DRUG DELIVERY SYSTEM:

Adoption of electronic health records (EHRs) and advances in ocular imaging technology have revolutionized healthcare delivery in ophthalmology and resulted in significantly increased data available for clinical care and research. The American Academy of Ophthalmology (AAO) and National Institutes of Health (NIH) have supported this movement with the development of large, processed EHR based datasets such as the Intelligent Research in Sight (IRIS) Registry and the All of Us research program . Research efforts using these datasets have largely focused on retrospective association analysis and trends in care. Large datasets have also been used to develop predictive artificial intelligence (AI) models. The majority of these applications within ophthalmology have focused on image based AI including diagnosis of diabetic retinopathy, age-related macular degeneration, retinopathy of prematurity , and glaucoma among others [1]. Ophthalmologists have conventionally been trained to fear the term 'NLP', for it indicates 'no light perception' – a level of vision where the patient is effectively blind. In a more modern context, and outside of ophthalmology, however, NLP has grown to most commonly mean 'natural language processing' – a subset of machine intelligence focused on the interaction of human language with computer systems. For as long as computers have existed, NLP has been an area of interest, with Alan Turing's proposal of what is now called the 'Turing Test' – an experiment for determining whether the language generated by a computer is distinguishable from that produced by a human – having been in place since 1950 [9].

The rapid adoption of electronic health records (EHRs) in recent decades has generated large volumes of clinical data with potential to support secondary use in research. Indeed, a recurring justification for HER adoption has been to support the collection and analysis of "big data" to gain meaningful insights. The clinical research community has expressed growing interest in developing effective technique store use clinical data from EHRs, in part because of the benefits of secondary data reuse over primary data collection. However, there are challenges associated with reusing HER data, particularly because of its complexity and heterogeneity. For example, in ophthalmology, patient data contained in EHRs may include fields as diverse as demographic information, diagnosis , laboratory tests, prescriptions, eye examinations, imaging, and surgical records. Interpreting these heterogeneous data requires strategies such as information extraction, dimension reduction, and predictive modeling typical of machine learning and, more broadly, artificial intelligence (AI) techniques [10] .

2. CARDIO VASCULAR DRUG DELIVERY SYSTEM:

Heart failure (HF) is a global public health problem that affects an estimated 64 million people worldwide. In the US alone, more than 6 million patients have been diagnosed with HF and in excess of 1 million annual hospital admissions have had HF as a primary discharge diagnosis. Despite the increasing burden of hospitalizations for worsening HF (WHF), our understanding of epidemiological factors, patient clinical characteristics, care management, and outcomes in patients admitted for WHF is based almost entirely on observational studies relying on administrative or claims data and/or national reporting databases or quality improvement registries. However, these data sources are intrinsically limited by the accuracy and completeness of diagnostic coding and/or voluntary reporting. Furthermore, signs and symptoms of HF inherently lack sensitivity and specificity, and overlap occurs with other common cardiovascular and non cardiovascular conditions, raising the possibility of misclassification [11].

Heart disease is the leading cause of death in the United States, the UK, and worldwide. It causes more than 73,000 and 600,000 deaths per year in the UK and the US, respectively. Heart disease caused the death of about 1 in 6 men and 1 in 10 women. Heart disease has a number of common forms such as Coronary Artery Disease (CAD). According to the World Health Organization, risk factors of a specific disease are any attributes that raise the probability that a person may get that disease. There are several risk factors for CAD and heart disease such as Diabetes, CAD, Hyperlipidemia, Hypertension, Smoking, Family history of CAD, Obesity, and Medications associated with the mentioned chronic diseases. Each heart risk factor should be specified with indicator and time attributes except for a family history of CAD and smoking status. Each indicator attribute reflects the implications of the risk factor in the clinical text. It is essential to detect risk factors mentioned in narrative clinical notes for heart disease prediction and prevention which is considered an important challenge. Our goal is to develop a model that can detect and predict the progression of heart disease and CAD from clinical notes. The prediction of heart disease risk factor using clinical and statistical approaches has attracted a lot of attention over the past ten years because this process is very complex. Several techniques have been applied to clinical concept extraction such as simple pattern matching, statistical systems, and machine learning. Although these techniques have achieved better results, it is difficult to apply such statistical models to analyze the EHR data due to the time-consuming process of processing large amounts of data, their usage of several statistical and structural assumptions, and custom features/markers [12].

The development or application of NLP methods for clinical text (EHRs), and were clinically focused on a patient population with existing cardiac disease. Studies were excluded for the following reasons:

(1) Not published or available in English,

(2) Duplicate studies,

(3) Aspects of the same study published by the same research group in multiple publications,

(4) Lacking a description of NLP methods or applications

(5) Focus on patients without existing cardiac disease, such as patients with cardiac risk factors but no diagnosed disease; while these do represent additional areas of interest in NLP, they were felt to be beyond the scope of this work. Following this strategy, three reviewers (MRT, AV, DS) performed two rounds of study selection: title and abstract screening followed by full-text review. Each article was screened by two independent reviewers and disagreements were discussed among the three reviewers until consensus was achieved [13].

The use of NLP word vectorization algorithms and logistic regression (LR) to predict eight ICD-10 codes related to common cardiovascular diseases from free text outpatient progress notes . We compared both interpretable models and less-interpretable models with regard to their performances on the ICD-10 code prediction tasks. The proposed models show good classification performance on eight ICD-10 codes in two Stanford cohorts and the models generalized well on the MIMIC-III (Medical Information Mart for Intensive Care III) dataset. Additionally, the most interpret able models also showed the best performance on all datasets [14].

Semi-structured data involve all sorts of text-based information such as medication use and dosage, as well as parameters collected on continuous scales as visual analogue scale scores. Imaging data are considered unstructured, given their large inter-individual variability. One way to reduce this variability is to collect this type of data in a standardized fashion, e.g. by using the same ECG device or identical cardiovascular magnetic resonance scanners with standardized imaging protocols. However, this would premise that the data analyzed have either been collected within prospective trials, or that retrospectively ascertained data derive from tertiary centers applying standardized, up-to-date technologies [15].

3. PULMONARY DRUG DELIVERY SYSTEM:

Chest computed tomography (CT) is widely used for lung cancer screening and respiratory disease diagnosis. With the increasing utilization of CT for both screening and diagnostic purposes, clinicians are identifying and managing patients with pulmonary nodules more frequently. In lung cancer screening studies, nodules were found in approximately 20% of participants, and a population-based study of diagnostic CT scans identified incidentally detected nodules in as many as 31% of all scans. Extrapolating from the latter finding, the authors estimated that approximately 1.6 million Americans would have a nodule identified annually. However, most pulmonary nodules are benign and harmless; only 5% of nodules are caused by lung cancer, and distinguishing malignant from benign nodules remains challenging (16).

Pulmonary function testing (PFT) provides objective, quantifiable measurements of lung function and is a cornerstone of diagnosis and monitoring. Spirometry, the most commonly used test, is the measurement of the movement of air into and out of the lungs during breathing. It includes the measured forced expiratory volume in one second (FEV1), forced vital capacity (FVC).

We developed a natural language processing (NLP) tool to extract FEV1 and FVC pre- and post BDC from medical notes in order to complement data obtained from the electronically captured BDC reports, to more completely assess bronchodilator response in our asthma population from the Rocky Mountain Network of the Veterans Affairs Health (VHA) System. The goal of this study was to validate the NLP against expert review for accuracy in identifying spirometric values and bronchodilator responsiveness, and to estimate the number of additional values identified with NLP plus structured data compared to the structured data alone (17).

COPD is the third leading cause of death worldwide but is often under diagnosed. Airway inflammation, air trapping and emphysema may be secondary to smoking or environmental exposures and people with COPD are at increased risk of respiratory infections and cancer. Studies report a twofold to fourfold increase in lung cancer risk in patients with COPD compared to those without airflow obstruction. Lung cancer is the highest mortality cancer in the US and is often discovered at distant stage. The National Lung Screening Trial (NLST) showed a 20% reduction in lung cancer mortality for subjects imaged with CT compared to chest radiograph. However, although lung cancer screening with low-dose computed tomography (LDCT) is now recommended, few patients actually receive CT screening exams: for 2010–2015 fewer than 4% and in 2016 fewer than 2% of those eligible (18). The NLP pipeline regresses raw text to the same two positive real values, FEV1/FVC and FEV1. We experiment with two strong NLP models: BiLSTM, which represents a standard design of recurrent neural networks, and RoBERTa, which represents recent state-of-the-art trans former architecture. With the BiLSTM model, we use the Common Crawl version of GloVe embeddings with 100 dimensions, combined with character-level embeddings convoluted by a 2-dimensional filter of size 5 (18).

CURRENT PHARMACEUTICAL CHALLENGES :

The Linguistic String Project-Medical Language Processor is one the large scale projects of NLP in the field of medicine. The LSPMLP helps enabling physicians to extract and summarize information of any signs or symptoms, drug dosage and response data with the aim of identifying possible side effects of any medicine while highlighting or flagging data items. The National Library of Medicine is developing The Specialist System. It is expected to function as an Information Extraction tool for Biomedical Knowledge Bases, particularly Medline abstracts (19). Evaluation of NLP systems is still typically performed with standard statistical metrics based on intrinsic criteria, not necessarily optimal for the clinical research problem at hand. To address such issues, it is important to identify which level of analysis is appropriate, and model the problem accordingly. Enriching informatics approaches with novel data sources, 280 using evaluation metrics that capture novel aspects such as model interpretability or time sensitivity, and developing NLP solutions with the clinical end-users in mind could lead to considerable advances in this field (20). A powerful branch of machine learning called deep learning has shown impressive results in deciphering complicated medical imaging data, including CT, MRI, and X-ray image . Radiologists and other healthcare professionals can better diagnose and plan treatments for diseases including cancer, cardiovascular disease, and neurological disorders by using deep neural networks, which provide accurate and efficient detection of abnormalities and diseases. The way medical professionals engage with enormous volumes of unstructured textual data, such as clinical notes, electronic health records (EHRs), and medical literature, has been completely transformed by natural language processing (NLP) technologies. In order to improve documentation accuracy, clinical coding, and information retrieval procedures, natural language processing (NLP) algorithms extract useful clinical information, identify trends in patient narratives, and aid in the semantic interpretation of medical papers (21).

CONCLUSION

The utility of NLP is promising in improving medication safety with regard to the ability to identify adverse events and in the efficiency in identifying these events. This review identifies NLP systems that have been designed to help improve medication safety while also identifying the challenges that clinical NLP introduces. Continued efforts to improve the utility of clinical NLP include addressing the challenges of characterizing the context of adverse events. Involvement of pharmacists in development of NLP systems will enhance the ability of these systems to improve medication safety [22].

REFERENCES

1. **Chen JS, Baxter SL**. Applications of natural language processing in ophthalmology: present and future. Frontiers in Medicine. **2022 Aug 8;9: 906554.**
2. **Vora LK, Gholap AD, Jetha K, Thakur RR, Solanki HK, Chavda VP.** Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics. **2023 Jul 10; 15(7):1916.**
3. **Visan AI, Negut I**. Integrating artificial intelligence for drug discovery in the context of revolutionizing drug delivery. Life. **2024 Feb 7; 14(2):233.**
4. **Rojas-Carabali W, Agrawal R, Gutierrez-Sinisterra L, Baxter SL, Cifuentes-González C, Wei YC, Arputhan AJ, Kannapiran P, Wong S, Lee B, de-la-Torre A.** Natural Language Processing in Medicine and Ophthalmology: A Review for the 21st-century clinician. Asia-Pacific Journal of Ophthalmology. **2024 Jul 25:100084.**
5. **Keskinbora K, Güven F.** Artificial intelligence and ophthalmology. Turkish journal of ophthalmology. **2020 Mar 5;50 (1):37.**
6. **Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB.** Natural language processing in medicine: a review. Trends in Anaesthesia and Critical Care. **2021 Jun 1;38:4-9.**
7. **Chary M, Parikh S, Manini AF, Boyer EW, Radeos M.** A review of natural language processing in medical education. Western Journal of Emergency Medicine. **2018 Dec 12;20(1):78.**
8. **Iroju OG, Olaleke JO.** A systematic review of natural language processing in healthcare. International Journal of Information Technology and Computer Science. **2015 Jul 8;8(8):44-50.**
9. **Nath S, Marie A, Ellershaw S, Korot E, Keane PA.** New meaning for NLP: the trials and tribulations of natural language processing with GPT-3 in ophthalmology. British Journal of Ophthalmology. **2022 Jul 1;106(7):889-92.**
10. **Lin WC, Chen JS, Chiang MF, Hribar MR**. Applications of artificial intelligence to electronic health record data in ophthalmology. Translational vision science & technology**. 2020 Jan 28;9(2):13-.**
11. **Ambrosy AP, Parikh RV, Sung SH, Narayanan A, Masson R, Lam PQ, Kheder K, Iwahashi A, Hardwick AB, Fitzpatrick JK, Avula HR**. A natural language processing–based approach for identifying hospitalizations for worsening heart failure within an integrated health care delivery system. JAMA Network Open. **2021 Nov 1;4(11):e2135152-.**
12. **Houssein EH, Mohamed RE, Ali AA.** Heart disease risk factors detection from electronic health records using advanced NLP and deep learning techniques. Scientific Reports**. 2023 May 3;13(1):7173.**
13. **Turchioe MR, Volodarskiy A, Pathak J, Wright DN, Tcheng JE, Slotwiner D.** Systematic review of current natural language processing methods and applications in cardiology. Heart. **2022 Jun 1;108(12):909-16.**
14. **Zhan X, Humbert-Droz M, Mukherjee P, Gevaert O.** Structuring clinical text with AI: Old versus new natural language processing techniques evaluated on eight common cardiovascular diseases. Patterns. **2021 Jul 9;2(7).**
15. **Lüscher TF, Wenzl FA, D'Ascenzo F, Friedman PA, Antoniades C.** Artificial intelligence in cardiovascular medicine: clinical applications. European heart journal. **2024 Oct 21;45(40):4291-304.**
16. **Zheng C, Huang BZ, Agazaryan AA, Creekmur B, Osuj TA, Gould MK.** Natural language processing to identify pulmonary nodules and extract nodule characteristics from radiology reports. Chest. **2021 Nov 1;160(5):1902-14.**
17. **Sauer BC, Jones BE, Globe G, Leng J, Lu CC, He T, Teng CC, Sullivan P, Zeng Q.** Performance of a natural language processing (NLP) tool to extract pulmonary function test (PFT) reports from structured and semistructured Veteran Affairs (VA) data. eGEMs. **2016 Jun 1;4(1):1217.**
18. **Schroeder JD, Bigolin Lanfredi R, Li T, Chan J, Vachet C, Paine III R, Srikumar V, Tasdizen T**. Prediction of obstructive lung disease from chest radiographs via deep learning trained on pulmonary function data. International Journal of Chronic Obstructive Pulmonary Disease. **2020 Jan 5:3455-66.**
19. **Khurana D, Koli A, Khatter K, Singh S.** Natural language processing: state of the art, current trends and challenges. Multimedia tools and applications. **2023 Jan;82(3):3713-44.**
20. **Velupillai S, Suominen H, Liakata M, Roberts A, Shah AD, Morley K, Osborn D, Hayes J, Stewart R, Downs J, Chapman W.** Using clinical natural language processing for health outcomes research: overview and actionable suggestions for future advances. Journal of biomedical informatics. **2018 Dec 1;88:11-9.**
21. **Ahire YS, Patil JH, Chordiya HN, Deore RA, Bairagi V.** Advanced Applications of Artificial Intelligence in Pharmacovigilance: Current Trends and Future Perspectives. J. Pharm. Res. **2024 Jan;23(1):23-33.**
22. **Wong A, Plasek JM, Montecalvo SP, Zhou L**. Natural language processing and its implications for the future of medication safety: a narrative review of recent advances and challenges. Pharmacotherapy: The Journal of sHuman Pharmacology and Drug Therapy. **2018 Aug;38(8):822-41.**

How to cite this article:
Nandhini. A et al. Ijppr.Human, 2025; Vol. 31 (5): 230-236.

Conflict of Interest Statement: All authors have nothing else to disclose.