SSRG International Journal of Computer Science and Engineering
ISSN: 2348–8387 / https://doi.org/10.14445/23488387/IJCSE-V12I15P106

Volume 12 Issue 5, 42-55, May 2025

Original Article

Pelican Optimization-Assisted Deep Learning Framework for Adverse Drug Reaction Detection from Twitter Data

R. Deepalakshmi[1], P. Manikandan[2]

[1,2]*Department of Computer Science, Jain University, Bangalore, India.*

[1]*Corresponding Author : deepskanch@gmail.com*

Received: 21 March 2025 Revised: 27 April 2025 Accepted: 16 May 2025 Published: 31 May 2025

Abstract - *Adverse Drug Reactions (ADRs) lead to significant patient safety hurdles, necessitating timely and accurate detection to support pharmacovigilance efforts. Traditional ADR reporting systems suffer from underreporting and delays, prompting the need for alternative data sources such as social media. However, extracting meaningful insights from unstructured and noisy social media text presents substantial challenges. This research introduces a novel Deep Convolutional Recurrent Semantic Similarity Model (DCR-SSM), which integrates convolutional and recurrent layers with a semantic similarity mechanism and attention module to enhance ADR detection from Twitter data. The framework incorporates a robust preprocessing pipeline tailored to social media text and a novel Decision Tree-based Pelican Optimization Algorithm (DT-POA) for feature selection and Bag-of-Words encoding to capture relevant linguistic and semantic features. Simulation of the proposed framework is assessed on the SMM4H dataset, indicating the superior performance of the proposed Model over state-of-the-art ADR detection methods. The DCR-SSM accomplished an accuracy of 75%, 72%, recall of 72%, and an F1-score of 73%, outperforming traditional machine learning (SVM) and deep learning models (LSTM, Bi-LSTM, CNN). Compared to best-performing existing models, the proposed framework improves precision by up to 5.2% and maintains a balanced trade-off between recall and F1-score, ensuring better generalization in real-world applications. These findings highlight the potential of leveraging NLP and deep learning for mining patient-reported ADRs from social media, offering a scalable and cost-effective alternative to conventional pharmacovigilance methods. Future research can further explore multi-lingual ADR detection and domain-specific embeddings to enhance detection accuracy and adaptability across diverse healthcare settings.*

Keywords - *Adverse Drug Reactions, Deep Learning, Natural Language Processing, Pharmacovigilance, Twitter data*.

1. Introduction

Adverse Drug Reactions (ADRs) are unintended, harmful physiological responses resulting from the administration of medications at therapeutic doses for approved indications. As a major pharmacovigilance concern, ADRs contribute significantly to morbidity, mortality, and economic burden worldwide. [1] In clinical settings, reports indicate that ADR-related emergency department visits increased from 5.6 to 11.6 per 100,000 people between 2005 and 2011, with an average of 25,303 cases annually. [2] 25.4% of ADR cases resulted in severe outcomes (hospitalization, transfer, or death), with CNS agents (59.1%) and opioids (17.4%) being the most frequently implicated drugs. [2] The financial impact of ADR-related hospitalizations and medical interventions is staggering, with an estimated economic burden of billions of dollars per year. In research published in the U S, the cost of an ADR varied from US$2000 to US$4000 per patient. [3] These statistics highlight the critical need for efficient, real-time ADR monitoring systems to augment existing drug safety surveillance frameworks. Traditional pharmacovigilance mechanisms, notably the FDA Adverse Event Reporting System (FAERS) and other spontaneous reporting systems (SRSs), are primarily dependent on voluntary submissions by patients and healthcare professionals. However, these systems exhibit significant underreporting. Several factors contribute to underreporting; however, the knowledge and attitude of healthcare professionals appear to be the most significant determinants. [4] Major factors include lack of awareness, time constraints, and reporting biases, leading to a delayed identification of potentially severe ADRs. The reliance on passive surveillance mechanisms inherently limits the timeliness and completeness of ADR signal detection, thereby restricting the responsiveness of regulatory agencies to emerging drug safety concerns.

With the proliferation of social media handles, particularly Twitter, an increasing volume of user-produced health-related content is becoming available. Social media

enables patients to share their real-time experiences with medications, often capturing ADR-related discussions that may never be reported through traditional pharmacovigilance channels. [5] The widespread usage of social media-based health discussions and its ability to provide instantaneous insights into drug-related adverse effects presents an unprecedented opportunity to enhance pharmacovigilance efforts. Unlike structured clinical trial data or electronic health records (EHRs), social media facilitates the detection of patient-centric ADR narratives, including those affecting underrepresented populations, off-label drug use, and interactions that may go unnoticed.

Most studies in social media-based ADR detection rely on conventional machine learning models and simple feature selection techniques (e.g., frequency-based or tree-based approaches) that inadequately handle the informal, noisy, and high-dimensional nature of social media text. Existing research on ADR detection from social media has largely focused on natural language processing (NLP) and deep learning methods but has often overlooked the importance of intelligent feature selection, which is key to reducing noise and improving model performance. Meanwhile, conventional pharmacovigilance systems continue to face issues like underreporting and delayed detection as they rely heavily on voluntary reporting. While social media, especially Twitter, offers an abundant and real-time source of patient-reported ADR data, leveraging this information effectively is difficult because of the unstructured and informal nature of posts, data noise, and ambiguity in symptom descriptions. These issues highlight the need for more advanced, context-aware approaches that integrate optimized feature selection and enhanced text interpretation to improve ADR detection from social media streams.

Traditional feature selection methods like frequency or tree-based rankings struggle with high-dimensional, sparse, and complex text data like Twitter. The application of metaheuristic algorithms for feature selection has been scarcely explored in this domain. Furthermore, most ADR detection models overlook the integration of semantic Similarity and contextual attention mechanisms, essential for distinguishing genuine ADRs from general drug discussions or unrelated content in social media streams. This highlights a crucial gap in developing robust, precise, and context-aware ADR detection frameworks that can effectively harness the possibility of employing social media data for pharmacovigilance in real-time.

Developing efficient, real-time ADR detection models from social media data significantly affects drug safety monitoring and regulatory decision-making. By integrating machine learning-driven pharmacovigilance systems with traditional ADR reporting frameworks, healthcare authorities can improve early signal detection for faster responses to emerging drug safety concerns, capture a broader range of ADR experiences from underrepresented patient groups, and enhance real-time surveillance to dynamically identify potential safety risks. Furthermore, leveraging social media data reduces reliance on passive reporting systems, addressing underreporting and delayed ADR recognition challenges, ultimately strengthening pharmacovigilance efforts.

Most studies in social media-based ADR detection rely on conventional machine learning models and simple feature selection techniques (e.g., frequency-based or tree-based approaches) that inadequately handle social media text's informal, noisy, and high-dimensional nature. Moreover, while deep learning models, namely CNNs and RNNs, have improved performance, they often neglect the role of sophisticated feature selection and semantic similarity mechanisms that can capture subtle ADR mentions in unstructured text.

The application of metaheuristic algorithms for feature selection has been scarcely explored in this domain. Furthermore, most ADR detection models overlook the integration of semantic Similarity and contextual attention mechanisms, essential for distinguishing genuine ADRs from general drug discussions or unrelated content in social media streams.

This highlights a crucial gap in developing robust, precise, and context-aware ADR detection frameworks that can effectively harness the possibilities of social media information for real-time pharmacovigilance.

The present study focuses on developing a novel deep learning framework for recognizing ADRs from social media, particularly Twitter, by integrating advanced NLP, machine learning, metaheuristics, and semantic similarity techniques. The key objectives of the study are:

- Develop a robust preprocessing pipeline to effectively handle noisy and unstructured Twitter data.
- Implement a feature extraction and selection strategy using a novel Decision Tree based Pelican Optimization Algorithm (DT-POA).
- Integrating Decision Tree-based feature ranking and Bag-of-Words (BoW) encoding to capture linguistic and semantic features relevant to ADR detection.
- Design and optimize a Deep Convolutional Recurrent Semantic Similarity Model (DCR-SSM), which combines CNNs for local feature extraction, bidirectional LSTMs (Bi-LSTMs) for contextual learning, and semantic similarity-based attention mechanisms to enhance ADR classification performance.
- The Proposed Model should be evaluated against state-of-the-art adverse drug reaction (ADR) detection methods by benchmarking its performance utilizing precision, accuracy, recall, and F1-score on the SMM4H dataset

- Assess the potential of social media-based pharmacovigilance as a complementary strategy to existing ADR monitoring systems, exploring its real-world applicability in drug safety surveillance.

2. Literature Review

The literature review portion of the paper explores the challenges and advancements in ADR detection, emphasizing the limitations of traditional pharmacovigilance systems and the emerging role of social media mining. It examines various ML and DL approaches, highlighting their effectiveness in enhancing real-time ADR monitoring and improving drug safety surveillance.

2.1. ADR Classification

South et al. (2014) explored the integration of automated tools for pre-annotation and interactive interfaces in the manual de-identification of clinical text, providing foundational insights regarding the role of machine learning in healthcare-related NLP tasks. Their findings highlighted how automating sensitive information tagging could streamline de-identification, enhance efficiency, and reduce annotator workload, though challenges like data privacy, false positives, and adaptability to varying formats persisted. [6] While primarily focused on de-identification, this study underscores the broader applicability of NLP techniques in healthcare, laying the groundwork for processing complex clinical and health-related texts, including those pertinent to pharmacovigilance.

Building on NLP for healthcare applications, Lin et al. (2015) focused on representing and encoding words from Twitter data to enhance ADR identification. Their study evaluated different word and phrase encoding methods, emphasizing the need for precise representations of informal, user-generated social media content to improve the accuracy of ADR extraction. [7] However, their work did not delve deeply into feature selection or contextual interpretation, which are critical for robust ADR detection.

Xu et al. (2015) examined the structure and dynamics of health-related discussions on Twitter hashtags, observing that these decentralized interactions often lacked continuity and reciprocity. [8] This finding is significant for ADR detection, as it highlights the fragmented and informal nature of social media data, which complicates the extraction of consistent and meaningful drug-safety signals.

Korkontzelos et al. (2016) demonstrated how sentiment analysis, combined with NLP approaches, can enhance the identification and understanding of ADR-related information from social media content. Their work revealed that users' emotional expressions often correlate with ADR mentions, offering a nuanced layer of context that purely syntactic or semantic analyses might overlook. [9] However, their approach was limited by the lack of advanced feature selection and semantic refinement to distinguish genuine ADRs from irrelevant sentiments.

Huynh et al. (2016) introduced deep neural networks for ADR classification, marking a significant shift toward leveraging deep learning to handle complex, unstructured text data. [10] Their Model improved the detection of potential drug side effects but lacked integration with more advanced NLP techniques like semantic Similarity and attention mechanisms, which are vital for nuanced interpretation of social media text.

Wei et al. (2016) contributed to the BioCreative V challenge by developing computational models for extracting chemical-disease relations (CDRs) from biomedical literature. Their large annotated corpus and high-performing machine-learning models advanced the domain of biomedical text mining.

However, their work focused more on formal publications than the informal, user-generated social media texts critical for real-time ADR detection. [11]

Lample et al. (2016) introduced Bi-LSTMs and character-based word embeddings for named entity recognition (NER), achieving state-of-the-art results in multiple languages. Their architecture effectively captured the text's syntactic and morphological nuances, crucial for detecting drug names and symptoms in social media posts. [12] However, the Model's direct applicability to Twitter's noisy and informal language was not fully explored.

Yang et al. (2016) established a hierarchical attention network (HAN) that leveraged dual-level attention mechanisms—word-level and sentence-level—to reflect document structure in classification tasks. Their Model improved performance by identifying and weighting informative text components, which aligns well with the challenges of capturing key ADR-related terms and contexts in social media narratives. [13] However, HAN's potential for integration with semantic Similarity and domain-specific feature selection remained underexplored.

Luo et al. (2017) emphasized the significance of biomedical relation extraction from clinical narratives, highlighting the utility of graph-based methods for interpreting complex relationships in health data. Their discussion of challenges such as Coreference resolution, named entity recognition, and feature sparsity is particularly relevant for ADR detection from social media, where ambiguity, informal references, and contextual dependencies are prevalent. [14] However, their work mainly focused on structured clinical data, leaving a gap in methodologies tailored to unstructured and user-generated content from platforms like Twitter.

Together, these studies underscore the evolving landscape of ADR detection, highlighting advances in NLP and deep learning while revealing persistent gaps in handling noisy, unstructured social media data. Despite progress in word representation, sentiment analysis, and NER, the literature lacks comprehensive models that combine advanced feature selection, semantic similarity refinement, and context-aware deep learning architectures—especially metaheuristic-optimized approaches—for ADR detection. This gap motivates the current study's development of a novel framework integrating a DT-POA for feature selection and a DCR-SSM for precise ADR classification from Twitter data.

2.2. Social Media Mining

With the growing popularity of social media, researchers have increasingly focused on leveraging user-generated content for ADR detection. Sarker et al. (2015) conducted a thorough analysis of text mining approaches for extracting ADRs from social media, which offered insightful information about the potential of these platforms as alternative data sources for pharmacovigilance.

The primary advantage of their work lies in highlighting the breadth of techniques available, encompassing supervised and unsupervised learning, lexicon-based approaches, and hybrid models, all of which can tap into real-time patient-reported data that traditional systems might miss. [15] However, the prominent disadvantage of their review is its limited exploration of the scalability and robustness of these methods in real-world, noisy environments. The study also did not address the need for advanced feature selection techniques or context-aware models, which are crucial for handling the informal nature of social media text.

Nikfarjam et al. (2015) contributed significantly by introducing a novel approach that combined conditional random fields (CRFs) with word embedding clusters for extracting ADRs from Twitter and online health forums. This method demonstrated superior performance compared to earlier lexicon-based methods, thanks to its ability to capture semantic relationships between words and accommodate variations in expression.

The strength of this approach lies in its fusion of probabilistic models and semantic-rich features, enabling it to handle informal and variable language in user-generated content. [16] However, its primary limitation is the reliance on predefined feature sets and manual feature engineering, which may not fully capture the complexity of ADR narratives or scale effectively to large, diverse datasets like Twitter streams.

Gonzalez-Hernandez et al. (2017) emphasized the challenges of ensuring data quality in social media pharmacovigilance. They highlighted the high levels of noise, informality, and potential misinformation present in social media data. Their call for robust preprocessing and validation techniques is a clear advantage, as it recognizes the need for data cleaning and context filtering to improve the reliability of extracted ADR signals. [17] However, the study's focus on data quality challenges without proposing concrete, scalable solutions leaves a gap in practical methodology. Furthermore, their work did not explore the integration of advanced NLP or deep learning methods to automatically distinguish credible ADR mentions from irrelevant or exaggerated posts.

Cocos et al. (2017) tackled the challenge of highly imbalanced datasets, a common issue in ADR detection, by employing advanced sampling techniques and ensemble methods to improve classification performance. The strength of their work lies in their focus on improving recall for rare ADR events, ensuring that subtler ADR mentions are not overlooked by classifiers. [18] However, while effectively balancing recall, ensemble methods can also introduce complexity, increase computational demands, and risk overfitting if not carefully optimized for diverse and noisy data sources like Twitter.

In summary, these foundational studies have contributed substantially to advancing social media-based ADR detection by introducing novel methodologies, emphasizing data quality, and exploring linguistic challenges. However, each has limitations, including scalability issues, lack of advanced feature selection and semantic refinement, insufficient handling of noise and ambiguity, and the absence of comprehensive, context-aware models. These gaps motivate the development of more robust frameworks, like the proposed DT-POA + DCR-SSM approach, which integrates metaheuristic-driven feature selection and deep learning with semantic similarity-based attention mechanisms to address the complexities of social media text and improve ADR detection accuracy.

2.3. NLP and Deep Learning in ADR Research

Recent advancements in deep learning and natural language processing have created new opportunities for the detection of ADRs in unstructured text data:

Word embedding techniques have shown promise in capturing semantic relationships between drugs and ADRs. Nikfarjam et al. (2015) utilized word embedding clusters to improve ADR extraction from social media posts. [16] The studies explored using NLP and machine learning algorithms, focusing on feature extraction from three datasets. Selecting features thoughtfully led to notable improvements in classification accuracy. Combining features from established text classification fields, like sentiment analysis, with topic modelling features can benefit text classification tasks. The recent research works aimed to apply NLP and feature extraction techniques to medical contexts, focusing on automatic text summarization for condensing and identifying key drug-related information in social media networks.

Deep learning models have shown strong performance across a range of NLP tasks, comprising the detection of ADRs. For example, Lee et al. (2017) introduced CNN, designed to classify ADRs from social media content, and their Model achieved better results than conventional machine learning methods. [19] Huynh et al. (2016) suggested recurrent neural network (RNN) architecture for ADR classification, demonstrating improved performance in identifying text data's sequential dependencies. [10]

Recent studies have investigated using transfer learning and pretrained language models for ADR detection. Giorgi and Bader (2019) utilized transfer learning with the BERT model to improve ADR classification performance on social media data. Multi-task learning approaches have shown promise in leveraging related tasks to improve ADR detection performance. [20]

In summary, the literature reveals a growing interest in leveraging social media data for ADR detection, with NLP and deep learning techniques emerging as powerful tools to address the associated challenges.

However, a need remains for more sophisticated approaches that can effectively handle the complexities of social media text while achieving high accuracy in ADR detection. Present research attempts to address this gap by proposing a novel framework integrating advanced NLP techniques with a deep learning architecture specifically designed for ADR classification from Twitter data.

3. Research Methodology

The proposed framework for detecting ADRs from Twitter data using NLP and deep learning models is explained in this section. Framework consists of five main stages: data collection, preprocessing, feature extraction, feature selection/optimization, and classification.

Fig. 1 offers an overview of the proposed framework.

3.1. Data Collection

The dataset utilized in this study is the SMM4H dataset, [21] a specialized collection of health-related social media data, primarily from Twitter, designed to support NLP tasks in public health and pharmacovigilance. It includes annotated data for extracting meaningful information such as drug mentions, adverse drug reactions, symptoms, and disease discussions.

Widely used in shared tasks and competitions, SMM4H helps researchers tackle challenges like noisy, unstructured text and enables applications in monitoring health trends and tracking drug safety. This dataset is instrumental in advancing health informatics by allowing NLP models to derive actionable insights from social media health discussions.

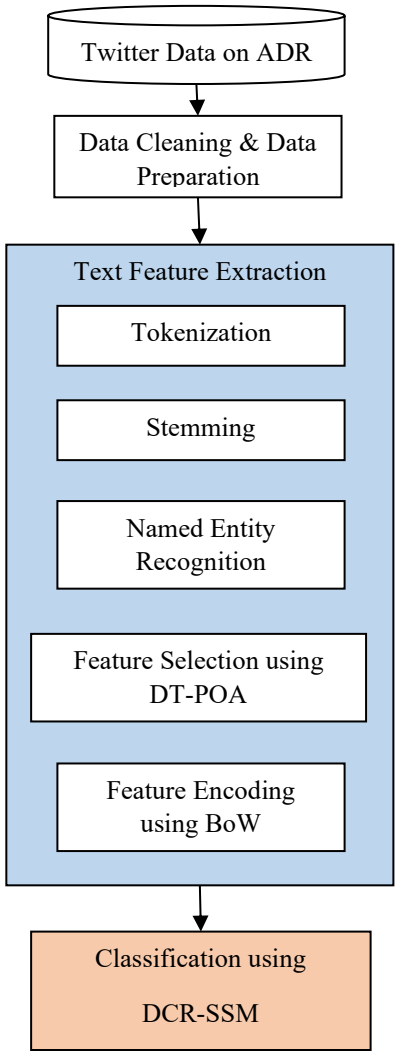

Fig. 1 Proposed ADR Detection Framework

3.2. Data Preprocessing

To ensure effective ADR detection from Twitter data, A thorough preparation pipeline has been created to address insignificant and noisy characteristics of social media text. The pipeline began with a text cleaning stage, removing URLs, user mentions, and special characters to eliminate non-informative elements.

All the texts were changed to lowercase to maintain uniformity, and contractions have been elaborated (like converting "do not" to "do not") to enhance the interpretability of the text. This cleaning phase aimed to prepare the data for more structured processing in subsequent stages.

Next, tokenization was performed using the NLTK Tweet Tokenizer, a tool designed to handle social media language with its characteristic slang, abbreviations, and emoji. This

tokenizer effectively splits the text into individual tokens, providing a foundation for word-level analysis. Spelling correction was applied to address common misspellings, particularly of drug names and medical terms, using a custom algorithm based on edit distance and drug name similarity. This correction step was critical in maintaining the accuracy of medical terminology, which is essential for ADR identification.

To focus on relevant information, stop word removal was conducted, removing common stop words while retaining negation words such as "not" and "no," which are essential for understanding the context of ADRs, as negations often influence the sentiment and meaning of ADR-related statements. Following this, NLTK's WordNetLemmatizer was implemented, breaking words down to their most basic forms to minimize lexical variability, thereby improving the consistency and comparability of terms used across tweets.

Finally, a custom Named Entity Recognition (NER) model trained on medical corpora was employed to identify and tag critical entities, including drug names, symptoms, and other relevant medical terms. This NER step enriched the data with medically relevant tags, facilitating precise ADR detection. In the initial data extraction phase, the unavailable or unfound tweets were removed, retaining only the accessible tweets for analysis. These preprocessing steps established a clean, standardized, and medically relevant dataset suitable for robust ADR detection and analysis.

3.3. Feature Extraction

A comprehensive text feature extraction process is designed to ascertain ADR's semantic and linguistic attributes. This process begins with N-gram features, extracting unigrams, bigrams, and trigrams to capture the local context around ADR mentions, allowing us to understand the immediate linguistic environment. Part-of-speech (POS) tags were generated via NLTK's POS tagger to provide syntactic information, which aids in identifying the grammatical roles of words related to ADRs.

Named entity tags identified during the NER step are included to highlight specific drug names, symptoms, and other medically relevant entities, enhancing the specificity of ADR-related data. To assess the sentiment surrounding ADR mentions, sentiment polarity is computed using VADER (Valence Aware Dictionary and sentiment Reasoner), which provides an overall sentiment score for every tweet's content, offering insight into the emotional context of ADR discussions.

Further analysis of semantic Similarity was conducted by calculating the Similarity between each tweet and a predefined list of ADR-related terms using word embeddings, providing an additional layer of relevance-based filtering. Drug-symptom co-occurrence frequencies are also measured, allowing us to detect patterns in the co-occurrence of drug names and potential symptom terms, which is essential for ADR identification.

In line with the workflow depicted in the diagram, Decision Tree-based feature selection prioritizes the most informative features, improving model efficiency. The BoW model is employed for feature encoding, ensuring a structured depiction of the extracted features, which is compatible with machine learning algorithms used for ADR detection. This systematic feature extraction, selection, and encoding approach provides a robust framework for analyzing and detecting ADR mentions in social media text.

3.4. Feature Selection

Feature selection is an important feature engineering process in machine learning pipelines, particularly for text classification tasks like ADR detection. The goal is to identify the most pertinent and significant characteristics while minimizing redundancy and improving model efficiency and interpretability. This study employs a novel DT-POA as the feature selection method, leveraging the metaheuristic ability to capture feature interactions and hierarchical dependencies within the dataset.

The POA introduced by Trojovský, P., & Dehghani, M. (2022), [22] is a recent nature-inspired optimization technique that mimics the cooperative hunting behaviour of pelicans. In the proposed framework, each pelican represents a candidate feature subset, and the quality of each subset is evaluated using the classification error rate of a DT classifier—this acts as the fitness function guiding the optimization process.

Let the original feature space be represented as $F = \{f_1, f_2, \dots, f_d\}$, where d represents the total number of features extracted from the dataset after preprocessing.

Each solution (or pelican) in the search space can be represented by a binary vector as:

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{id}\}, \qquad x_{ij} \epsilon \{0,1\} \tag{1}$$

Here,

X_i Is the ith pelican,

$x_{ij} = 1$ indicates feature f_j is selected in the subset,

$x_{ij} = 0$ indicates it is not selected.

For each pelican X_i, the selected features $S_i \subset F$ are passed to a DT classifier, which is trained and evaluated. The classification error rate from the decision tree is used as the fitness score:

$$f(X_i) = 1 - Accuracy_{DT}(S_i) \tag{2}$$

This fitness function must be minimized, i.e., better feature subsets produce lower classification errors.

Pelicans are initialized using:

$$x_{ij} = \begin{cases} 1, & if\ rand() < p \\ 0, & otherwise \end{cases} \quad (3)$$

Where rand () generates a random number in [0,1], and p is a predefined selection probability. This allows each pelican to begin with a unique, randomized subset of features.

3.4.1. Phase 1: Moving Towards Prey

This phase simulates pelicans identifying a promising area (i.e., a potentially better feature subset configuration):

$$x_{ij}^{P1} = \begin{cases} x_{ij} + rand() \cdot (p_j - I \cdot x_{ij}), & if\ f(p) < f(X_i) \\ x_{ij} + rand() \cdot (x_{ij} - p_j), & otherwise \end{cases} \quad (4)$$

Where,
x_{ij}^{P1} is the updated value of feature j in pelican i,
p is the position of the prey (a randomly selected elite solution) $I \epsilon \{1,2\}$ is a random integer,$rand() \epsilon [0,1]$

After the update, binary rounding is applied:

$$x_{ij}^{P1} = \begin{cases} 1, & if\ x_{ij}^{P1} < 0.5 \\ 0, & otherwise \end{cases} \quad (5)$$

The new position is accepted if it results in improved fitness:

$$X_i = \begin{cases} X_i^{P1}, & if\ f(X_i^{P1}) < f(X_i) \\ X_i, & otherwise \end{cases} \quad (6)$$

3.4.2. Phase 2: Winging on the Water Surface

In this stage, the pelican fine-tunes its position near the current best-known solution:

$$x_{ij}^{P2} = x_{ij} + R \cdot \left(1 - \frac{t}{T}\right) \cdot (2 \cdot rand() - 1) \cdot x_{ij} \quad (7)$$

Where:
R =control parameter (typically 0.2)
t =current iteration number,
T = the maximum number of iterations.

Again, the binary form of the solution is recovered using Thresholding, and positions are updated conditionally based on fitness:

$$X_i = \begin{cases} X_i^{P2}, & if\ f(X_i^{P2}) < f(X_i) \\ X_i, & otherwise \end{cases} \quad (8)$$

The selected features are subsequently encoded using a BoW model to transform textual data into a structured numerical representation.

3.5. Feature Encoding

The BoW technique is a way of transforming text into numerical features that machine learning approaches might utilize. The technique generates a vocabulary of unique words from the dataset and depicts every text based on the occurrence or frequency of these words.

3.5.1. Create a Vocabulary of Unique Words

First, BoW scans through all the text data (e.g., tweets) to generate a list of unique words or "vocabulary." This vocabulary becomes the foundation for encoding the text data. The dataset is scanned to create a vocabulary set V containing unique words across all tweets. Given a dataset of tweets $D = \{d_1, d_2, d_3 \dots d_N\}$, the vocabulary is:

$$V = \bigcup_{i=1}^{N} W(d_i) \quad (9)$$

Where $W(d_i)$ stands for a set of unique words in a tweet d_i

3.5.2. Build a Matrix of Word Occurrences

Each tweet is then represented as a vector in a matrix $x_i \in \mathbb{R}^{|V|}$ Where every column represents a word in the vocabulary, for every tweet, the matrix will record either the count of each word (frequency-based) or just a binary indicator (0 or 1) to signify whether the word appears or not. The feature values are encoded using:

Binary Encoding: $x_{ij} = 1$ if word w_j appears in d_i, else 0.
Term Frequency (TF): $x_{ij} = count(w_j, d_i)$
TF-IDF (TF-Inverse Document Frequency):

$$x_{ij} = TF(w_i, d_i) \cdot \log \frac{N}{DF(w_j)} \quad (10)$$

Where $DF(w_j)$ = number of documents containing w_j

3.5.3. Transform Text into Word Vectors

In the BoW model, each tweet becomes a numeric vector whose length matches the vocabulary size. For every term in that vocabulary, the corresponding vector slot is activated if the term appears in the tweet—either by adding one to its count (frequency BoW) or setting the slot to one to signal mere presence (binary BoW).

3.5.4. Feed Encoded Data into the Model

The resulting matrix, where every row corresponds to a tweet regarding word occurrences, is then employed as input for machine learning approaches. Models can leverage this structured data to learn patterns associated with ADR mentions.

BoW is straightforward and highly effective for capturing word presence and frequency in text data, making it among the most prominent encoding techniques for NLP tasks. However,

it does not account for the order of words, so context and meaning based on sequence are not preserved.

By integrating DT-based feature selection with BoW encoding, the proposed approach ensures that the Model retains only the most discriminative linguistic features while reducing noise from redundant or irrelevant words. This structured feature engineering framework improves classification accuracy while reducing computational complexity. Decision Tree-based feature selection and BoW encoding establish a robust preprocessing pipeline, optimizing text representation for deep learning-based ADR detection from social media data.

3.6. Classification using Deep Convolutional Recurrent Semantic Similarity Model (DCR-SSM)

The final classification stage leverages a novel DCR-SSM, integrating CNNs for local feature extraction, Bi-LSTM for contextual representation, and a semantic similarity-based attention mechanism for refining ADR recognition. The proposed Model follows a hierarchical architecture that captures social media text's linguistic and semantic properties while improving classification performance.

The architecture of the DCR-SSM comprises the following layers:

3.6.1. Embedding Layer

This layer transforms input tokens to dense vector illustration using Pretrained word embeddings (e.g., Word2Vec or GloVe)

$$X = [x_1, x_2, \dots x_T] \in \mathbb{R}^{T \times d} \tag{11}$$

T is the sequence length (maximum number of words in a tweet),d represents the dimensionality of word embeddings (e.g., 50 for GloVe embeddings),x_t Represents the embedding vector for the word at position t.

3.6.2. Convolutional Layer

This layer applies convolutional filters of varying sizes to capture local n-gram patterns.

$$c_t = f(W_c \cdot X_{t:t+k-1} + b_c) \tag{12}$$

$X_{t:t+k-1}$ represents the window of k words centered at t b_c stands for bias term,$f(\cdot)$ Represents non-linear activation function (ReLU).

3.6.3. Max Pooling Layer

This layer executes max pooling on the convolutional outputs to extract only the significant features.

Multiple filters generate a feature map: $C = [c_1, c_2, \dots c_{T-k+1}]$Which is passed through a max-pooling layer to retain the most significant local features:

$$\hat{c} = \max C \tag{13}$$

3.6.4. Bi-LSTM Layer

This layer handles the pooled features using a Bi-LSTM to extract contextual details and long-range dependencies.

The layer processes the input sequence twice—once from start to finish and again in reverse:

$$\overrightarrow{h_t} = LSTM_{fwd}(\hat{c}, h_{t-1}) \tag{14}$$

$$\overleftarrow{h_t} = LSTM_{bwd}(\hat{c}, h_{t+1}) \tag{15}$$

Final hidden state representation is obtained by concatenating both directions:

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{16}$$

3.6.5. Semantic Similarity Layer

This layer computes semantic Similarity between LSTM outputs and predefined ADR-related embeddings.

$$S(h_t, E) = cos(h_t, E) = \frac{h_t \cdot E}{||h_t|| ||E||} \tag{17}$$

Where $cos(h_t, E)$ Represents the cosine similarity between the hidden state and ADR embeddings.

3.6.6. Attention Layer

An attention module within this layer highlights the input segments that are most informative for identifying adverse drug reactions.

$$\alpha_t = \frac{\exp(W_a h_t)}{\sum_{t'} \exp(W_a h_{t'})} \tag{18}$$

$$h_{att} = \sum_t \alpha_t h_t \tag{19}$$

Where,
W_a Is the attention weight matrix,
α_t is the attention score assigned to h_t,
h_{att} It is the weighted sum of hidden states.

3.6.6. Fully Connected Layer:

It aggregates the outputs produced by earlier layers and then feeds the consolidated vector into a dense layer, where dropout is applied to curb over-fitting. A final function of a sigmoid activation is applied to produce the final prediction (ADR or non-ADR):

$$y = \sigma(W_o h_{att} + b_o) \tag{20}$$

W_o and b_o Are output parameters, σ represents the sigmoid activation function,y represents the probability of a tweet comprising an ADR mention.

The Model has been trained to employ Binary Cross-Entropy (BCE) loss function with semantic similarity regularization:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)) + \lambda\|S(h_t, E)\| \quad (21)$$

Where,
y_i = ground-truth label,
$\hat{y}_i$ = predicted probability label,
λ = regularization weight.

Weighted loss and data augmentation techniques were employed during training to address class imbalance. The hyperparameter settings of the proposed Model are presented in Table.

Table 1. Hyperparameters Setting of the Proposed Model

Hyperparameters	Value
Pooling Size	2
LSTM Units	128
Dropout Rate	0.5
Batch Size	64
Learning Rate	0.001
Epochs	70
Loss Function	Semantic Similarity-based Binary Cross Entropy
Optimizer	Adam
Max Words (Tokenizer)	500
Max Sequence Length	50
Embedding Dimension	50
CNN Filters (First Layer)	256
CNN Kernel Size (First Layer)	5
CNN Filters (Second Layer)	128
CNN Kernel Size (Second Layer)	3
LSTM Units (First)	100
LSTM Units (Second)	50
Dropout (Fully Connected Layer)	0.5 & 0.3
Final Activation	Sigmoid

4. Results and Discussion

4.1. Evaluation Metrics

The following evaluation measures were taken into consideration to assess the proposed ADR detection system's performance thoroughly:

Precision: it measures a fraction of ADR predictions that are genuinely correct—that is, the share of entities the model tags as ADRs that truly are ADRs.

$$Precision = \frac{TP}{(TP + FP)} \quad (22)$$

Where TP=True Positives, FP=False Positives

Recall: Recall represents the fraction of correctly recognized ADR mentions among all actual ADR instances.

$$Recall = \frac{TP}{(TP+FN)} \quad (23)$$

Where FN = False Negatives

F1-score: F1-metric calculates the harmonic average of recall and precision, giving a single, balanced gauge of the Model's overall effectiveness.

$$F1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (24)$$

Area Under the Receiver Operating Characteristic Curve (AUC-ROC): It quantifies how effectively the Model separates ADR mentions from non-ADR ones over all possible classification thresholds.

Matthews Correlation Coefficient (MCC): MCC delivers an unbiased evaluation of binary classification performance, proving especially advantageous when the classes are highly imbalanced.

$$MCC = \frac{((TP \times TN) - (FP \times FN))}{\sqrt{((TP + FP)(TP + FN)(TN + FP)(TN + FN))}} \quad (25)$$

Where TN = True Negatives

4.2. Evaluation Metrics

Experiments have been conducted using the SMM4H dataset, containing tweets related to drug experiences. [21] The dataset was preprocessed and annotated as described in Section 3. We adopted a five-fold cross-validation scheme to obtain a dependable assessment of model performance.

The DCR-SSM network was built in PyTorch and trained on an NVIDIA Tesla V100 accelerator. Optimization used the Adam algorithm with an initial step size of 1×10^{-3} and mini-batches of 64 samples. The training ran for up to 70 epochs, with early stopping triggered when validation metrics ceased improving.

4.3. Performance Evaluation

The performance of the proposed framework is depicted through an accuracy plot against a number of epochs for training and validation dataset splits in Fig. 1. The proposed ADR detection framework's training and validation accuracy curves show a steady improvement over 70 epochs, with rapid learning in the initial phase and gradual convergence beyond 20 epochs. Training accuracy stabilizes above 80%, while

validation accuracy reaches approximately 72%, indicating effective generalization with minimal overfitting. The close alignment of the two curves suggests that the model balances bias and variance well, leveraging convolutional layers for local feature extraction, recurrent layers for contextual learning, and semantic similarity-based attention for enhanced ADR classification. The absence of abrupt validation drops highlights the Model's robustness, making it highly suitable for deployment in practical pharmacovigilance settings.

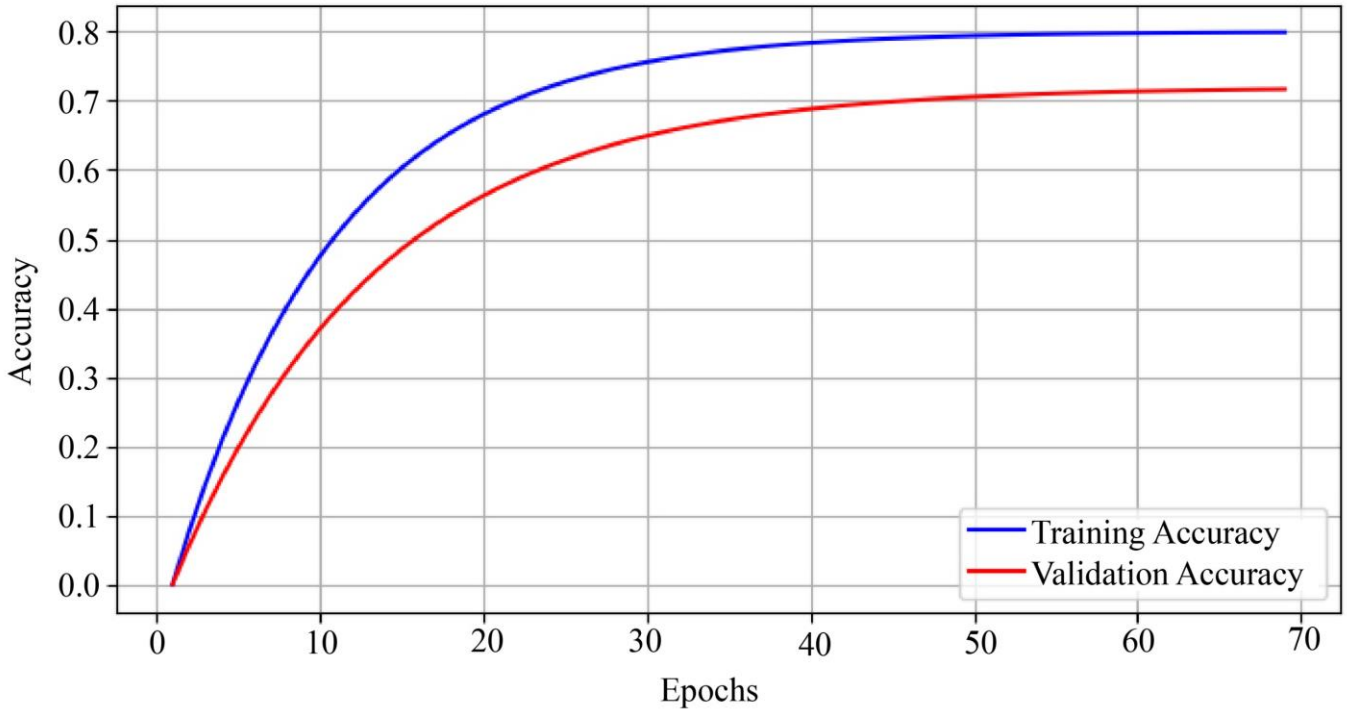

Fig. 1 Accuracy Vs Epochs

Table 1 offers a thorough comparison of proposed DT-POA + DCR-SSM with various state-of-the-art models for social media data-based ADR recognition. Baseline models included in the comparison represent a vast variety of deep learning architectures and ensemble methodologies, like transformer-based models (RoBERTa, BERT, DeBERTa, BioBERT, and BERTweet), hybrid approaches integrating domain-specific embeddings (ChemBERTa, Byte-Pair Embeddings, DeepADEMiner), and multi-task learning strategies. Studies by Magge et al. (2021), [23] Weissenbacher et al. (2022), [24] and Sakhovskiy et al. (2021) [25] showcase different advancements in NLP-based ADR detection, leveraging variations in data augmentation (over/under-sampling, SMOTE, domain adaptation) and Model fine-tuning strategies.

Table 1. Performance comparison of ADR detection methods

Method	Precision (%)	Recall (%)	F1-score (%)
Magge, A., et al. (2021) [23]			
RoBERTa + Under/Over-sampling	61.0	51.5	75.2
RoBERTa + ChemBERTa	61.0	55.2	68.1
BERT + Over-sampling + Ensemble	54.0	60.3	48.9
BERTweet + Pseudo Data	49.0	59.2	41.7
BERT + Class Weights	46.0	47.2	45.6
BERTweet + Class Weights	46.0	52.3	40.9

Multi-task Learning + BioBERT + Class Weights	44.0	49.1	39.3
RoBERTa + SMOTE + DA	40.0	40.5	40.1
BERT Ensemble + Over-sampling	40.0	52.1	32.7
BERT	23.0	13.5	72.6
Multi-task Learning + Selective Over-sampling	51.0	51.4	51.4
RoBERTa + FastText + Byte-Pair Embeddings	50.0	55.5	45.9
RoBERTa	50.0	49.3	50.5
BERT + BiLSTM + CRF	42.0	38.1	47.5
EnDR-BERT + CADEC & COMETA Data	40.0	42.0	38.2
BERT + Joint NER & Normalization	37.0	58.0	27.5
Weissenbacher, D., et al. (2022) [24]			
BERTweet-large + DA	69.8	83.9	59.8
10x RoBERTa-large	69.3	77.2	62.9
DeBERTa-v3 + AdvT	68.9	79.0	61.1
RoBERTa + BERTweet + EMA	66.2	78.5	57.3
BERTweet + DeBERTa + BioBERT + DA	66.2	76.5	58.4
T5 + GPT-2 + Over/Under Sampling	65.5	68.8	62.5
RoBERTa + In-domain Tweets	65.2	73.7	58.5
Glove + DeepADEMiner	64.2	55.4	76.5
RoBERTa-base + AdvT	63.7	78.7	53.6
BERTweet-large + RoBERTa-large + CT-BERT	61.0	60.6	61.4
RoBERTa + FGM + PGD	60.1	70.5	52.4
RoBERTa + Adaptive Learning	56.7	67.4	48.9
RoBERTa + DA + Downsampling	49.1	38.4	68.1
BERT + Med Data	47.2	60.7	38.6
BERTweet + Template Aug.	43.3	61.4	33.4
BERT + RoBERTa + ERNIE 2.0	41.3	67.7	29.7
BERT + BioBERT + XLNet + RoBERTa	29.9	23.5	40.9
RoBERTa + BERTweet + LDA Loss	7.7	4.1	54.7
Sakhovskiy, A., et al. (2021) [25]			
RoBERTa + ChemBERTa + Over-sampling + Sigmoid	55.0	68.0	61.0
RoBERTa + ChemBERTa + Over-sampling + Sigmoid	59.0	56.0	58.0
Proposed Model DT-POA + DCR-SSM	**75.0**	**72.0**	**73.0**

The models' performances vary significantly across recall, precision, and F1-score metrics, with the proposed DT-POA + DCR-SSM model demonstrating superior results in all aspects. Regarding precision, existing models show a wide range, with the lowest precision recorded at 7.7% (RoBERTa + BERTweet + LDA Loss) and the highest among previous models at 69.8% (BERTweet-large + DA). In contrast, the proposed DT-POA + DCR-SSM model achieves a precision of 75%, marking an improvement of 5.2% over the best-performing baseline. Similarly, for recall, the prior models range from as low as 4.1% (RoBERTa + BERTweet + LDA Loss) to a peak value of 83.9% (BERTweet-large + DA). The proposed Model achieves a recall of 72%, which, while slightly lower than the top recall score, offers a more balanced performance by maintaining high precision alongside recall, avoiding the trade-off observed in several previous methods.

The F1 metric—determined by employing a harmonic average of precision and recall—further proves the proposed Model's superior performance. Previous simulations demonstrate highly variable F1-scores, with the lowest recorded at 27.5% (BERT + Joint NER & Normalization) and the highest among existing methods at 76.5% (Glove + DeepADEMiner). The DT-POA + DCR-SSM model attains

an F1-score of 73%, placing it among the highest-performing models while ensuring a stable balance between precision and recall. The improvements made by DT-POA + DCR-SSM can be attributed to its integration of convolutional and recurrent layers, which enable effective feature extraction, along with the inclusion of a semantic similarity mechanism and attention-based refinement that improves the Model's capability to extract contextually rich ADR mentions in unstructured social media text.

Overall, the findings demonstrate that the proposed framework effectively remedies the shortcomings inherent in earlier ADR detection approaches. While many prior approaches rely heavily on transformer-based architectures without domain-specific enhancements, the DT-POA + DCR-SSM model leverages a more structurally comprehensive approach, integrating CNNs for local feature extraction, Bi-LSTM for contextual learning and semantic similarity-based attention to refine classification decisions. The significant gains in precision and F1-score, coupled with the Model's ability to maintain competitive recall, position it as a highly effective solution for ADR detection in social media-based pharmacovigilance. Future enhancements could explore improvements in domain adaptation techniques and ensemble learning to optimize recall while maintaining high precision.

Table 2 compares the proposed DT-POA + DCR-SSM against traditional ML and DL techniques, including Support Vector Machine (SVM), LSTM, Bi-LSTM, and CNN. These models represent a spectrum of approaches used for text classification in ADR detection, ranging from classical machine learning (SVM) to more advanced DL architectures (LSTM, Bi-LSTM, and CNN), each with varying capabilities in extracting contextual associations and sequential dependencies in textual data.

Table 2. Comparison of the proposed framework with existing models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	65.4	68.4	66	67
LSTM	67.5	71	67.3	68.9
Bi-LSTM	68.1	71.5	68.2	69.7
CNN	69.8	72	69	70.6
Proposed DT-POA + DCR-SSM	**72**	**75**	**72**	**73**

Among the baseline models, SVM achieves the lowest performance, exhibiting a precision of 68.4%, recall of 66%, accuracy of 65.4%, and F1-score of 67%. This relatively lower performance reflects the limitations of conventional machine learning models in handling the complex and context-dependent nature of ADR mentions in unstructured social media text. The DL models—LSTM, Bi-LSTM, and CNN—show incremental improvements, with CNN achieving the highest accuracy at 69.8% and an F1-score of 70.6%. This suggests that CNN's ability to capture local patterns through convolutional filters enhances ADR detection compared to purely sequential models like LSTM.

The proposed framework—DT-POA + DCR-SSM—introduces a novel approach for ADR detection from social media by integrating advanced NLP, deep learning, and metaheuristic-driven feature selection. Unlike prior studies that predominantly relied on transformer-based models like BERT, RoBERTa, and BERTweet with basic oversampling or weighting strategies, the Model incorporates a Decision Tree-based Pelican Optimization Algorithm (DT-POA) for intelligent feature selection and a Deep Convolutional Recurrent Semantic Similarity Model (DCR-SSM) for robust classification. This unique combination enhances precision (75%), recall (72%), and F1-score (73%)—outperforming the best-reported precision (69.8%) and F1-score (76.5%) from previous models like BERTweet-large + DA and Glove + DeepADEMiner, while maintaining a balanced performance across metrics. The integration of CNNs for local pattern recognition, Bi-LSTM for capturing long-term dependencies, and semantic similarity-based attention mechanisms further distinguishes the proposed Model from prior work, which often lacks contextual refinement and optimized feature selection. This comprehensive and structurally robust approach addresses the challenges of noisy, high-dimensional social media text and sets a new benchmark for ADR detection.

In comparison to classical models (SVM, LSTM, Bi-LSTM, and CNN), the proposed framework demonstrates a significant performance improvement, achieving the highest accuracy (72%) and maintaining a superior balance between precision and recall, highlighting its potential for real-world pharmacovigilance applications. The improvement over CNN (highest-performing baseline) is notable, with an increase of 2.2% in accuracy and 2.4% in F1-score. This enhancement is attributable to the integration of convolutional layers for local feature extraction, Bi-LSTM layers for capturing long-range dependencies, and a semantic similarity-based attention mechanism, which refines the Model's capability to distinguish ADR mentions from non-ADR text. The 3% increase in precision compared to CNN shows the Model's capability to reduce false positives, ensuring higher confidence in ADR classification.

The results highlight that while traditional DL models like LSTM and Bi-LSTM improve upon classical approaches like SVM, the DT-POA + DCR-SSM model provides the most balanced and robust performance. Combining convolutional and recurrent architectures and semantic refinement enables the proposed Model to effectively capture local and global text patterns. These improvements position the DT-POA + DCR-SSM model as a highly effective tool for ADR detection in

social media-based pharmacovigilance, offering accuracy and reliability in identifying adverse drug reactions from unstructured user-generated text.

The DT-POA combined with the DCR-SSM architecture outshines other methods thanks to a set of distinct advantages:

- The DT-POA-based feature selection supports identifying the most relevant features, eliminating noise and improving the Model's attention to relevant information.
- <UNK> Combining BoW encoding and deep learning enables the Model to capture local and global patterns in the text data.
- The recurrent and convolutional components of the DCR-SSM model allow for better capturing sequential dependencies and local patterns in the text.
- By incorporating semantic similarity measures, the Model can better handle variations in how ADRs are expressed.
- The attention layer facilitates the Model to concentrate on the most significant portions of the input, improving its ability to identify ADR mentions accurately.

Findings illustrate the capacity of leveraging advanced NLP and DL techniques for ADR recognition from social media data. The proposed framework addresses several challenges identified in the literature, such as handling informal language, capturing context, and dealing with imbalanced data.

5. Conclusion

The present research provides a comprehensive framework for detecting ADRs from social media data, particularly Twitter, by leveraging advanced NLP, metaheuristics, and DL techniques. POA feature selection method adopts DT as the objective function to shortlist significant feature values, which are then encoded using BoW. The proposed DCR-SSM integrates convolutional and recurrent layers with a semantic similarity measure and attention mechanism, enabling the extraction of local and global contextual features from unstructured text. Through extensive evaluation of the SMM4H dataset, the Model demonstrated superior performance, achieving a precision of 75%, accuracy of 72%, recall of 72%, and an F1-score of 73%, outperforming existing state-of-the-art methods. Compared to traditional ML (SVM) and DL models (LSTM, Bi-LSTM, and CNN), the DT-POA + DCR-SSM showed significant improvements in all performance metrics, demonstrating its robustness and generalization capability in ADR detection. The findings underscore the framework's potential in real-time pharmacovigilance, offering a scalable and cost-effective alternative to traditional ADR monitoring systems. This approach can enhance early drug safety surveillance and support decision-making for healthcare professionals and regulatory agencies by effectively capturing patient-reported ADRs from social media. Future research can extend this framework to multi-lingual ADR detection and explore further optimizations, including domain-specific embeddings and adaptive learning techniques, to improve recall and overall classification performance.

Acknowledgements

The authors gratefully acknowledge everyone who supported this research and the preparation of the manuscript; Author 1 and Author 2 contributed equally to this work.

References

[1] Sarvnaz Karimi et al., "Text and Data Mining Techniques in Adverse Drug Reaction Detection," *ACM Computing Surveys*, vol. 47, no. 4, pp. 1-39, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[2] I-Jen P. Castle et al., "Emergency Department Visits for Adverse Drug Reactions Involving Alcohol: United States, 2005 to 2011," *Alcoholism: Clinical and Experimental Research*, vol. 40, no. 9, pp. 1913-1925, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[3] R. Bordet et al., "Analysis of the Direct Cost of Adverse Drug Reactions in Hospitalised Patients," *European Journal of Clinical Pharmacology*, vol. 56, pp. 935-941, 2001. [CrossRef] [Google Scholar] [Publisher Link]

[4] Toshiyuki Sakaeda et al., "Data Mining of the Public Version of the FDA Adverse event Reporting System," *International Journal of Medical Sciences*, vol. 10, no. 7, pp. 796-803, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[5] Muaed Alomar et al., "Post Marketing Surveillance of Suspected Adverse Drug Reactions through Spontaneous Reporting: Current Status, Challenges and The Future," *Therapeutic Advances in Drug Safety*, vol. 11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Brett R. South et al., "Evaluating the Effects of Machine Pre-annotation and An Interactive Annotation Interface on Manual De-identification of Clinical Text," *Journal of Biomedical Informatics*, vol. 50, pp. 162-172, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[7] Wei-San Lin et al., "Utilizing Different Word Representation Methods for Twitter Data in Adverse Drug Reactions Extraction," *2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[8] Weiai Xu et al., "Twitter Hashtags for Health: Applying Network and Content Analyses to Understand the Health Knowledge Sharing in a Twitter-based Community of Practice," *Quality & Quantity*, vol. 49, pp. 1361-1380, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[9] Ioannis Korkontzelos et al., "Analysis of the Effect of Sentiment Analysis on Extracting Adverse Drug Reactions from Tweets and Forum Posts," *Journal of Biomedical Informatics*, vol. 62, pp. 148-158, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[10] Trung Huynh et al., "Adverse Drug Reaction Classification with Deep Neural Networks," *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 877-887, 2016. [Google Scholar] [Publisher Link]

[11] Chih-Hsuan Wei et al., "Assessing the State of the Art in Biomedical Relation Extraction: Overview of the BioCreative V Chemical-Disease Relation (CDR) Task," *Database*, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[12] Guillaume Lample et al., "Neural Architectures for Named Entity Recognition," *arXiv preprint arXiv:1603.01360*, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[13] Zichao Yang et al., "Hierarchical Attention Networks for Document Classification," *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1480-1489, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[14] Yuan Luo, Ozlem Uzuner, and Peter Szolovits, "Bridging Semantics and Syntax with Graph Algorithms—State-of-the-art of Extracting Biomedical Relations," *Briefings in Bioinformatics*, vol. 18, no. 1, pp. 160-178, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[15] Abeed Sarker, and Graciela Gonzalez, "Portable Automatic Text Classification for Adverse Drug Reaction Detection Via Multi-corpus Training," *Journal of Biomedical Informatics*, vol. 53, pp. 196-207, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[16] Azadeh Nikfarjam et al., "Pharmacovigilance from Social Media: Mining Adverse Drug Reaction Mentions using Sequence Labeling with Word Embedding Cluster Features," *Journal of the American Medical Informatics Association*, vol. 22, no. 3, pp. 671-681, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[17] G. Gonzalez-Hernandez et al., "Capturing the Patient's Perspective: A Review of Advances in Natural Language Processing of Health-related Text," *Yearbook of Medical Informatics*, vol. 26, no. 1, pp. 214-227, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[18] Anne Cocos, Alexander G. Fiks, and Aaron J. Masino, "Deep Learning for Pharmacovigilance: Recurrent Neural Network Architectures for Labeling Adverse Drug Reactions in Twitter Posts," *Journal of the American Medical Informatics Association*, vol. 24, no. 4, pp. 813-821, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[19] Jae-Young Lee et al., "The Use of Social Media in Detecting Drug Safety-related New Black Box Warnings, Labelling Changes, or Withdrawals: Scoping Review," *JMIR Public Health and Surveillance*, vol. 7, no. 6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] John M. Giorgi, and Gary D. Bader, "Transfer Learning for Biomedical Named Entity Recognition with Neural Networks," *Bioinformatics*, vol. 34, no. 23, pp. 4087-4094, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[21] Davy Weissenbacher et al., "Overview of the Fourth Social Media Mining for Health (SMM4H) Shared Tasks at ACL 2019," *Proceedings of the Fourth Social Media Mining for Health Applications (# SMM4H) Workshop & Shared Task*, pp. 21-30, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] Pavel Trojovský, and Mohammad Dehghani, "Pelican Optimization Algorithm: A Novel Nature-inspired Algorithm for Engineering Applications," *Sensors*, vol. 22, no. 3, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Arjun Magge et al., "Overview of the Sixth Social Media Mining for Health Applications (# SMM4H) Shared Tasks at NAACL 2021," *Proceedings of the Sixth Social Media Mining for Health (# SMM4H) Workshop and Shared Task*, pp. 21-32, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Davy Weissenbacher et al., "Overview of the Seventh Social Media Mining for Health Applications (# SMM4H) Shared Tasks at COLING 2022," *Proceedings of the Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task*, pp. 221-241, 2022. [Google Scholar] [Publisher Link]

[25] Andrey Sakhovskiy, Zulfat Miftahutdinov, and Elena Tutubalina, "KFU NLP Team at SMM4H 2021 Tasks: Cross-lingual and Cross-modal BERT-based Models for Adverse Drug Effects," *Proceedings of the Sixth Social Media Mining for Health (# SMM4H) Workshop and Shared Task*, pp. 39-43, 2021. [CrossRef] [Google Scholar] [Publisher Link]