

An Attention-Enhanced Explainable Hybrid Deep Learning Framework for Multilingual Opioid Overdose Detection on Reddit

Muhammad Ahmad¹, Muhammad Usman¹, Umyh Habiba², Grigori Sidorov¹,
Venera Bayrasheva³, Nailya Kubysheva³, Ildar Batyrshin^{1,*}

¹ Instituto Politécnico Nacional, Centro de Investigación en Computación,
Mexico

² The Islamia University of Bahawalpur, Department of Computer Science,
Pakistan

³ Kazan Federal University,
Russian Federation

batyr1@cic.ipn.mx

Abstract. Opioid overdose remains a critical global public health challenge, with increasing morbidity and mortality driven by the misuse of prescription and synthetic opioids. Social media platforms provide real-time signals of opioid-related behaviors, yet automated detection of overdose-related content is hindered by informal language, multilingual discourse, and limited availability of fine-grained annotated datasets. Existing approaches are predominantly monolingual and offer limited interpretability, restricting their applicability in high-stakes public health settings. This study proposes a multilingual, explainable deep learning framework for opioid overdose detection using English and Spanish social media posts. We construct a novel bilingual dataset annotated with four clinically relevant route-of-administration categories through a hybrid LLM-assisted and expert-validated process. Two complementary modeling strategies are explored: a translation-based approach for cross-lingual standardization and a joint multilingual approach for shared representation learning. A hybrid CNN-BiLSTM architecture with a custom attention mechanism is introduced to capture both local lexical cues and long-range contextual dependencies, while explainable AI techniques provide interpretable model outputs. Experimental results demonstrate strong performance across all settings, achieving accuracies of 84% on English data, 79% on Spanish data, and 82% on the joint multilingual dataset. These findings highlight the potential of multilingual and interpretable NLP models for real-time opioid surveillance.

Keywords. Opioid overdose detection, public health surveillance, chatgpt, large language model, high dose,

machine learning, reddit, social media analytics, chronic pain.

1 Introduction

The misuse of prescription and illicit opioids has intensified into a persistent global public health emergency, driven by widespread over-prescription, increasing availability of potent synthetic opioids, and structural gaps in pain-management practices (Meier, 2003; Kulkarni et al., 2020; Schatman & Ziegler, 2017; Olsen, 2016).

Opioid-related morbidity and mortality continue to rise, placing substantial strain on healthcare systems, law enforcement, and community resources (Robinson et al., 2019; Peres et al., 2019; World Health Organization, 2009). As with other rapidly evolving public health threats (Gubler, 2002), timely detection of emerging opioid-related risks is essential for effective intervention and policy response.

Social media platforms have emerged as critical sources of real-time behavioral and epidemiological signals, offering insights into substance-use patterns, risk behaviors, and community-level concerns (Young et al., 2020; Kim et al., 2017; Weeks et al., 2002).

Individuals frequently disclose personal experiences, seek support, and discuss drug

availability and overdose incidents across platforms such as Reddit, Twitter, and YouTube (Balsamo, 2022; Carpenter et al., 2024). These digital traces provide large-scale, time-sensitive data that complement traditional surveillance systems.

However, extracting actionable information from social media text remains challenging due to substantial linguistic variability, including slang, abbreviations, coded expressions, and informal phrasing (Holbrook et al., 2024; Silwal & Dayton, 2024). Moreover, opioid-related discourse is inherently multilingual, with English and Spanish representing two of the most prominent languages in online discussions. Existing computational approaches overwhelmingly focus on English-only data, limiting their applicability across diverse user populations (Almeida et al., 2024; Lokala et al., 2024).

A second major limitation is the scarcity of high-quality, fine-grained annotated datasets for opioid-related NLP tasks. Most available datasets rely on keyword-based filtering or coarse-grained labels such as sentiment or topic categories (Carpenter et al., 2024; Almeida et al., 2024).

Recent work has explored more advanced classification schemes, including mixed-substance detection using large language models (Ahmad et al., 2025), but these efforts remain monolingual and do not address multilingual data integration. Fine-grained annotations capturing clinically meaningful information, such as overdose indicators or routes of administration, are rarely available, particularly in multilingual contexts.

A further challenge is the limited interpretability of deep learning models used in opioid-related text classification. Although such models achieve strong predictive performance (Lokala et al., 2024; Ahmad et al., 2025), their black-box nature restricts adoption in high-stakes public health settings, where transparency, accountability, and domain-expert validation are essential. As highlighted by Almeida et al. (2024), explainability remains an open challenge in opioid-related NLP research, and the absence of interpretable models limits their utility for clinical and policy decision-making.

This work addresses these challenges by developing a multilingual, fine-grained, and interpretable framework for opioid overdose detection using social media data. Specifically, we focus on classifying posts according to four clinically relevant routes of administration, intranasal, intravenous, oral, and non-overdose, across both English and Spanish content. To the best of our knowledge, no prior study has explored joint multilingual and translation-based modeling strategies for opioid overdose detection with route-specific annotations (Ahmad et al., 2025; Ahmad et al., 2025b; Ahmad et al., 2025c).

This work aims to create a novel machine learning framework for opioid crisis detection across multiple languages in social media posts. The following objectives have been devised to achieve these goals:

- Construct a large-scale English–Spanish opioid-related dataset using a hybrid annotation strategy that combines GPT-assisted labeling with expert human validation, producing fine-grained route-of-administration labels.
- Develop two complementary strategies, translation-based corpus unification and joint multilingual learning, to address cross-linguistic variability in opioid-related discourse.
- Develop a CNN–BiLSTM model augmented with a custom attention mechanism to capture both local lexical patterns and long-range contextual dependencies in multilingual text.
- Incorporate XAI techniques to provide interpretable model outputs, enabling domain experts to understand route-specific linguistic cues associated with overdose-related content.

The remainder of the paper is organized as follows. Section 2 reviews related work on opioid surveillance, multilingual NLP, and explainability. Section 3 describes the proposed methodology, including dataset construction, translation pipeline, and model architecture. Section 4 presents the experimental setup and evaluation results. Section 5 concludes the paper and outlines directions for future research.

2 State-of-the-Art

The application of natural language processing (NLP) to social media data has become a central approach for monitoring opioid- related behaviors and public health risks. Prior work has examined multiple dimensions of opioid discourse, including sentiment analysis, topic modeling, slang detection, and surveillance analytics. Sentiment analysis has been widely used to characterize public perceptions of opioid use and treatment (e.g., Carabot et al., 2023; Chenworth et al., 2021; Gandy et al., 2025; Ahmad et al., 2025), while distributional and contextual representation learning techniques, such as Word2Vec and transformer- based embeddings, have been applied to capture domain- specific language and informal expressions (e.g., Sarker et al., 2019; Holbrook et al., 2024; Lokala et al., 2024). Despite these advances, existing research remains constrained by monolingual focus, limited dataset granularity, and insufficient model interpretability.

2.1 Opioid Surveillance Using Social Media and NLP

A substantial body of work has leveraged platforms such as Reddit, Twitter, and YouTube for opioid surveillance and public health monitoring. Sentiment and perception analysis have been widely adopted to examine user attitudes toward opioid use and treatment. For instance, Chenworth et al. (2021) analyzed Twitter posts referencing methadone and buprenorphine- naloxone, identifying themes related to stigma, access, and treatment perceptions, with predominantly negative sentiment. Carabot et al. (2023) conducted a large- scale analysis of more than 250,000 opioid- related tweets, distinguishing between medical and nonmedical content and highlighting the prominence of fentanyl- related discussions. Gandy et al. (2025) evaluated automated sentiment analysis tools on opioid- related YouTube comments, demonstrating their utility while noting limitations in capturing nuanced public health discourse. Ahmad et al. (2025) extended this line of work by using large language models to detect mixed opioid and polysubstance use patterns, moving beyond coarse- grained

sentiment classification toward more context- aware analysis.

Beyond sentiment analysis, distributional and contextual representation learning techniques have been employed to capture opioid- related linguistic patterns. Sarker et al. (2019) demonstrated the utility of NLP for geolocation- centric monitoring of opioid discussions, highlighting the potential of social media for real- time surveillance. More recent studies have incorporated advanced deep learning architectures. Lokala et al. (2024) used transformer- based models to analyze sentiment, emotion, and topic associations in opioid- related content, providing deeper insights into user perceptions. Henson (2024) examined YouTube comments to analyze public discourse and policy implications surrounding fentanyl use. Holbrook et al. (2024) applied Word2Vec embeddings to identify opioid- related slang in Reddit discussions, significantly improving detection coverage and revealing polysubstance use patterns.

Collectively, these studies demonstrate that social media- based NLP methods provide valuable insights into opioid- related behaviors and perceptions. However, they remain largely limited to coarse- grained classification tasks and do not capture complex patterns such as overdose indicators, route- specific behaviors, or multilingual discourse.

2.2 Limitations of Monolingual Approaches

Despite methodological diversity, most existing studies rely exclusively on English- language datasets (e.g., Holbrook et al., 2024; Lokala et al., 2024). This monolingual bias restricts model generalizability and overlooks the multilingual nature of social media platforms. Spanish, in particular, represents a substantial portion of global online discourse, yet remains underrepresented in opioid- related NLP research. As a result, English- only models fail to capture linguistic nuances, slang variations, and cultural contexts present in non- English discussions. Despite the global nature of opioid- related discourse, multilingual or cross- lingual modeling strategies remain largely unexplored.

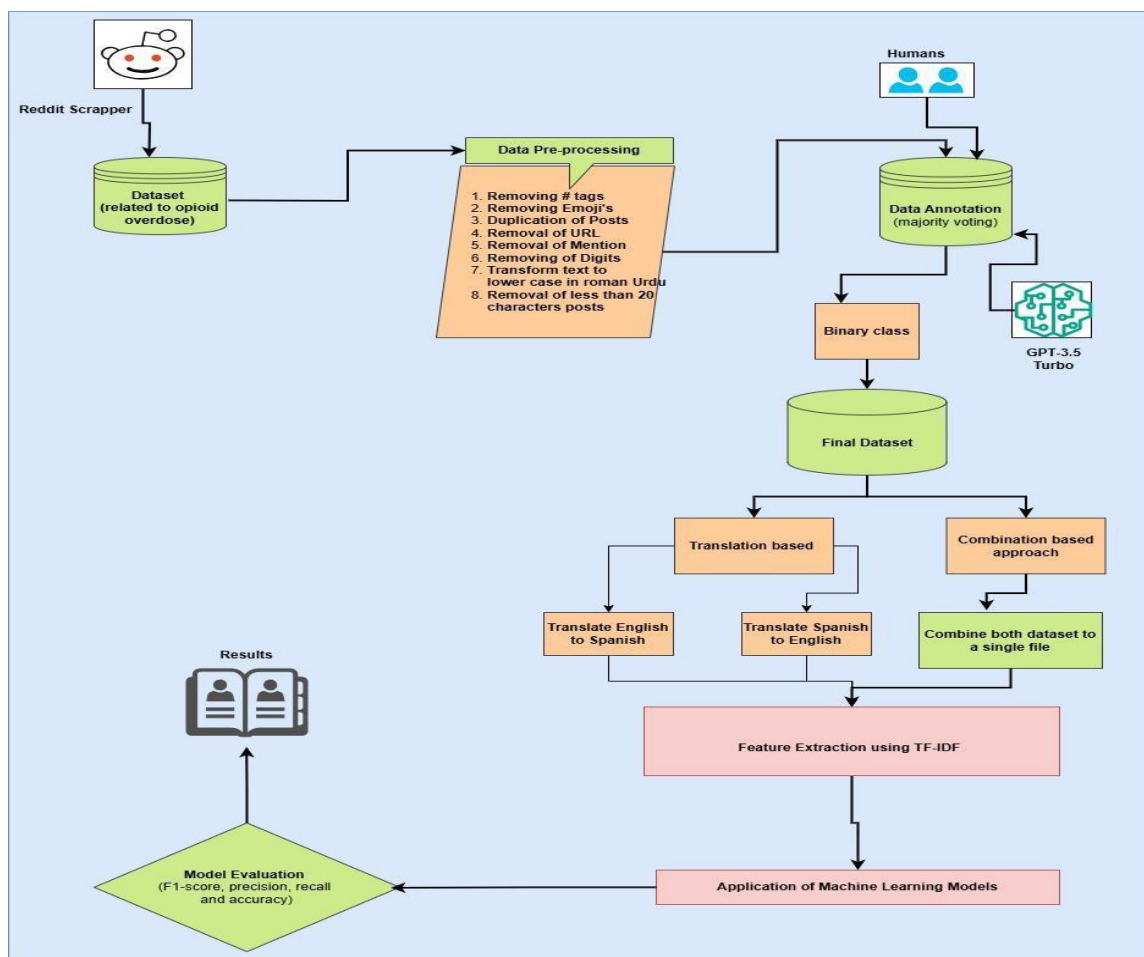


Fig. 1. Proposed architecture and design

2.3 Limited Availability of High-Quality Multilingual Datasets

A second major limitation is the scarcity of high-quality, fine-grained multilingual datasets for opioid-related NLP tasks. Existing datasets often rely on keyword-based filtering (e.g., Carabot et al., 2023) and weak supervision, focusing primarily on sentiment or topic categories. More recent work, such as Ahmad et al. (2025), has explored advanced classification schemes using large language models, but these efforts remain confined to single-language settings. Fine-grained annotations capturing clinically meaningful information, such as

overdose indicators or routes of administration, are rarely available, particularly in multilingual contexts. Moreover, multilingual dataset construction introduces challenges including translation noise, annotation consistency, and semantic ambiguity. These limitations hinder the development of robust multilingual models and impede progress toward real-world deployment.

2.4 Lack of Explainability

Although deep learning models have achieved strong performance in opioid-related text classification (e.g., Lokala et al., 2024; Ahmad et al., 2025), they typically operate as black-box systems with limited interpretability. This lack of

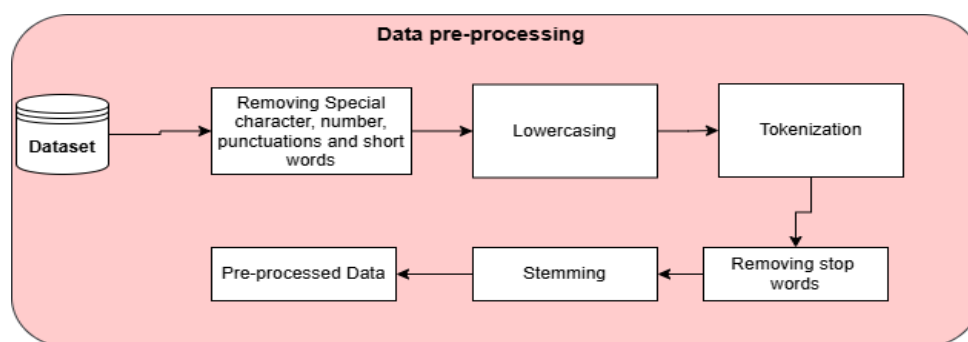


Fig. 2. Pre-processing steps used in this study

transparency is problematic in public health applications, where trust, accountability, and actionable insights are essential. Almeida et al. (2024) highlight interpretability as a persistent challenge in opioid-related NLP research. Without meaningful explanations, domain experts cannot validate model predictions or translate them into policy-relevant insights. This limitation restricts the practical adoption of such models in clinical and policy settings, where interpretability is essential for decision-making.

2.5 Summary of Research Gaps

In summary, existing literature is characterized by three major limitations: (i) monolingual focus with most studies rely on English-only data, limiting cross-linguistic applicability; (ii) limited multilingual datasets: lack of high-quality, fine-grained annotated multilingual data, and (iii) lack of explainability with minimal integration of XAI in opioid detection models. These limitations collectively highlight the need for a multilingual, fine-grained, and interpretable framework capable of capturing complex opioid-related patterns in diverse social media data—motivating the approach proposed in this work.

3 Methodology and Design

3.1 Dataset Construction

Reddit was selected as the primary data source because it hosts large volumes of opioid-related discourse distributed across topic-specific

communities. Using the Python Reddit API Wrapper (PRAW), approximately 80,000 posts were collected between January 2018 and February 2024 from subreddits known for discussions on opioid use and harm reduction, including r/opiates, r/OpiatesRecovery, and r/Drugs. To ensure adequate representation of Spanish-language content, a domain-specific multilingual keyword lexicon was constructed. This lexicon included formal drug names such as fentanilo, oxicodona, and morfina; common slang expressions such as chiva, perico, and pastillas; and route-specific terminology including esnifar (intranasal), inyectar (intravenous), introducir (rectal), and masticar (oral).

These terms were curated from Spanish harm-reduction resources, online drug-use forums, and community-maintained glossaries.

Posts were retrieved when they matched any term in the Spanish lexicon, after which language identification was performed using a language detection model to remove non-Spanish content, code-switched posts, and irrelevant matches (Ahmad et al., 2025) were incorporated to create a balanced multilingual corpus suitable for cross-lingual modeling. This combined dataset provided a linguistically diverse foundation for downstream annotation and model development.

3.2 Pre-Processing

All collected posts underwent a multilingual pre-processing pipeline designed to reduce noise while preserving opioid-related terminology and contextual cues. Each raw post x was first normalized through the removal of

Table 1. Example posts on opioid usage and routes of administration

Post	Label
English	
I started using heroin through this method, thinking it would be less risky. The initial high was intense, but short-lived. I overdosed a couple of times, and the nasal damage was significant over time	Intranasal
The first time I injected heroin, it was like a rush of relief and euphoria. It felt overwhelming, but after a few months, I overdosed twice. It was a terrifying experience not knowing if I'd wake up	Intravenous
I took Oxycodone orally, and the effects were slow but addictive. Over time, I increased my dosage, eventually overdosing on a higher-than-prescribed amount. It taught me how dangerous even oral use can be	Oral
Opioids are medications used to manage pain. They include drugs like morphine, oxycodone, and hydrocodone etc...	Not overdosed
Spanish	
Después de tomar analgésicos recetados para mi dolor crónico, comencé a depender cada vez más de ellos. No me di cuenta de lo fácil que era sufrir una sobredosis simplemente tomando demasiadas pastillas a la vez. La experiencia fue aterradora y me hizo darme cuenta de que los riesgos de los opioides orales son reales.	Oral
Una vez inhalé una mezcla de heroína y fentanilo, pensando que me daría el subidón perfecto. No me di cuenta de lo potente que era el fentanilo. Esto llevó a una sobredosis y apenas logré sobrevivir. Fue una llamada de atención de que los opioides pueden ser mortales en cualquier forma.	Intranasal
Inyectarme opioides se convirtió en mi manera de escapar de todo. Pasé de usarlos de vez en cuando a usarlos a diario. Tuve varias sobredosis, y cada vez fue más difícil regresar. El subidón era fugaz, y las consecuencias fueron devastadoras.	Intravenous

Table 2. Interpretation of the kappa values

Kappa Value	Interpretation
0.00 - 0.20	Poor Agreement
0.21 - 0.40	Fair Agreement
0.41 - 0.60	Moderate Agreement
0.61 - 0.80	Substantial Agreement
0.81 - 1.00	Almost Perfect Agreement

URLs, emojis, punctuation, and non- alphabetic characters.T

The text was then converted to lowercase to ensure case- invariant processing.

Tokenization was performed using a WordPiece- based subword tokenizer, which is well suited for handling slang, misspellings, and the morphologically rich structure of Spanish opioid terminology.

Following tokenization, language- specific stop words were removed using curated lists for both English and Spanish.

Stemming or lemmatization was applied using Snowball stemmers to reduce inflectional

variation and improve lexical consistency across posts.

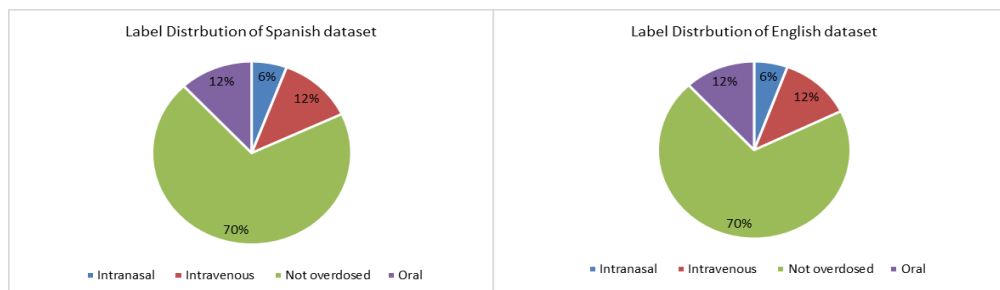
After applying the full pre- processing pipeline, a total of 5,053 posts, comprising 2,250 Spanish posts and 2,803 English posts, were retained for annotation and subsequent modeling.

3.3 Annotation

A hybrid annotation strategy was employed to generate fine- grained labels for the four overdose- related categories. The annotation process consisted of two stages.

Table 3. Dataset statistics

Language	Post	Total words	Total characters	Vocabulary size	Avg. Words per sentence	Avg. word per post	Avg characters per words	Avg. Character per post
English	5053	430111	1958913	18163	18.76	4.53	4.55	387.67
Spanish	5053	422637	2126932	22386	17.46	4.78	5.03	420.92

**Fig. 3.** Label distribution of multilingual dataset in English and Spanish

3.3.1 LLM-Assisted Initial Labeling

GPT- 3.5 Turbo was used to produce initial multi- class labels for all Spanish posts. The model was prompted with a structured schema defining the four classes and provided with examples to ensure consistency. LLM- assisted annotation enabled scalable labeling of nuanced opioid- related content, particularly for slang- rich Spanish posts.

3.3.2 Expert Human Validation

Two native Spanish- speaking PhD- level annotators independently reviewed all LLM- generated labels. Annotators followed a detailed guideline adapted from Ahmad et al. (2025), covering overdose cues, route- specific terminology, contextual indicators, and ambiguity resolution. Disagreements were adjudicated through consensus. The final annotated dataset comprised 5,053 posts across both languages.

3.4 Inter- Annotator Agreement

We employed Cohen's Kappa for our binary classification task because it is specifically designed to measure the agreement between two annotators. Since we used two human annotators

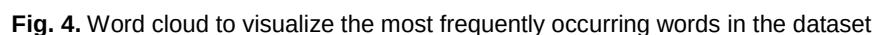
to verify the model's annotations, we utilized Cohen's Kappa to assess the level of agreement between them. In our case, the Cohen's Kappa value was 80%, which indicates a substantial level of agreement between the model's annotations.

3.5 Corpus Standardization and Statistics

To enable consistent multilingual analysis, the English and Spanish components of the dataset were standardized and examined for structural and linguistic differences.

Table 3 summarizes a comparative analysis of the 5,053 opioid- related posts and their translated counterparts. Although the number of posts is identical across languages, notable variations emerge in lexical and structural characteristics.

English posts contain a slightly higher total word count (430,111 vs. 422,637), whereas Spanish posts exhibit a substantially higher total character count (2,126,932 vs. 1,958,913). This discrepancy reflects the tendency of Spanish to employ longer word forms, a pattern further supported by the higher average word length observed in Spanish (5.03 characters) compared to English (4.55 characters).



multilingual dataset, confirming balanced representation across classes.

3.6 Ethical Considerations

This study was conducted using publicly available Reddit data, and all procedures adhered to platform policies governing the use of user-generated content.

Only anonymized text was collected, and no personal identifiers, such as usernames, profile metadata, or location information, were stored or analyzed at any stage of the research process.

The study was carried out solely for academic and public health research purposes, and all methodological steps align with widely accepted ethical standards for digital epidemiology and computational social science.

Computación y Sistemas, Vol. 30, No. 3, 2026, pp. 1809–1826
doi: 10.13053/CyS-30-3-6632

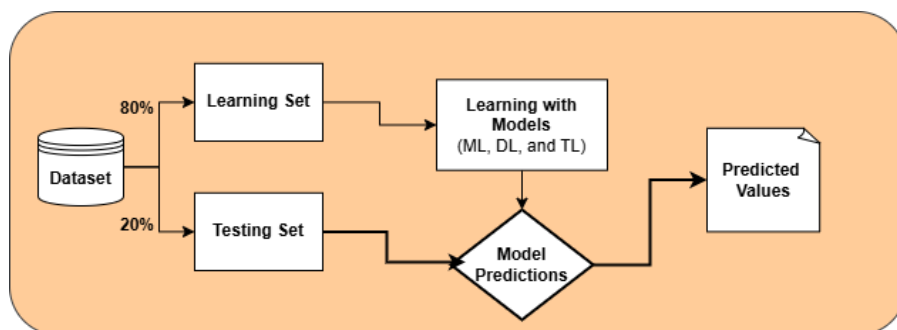


Fig. 5. Application of models, training and testing phase

3.7 Translation Steps

The translation-based component of the framework was designed to unify English and Spanish content into a consistent representation suitable for multilingual modeling. To achieve this, all posts were translated into a target language to create aligned corpora that support cross-lingual comparison and joint training. The resulting datasets were stored in standardized formats in which each translated post was paired with its corresponding route-of-administration label, enabling efficient downstream processing.

Before translation, each post underwent tokenization to segment the text into linguistically meaningful units. This step ensured that the translation models could more accurately interpret slang expressions, multi-word drug references, and route-specific terminology commonly found in opioid-related discourse. Following translation, all outputs were manually reviewed to identify and correct noisy or semantically inconsistent segments, particularly in cases involving idiomatic expressions or culturally specific slang that do not map cleanly across languages. This manual verification step was essential for preserving the intended meaning of the original posts and preventing semantic drift.

To maintain structural consistency across languages, Spanish posts were first translated into English to form a unified English-aligned corpus. A second corpus was then created by translating the English posts into Spanish, ensuring that each translated entry was aligned with its original counterpart. This bidirectional

translation process produced two parallel datasets that were comparable in content and quality supporting multilingual training, translation evaluation, and cross-lingual transfer learning.

Because deep learning models require inputs of consistent length, text length standardization was applied to the translated corpora. Posts exceeding the maximum allowable sequence length were truncated in a controlled manner to preserve the most informative segments while ensuring compatibility with the model architecture.

Through this translation pipeline, linguistic variation between English and Spanish was minimized, enabling consistent text processing and improving the robustness of the multilingual classification framework. Table 4 provides the pseudocode for the multilingual opioid-overdose classification workflow using the combined Spanish–English dataset within the WISE model architecture.

3.8 Model Training and Testing Phase

The classification framework was evaluated using both traditional machine learning models and deep learning architectures to assess their effectiveness in detecting opioid-related content across English, Spanish, and joint multilingual datasets. Classical machine learning models included Multinomial Naïve Bayes, K-Nearest Neighbors, Decision Trees, and AdaBoost. These models were trained using Term Frequency–Inverse Document Frequency (TF-IDF) representations, which convert each post into a weighted vector reflecting the relative importance

Table 4. Multilingual opioid overdose crises classification

Step	Procedure/Operation	Description
1	WISE-Multilingual-Opioid-Detection()	Procedure for multilingual opioid misuse classification using hybrid deep learning model
2	D1 ← Load-Spanish-Dataset (2,250 tweets)	Load Spanish dataset from Reddit
3	D2 ← Load-English-Dataset (2,803 tweets)	Load English dataset from Reddit
4	Eng-Translated-D1 ← Translate-To-English (D1)	Translate Spanish tweets into English
5	Spa-Translated-D2 ← Translate-To-Spanish (D2)	Translate English tweets into Spanish
6	Review-Translations (Eng-Translated-D1, Spa-Translated-D2)	Manually validate translations for slang, context, and accuracy
7	Combined-English ← Merge (Eng-Translated-D1, D2)	Merge Spanish-translated-to-English with original English
8	Combined-Spanish ← Merge (Spa-Translated-D2, D1)	Merge English-translated-to-Spanish with original Spanish
9	Joint-Multilingual-D ← Merge (Combined-English, Combined-Spanish)	Create a single multilingual dataset combining both language sets
10	Pre-Processed-D ← Preprocess(Joint-Multilingual-D)	Clean, normalize text (remove hashtags, URLs, emojis, punctuation, lowercase)
11	Tokenized-D ← Tokenizer(Pre-Processed-D)	Tokenize dataset using custom tokenizer (e.g., WordPiece or subword tokenization)
12	Embedding ← Embedding-Layer(Tokenized-D)	Convert tokens to embeddings (e.g., pretrained multilingual word vectors or trainable embeddings)
13	CNN-Output ← CNN(Embedding)	Apply 1D CNN to capture local n-gram patterns in the text
14	BiLSTM-Output ← BiLSTM(CNN-Output)	Feed CNN output into BiLSTM to capture bidirectional contextual dependencies
15	Attention-Output ← Custom-Attention(BiLSTM-Output)	Apply custom attention mechanism to focus on important features
16	Final-Predictions ← Dense-Layer(Attention-Output)	Fully connected layer for classification into 4 classes (routes of administration)
17	Train-Test-Split (80%-20%)	Split dataset into training and testing sets
18	Train WISE Model (Train-Data)	Train WISE model with categorical cross-entropy and Adam optimizer
19	Evaluate Model (Test-Data)	Evaluate model performance on unseen test data
20	Confusion-Matrix ← Generate-Results()	Generate confusion matrix from model predictions
21	Accuracy ← Compute-Accuracy(Confusion-Matrix)	Compute accuracy from confusion matrix
22	Precision ← Compute-Precision(Confusion-Matrix)	Compute precision metric
23	Recall ← Compute-Recall(Confusion-Matrix)	Compute recall metric
24	F1-Score ← Compute-F1(Confusion-Matrix)	Compute F1-score to measure balanced performance

of terms within the corpus. TF- IDF values were computed using standard formulations of term frequency and inverse document frequency, as shown in Equations (1)–(3).

Term Frequency (TF) is:

$$TF = \frac{\text{Number of times term } t \text{ appears in a document}}{\text{Total number of terms in the document}}. \quad (1)$$

Inverse Document Frequency (IDF) is:

$$IDF = \frac{\text{Number document in the corpus}}{\text{Number of document in the corpus contain terms}}. \quad (2)$$

TF-IDF is presented in equation 3:

$$TF - IDF = TF \times IDF. \quad (3)$$

These representations enable the classical models to capture discriminative lexical patterns associated with opioid misuse and overdose- related terminology.

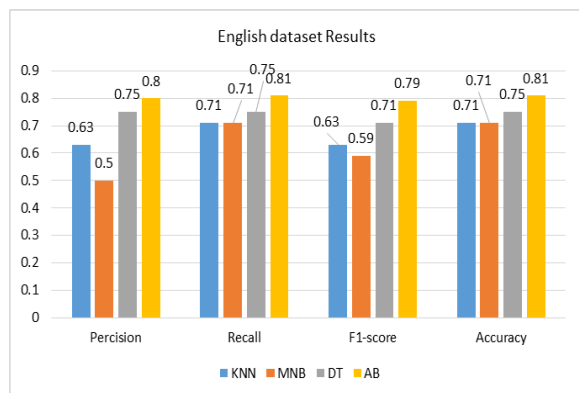


Fig. 6. Comparison of different machine learning models using the English dataset

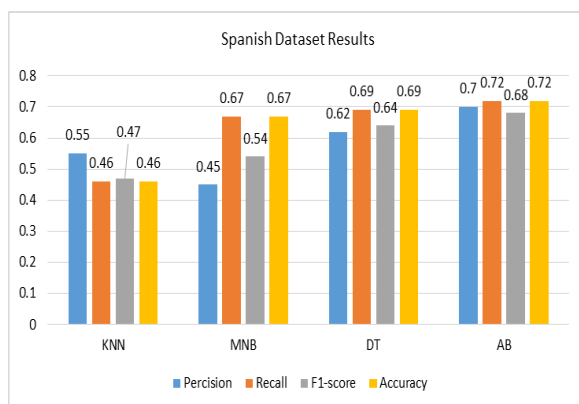


Fig. 7. Comparison of different machine learning models using the Spanish dataset

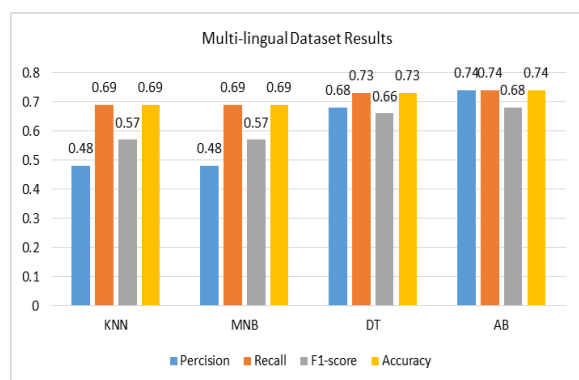


Fig. 8. Comparison of different machine learning models using the Multi-lingual dataset

In parallel, several deep learning architectures were explored to capture contextual and sequential dependencies that traditional models

may overlook. These included a Bidirectional Long Short- Term Memory (BiLSTM) network, a Convolutional Neural Network (CNN), and a hybrid CNN–BiLSTM model.

Each architecture was augmented with a custom attention mechanism designed to assign higher weights to semantically salient tokens within a post.

This attention layer enabled the models to focus on linguistically informative segments—such as route- specific expressions or overdose indicators—thereby improving their ability to detect nuanced opioid- related content in multilingual settings.

All models were trained and evaluated independently on three dataset configurations: the English corpus, the Spanish corpus, and the joint multilingual corpus.

For each configuration, the data was partitioned into an 80% training set and a 20% held- out test set. Model training was conducted using the respective feature representations and architecture- specific optimization procedures.

After training, each model was evaluated on the test set to assess its generalization performance.

Performance was quantified using four standard metrics: accuracy, precision, recall, and F1- score.

These metrics enabled a comprehensive assessment of both overall predictive performance and class- specific behavior.

The evaluation results provided insight into the comparative strengths of traditional and neural approaches, as well as the impact of multilingual training on model robustness.

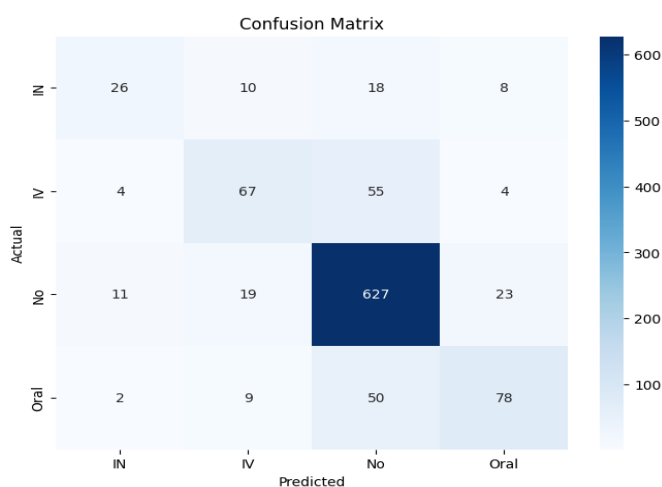
Collectively, this training and testing phase ensured that the proposed framework could reliably classify opioid overdose- related content across diverse linguistic inputs, supporting its applicability for real- time public health surveillance.

4 Results and Analysis

In this section, we will explore the results of the experiments conducted to assess the performance of the proposed methodology. The analysis focuses on the accuracy and

Table 7. Performance of attention-enhanced deep learning models

Language	Model	Precision	Recall	F1-score	Accuracy
English	BiLSTM	0.82	0.82	0.82	0.82
Multilingual		0.78	0.78	0.78	0.78
Spanish		0.75	0.77	0.76	0.77
English	WISE	0.84	0.84	0.84	0.84
Multilingual		0.81	0.82	0.81	0.82
Spanish		0.78	0.79	0.78	0.79
English	CNN	0.81	0.81	0.81	0.81
Multilingual		0.75	0.77	0.75	0.77
Spanish		0.73	0.75	0.74	0.75

**Fig. 9.** Confusion matrix of Spanish translation dataset

effectiveness of various machine learning models across joint multi-lingual and joint translation dataset.

4.1 Results for Machine Learning

Figure 6 shows the performance evaluation of four different machine learning models such as K-Nearest Neighbors (KNN), Multinomial Naïve Bayes (MNB), Decision Tree (DT), and AdaBoost (AB) in forecasting opioid overdose cases using joint translation including English dataset. These models are evaluated based on four different performance metrics such as Precision, Recall, F1-score, and accuracy. Among all models,

AdaBoost (AB) performed best and achieved the highest Precision=0.8, Recall 0.81, F1-score=0.79, and accuracy=0.81, showing its strong predictive ability.

On the other hand, Decision Tree (DT) follows with an F1-score of 0.71 and accuracy of 0.75 and KNN and MNB show relatively lower performance, with MNB having the lowest Precision (0.5) and F1-score (0.59), though both models share the same Recall (0.71) and accuracy (0.71).

These results suggest that boosting-based models like AdaBoost may be more effective in handling the dataset for opioid overdose prediction.

The Figure 7 displays the performance of four machine learning models—K-Nearest Neighbors (KNN), Multinomial Naïve Bayes (MNB), Decision Tree (DT), and AdaBoost (AB)—in predicting opioid overdose cases within a Spanish dataset. Among the models, AdaBoost (AB) performs the best, achieving the highest Precision (0.7), Recall (0.72), F1-score (0.68), and accuracy (0.72), suggesting it is the most effective in handling the dataset. Decision Tree (DT) follows with an F1-score of 0.64 and accuracy of 0.69, showing relatively strong predictive performance. MNB has a high Recall (0.67) but lower Precision (0.45) and an F1-score of 0.54, indicating it identifies more positive cases but with reduced accuracy. KNN performs the weakest, with the lowest Recall (0.46), F1-score (0.47), and accuracy (0.46), suggesting it struggles with distinguishing opioid overdose cases effectively. Overall, boosting-based models like AdaBoost seem to perform best in this dataset.

The Figure 8 presents the performance evaluation of four machine learning models—K-Nearest Neighbors (KNN), Multinomial Naïve Bayes (MNB), Decision Tree (DT), and AdaBoost (AB)—for opioid overdose prediction in a multilingual dataset. AdaBoost (AB) achieves the highest performance, with a Precision of 0.74, Recall of 0.74, F1-score of 0.68, and accuracy of 0.74, making it the most effective model. Decision Tree (DT) follows closely with an F1-score of 0.66 and accuracy of 0.73, showing strong predictive ability. KNN and MNB perform similarly, both having a Precision of 0.48, Recall of 0.69, F1-score of 0.57, and accuracy of 0.69. This indicates that boosting-based models like AdaBoost and tree-based approaches like Decision Tree yield better results for multilingual opioid overdose prediction compared to KNN and Naïve Bayes.

4.2 Results for Deep Learning Models

The procedure pipeline shown in Table 5 outlines a comprehensive process for building a deep learning-based text classification model. It begins with loading the dataset from an Excel file and proceeds to extract the raw text and their corresponding labels. These labels are then transformed into numeric form and further

converted into one-hot encoded vectors to prepare them for multi-class classification.

The text data is tokenized into sequences of word indices and padded to ensure uniform input length. The dataset is then split into training and testing sets. A custom attention mechanism is defined to help the model focus on the most relevant parts of each sequence by computing attention scores and context vectors. The model architecture includes an input layer, an embedding layer to capture word meanings, a convolutional layer for local feature extraction, a bidirectional LSTM for sequential understanding, and the attention layer to emphasize important information. After a dropout layer to prevent overfitting, the model ends with a softmax-activated dense layer for prediction. It is compiled using the Adam optimizer and categorical cross-entropy loss, and trained over 10 epochs with a portion of the training data reserved for validation to monitor performance during training.

4.3 Hyper-Parameter Tuning

To fine-tune our hybrid deep learning model incorporating a custom attention mechanism are shown in table 6. We carefully selected and configured various hyperparameters across its key components. The embedding layer uses 128-dimensional word embeddings to represent textual input. A tokenizer limits the vocabulary to 10,000 most frequent words and truncates or pads sequences to a maximum length of 100 tokens. The CNN layer applies 128 filters with a kernel size of 5 to extract local n-gram features. Following this, a BiLSTM layer with 64 units captures long-range contextual dependencies in both forward and backward directions.

The custom attention mechanism is designed to focus on informative parts of the sequence, utilizing a learnable weight matrix of shape (feature_dim, 1) and a bias vector of shape (sequence_len, 1), with a tanh activation function and softmax normalization for computing attention scores. A dropout layer with a rate of 0.5 is applied after the attention context vector to prevent overfitting. The output layer adapts its number of units based on the number of classes and uses a softmax activation for multi-class classification. For training, we use a batch size of

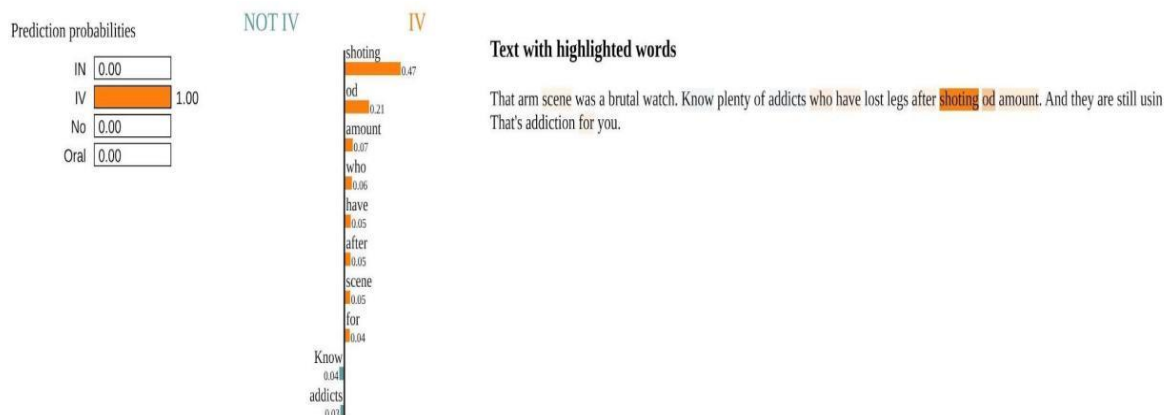


Fig. 12. LIME visualization illustrating the key textual features contributing to the proposed model's prediction of the intravenous (IV) route of opioid administration

32, 10 epochs, the Adam optimizer, and categorical_crossentropy as the loss function, with 10% of the training data reserved for validation. These configurations aim to ensure effective learning, prevent overfitting, and improve model generalization across multiclass NLP tasks.

4.4 Result for the Model

Table 7 presents a detailed comparison of three deep learning models—BiLSTM, CNN+BiLSTM, and CNN, each enhanced with a custom attention mechanism—for classifying opioid-related discussions in English, Spanish, and Multilingual datasets. The evaluation relies on four key metrics: Precision, Recall, F1-score, and Accuracy. These models were trained on a carefully curated multilingual dataset, reflecting real-world opioid-related discourse collected from Reddit.

Among all the models, CNN+BiLSTM demonstrates the best overall performance. For the English dataset, it achieves the highest scores across all metrics, with a Precision, Recall, F1-score, and Accuracy of 0.84.

It also performs well in the multilingual setting, achieving an accuracy of 0.82, and performs reasonably well on Spanish data with 0.79 accuracy. These results indicate that combining CNN and BiLSTM helps the model learn both

local patterns (via CNN) and long-term dependencies (via BiLSTM), making it highly effective for the task.

The BiLSTM-only model performs slightly below CNN+BiLSTM but still maintains solid results, especially for English, with an accuracy of 0.82. Its performance in multilingual (0.78) and Spanish (0.77) settings is lower but still competitive. Similarly, the CNN-only model shows comparable results for English (0.81) but performs slightly worse on multilingual (0.77) and Spanish (0.75) datasets.

In general, all models show the highest performance when trained and tested on English data, likely due to the higher availability and clarity of English content compared to Spanish or combined multilingual data.

The results suggest that while deep learning models can effectively classify opioid-related content across languages, performance is still affected by language-specific nuances and data characteristics.

The CNN+BiLSTM model, however, proves to be the most robust across all settings, making it a strong candidate for real-world multilingual opioid crisis monitoring systems. While Figure 9, Figure 10, and Figure 11 present the confusion matrices of our proposed hybrid CNN+BiLSTM model integrated with a custom attention mechanism. These confusion matrices provide a clear insight into how well the model distinguishes between

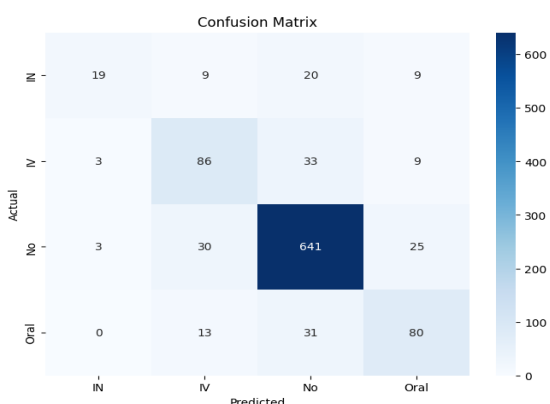


Fig. 10. Confusion matrix of joint multilingual dataset

classes and where it tends to misclassify, helping to analyze its overall performance.

4.5 Interpretability with LIME

Figure 12 illustrates the LIME (Local Interpretable Model-agnostic Explanations) analysis of the proposed model's prediction for the intravenous (IV) route of opioid administration. LIME provides local interpretability by identifying the words and phrases in a given social media post that most strongly influence the model's decision. It works by perturbing the input text and observing how these changes affect the model's output, thereby highlighting the features that contribute positively or negatively to a specific class prediction. This enables a transparent understanding of why the model classified the post as a particular route of administration rather than treating the prediction as a black box.

In the given text, the model predicts the intravenous (IV) route of administration. The LIME explanation shows that certain lexical cues have a strong positive contribution toward the IV class. Specifically, terms such as “shooting” (a colloquial spelling of shooting) and contextual references to severe limb damage (e.g., losing arms or legs) receive high positive weights. These features are well-known indicators of injection-based opioid use and are therefore strongly associated with intravenous administration in the training data.

From a semantic perspective, the word “shooting” is widely used in drug-related

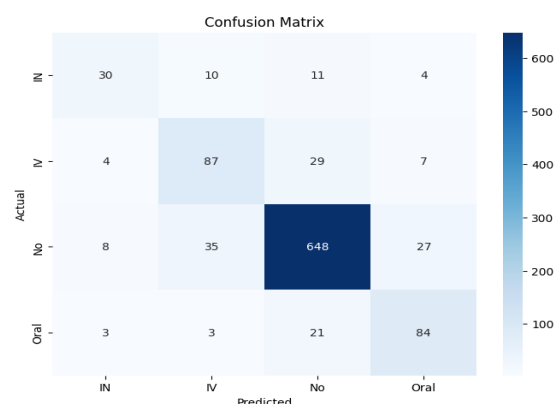


Fig. 11. Shows the confusion matrix of English translation dataset

discourse to describe injecting drugs directly into a vein. Unlike intranasal use (snorting) or smoking, injection is frequently linked with extreme physical consequences such as infections, abscesses, tissue necrosis, and limb loss. The model captures this association by attending more strongly to these cues through its custom attention mechanism, which amplifies medically and contextually salient terms while down-weighting less informative background words.

Overall, the LIME-based explainability confirms that the model's prediction of the intravenous route is not arbitrary. Instead, it is driven by meaningful, human-interpretable features that align with real-world drug-use terminology and clinical outcomes. This strengthens confidence in the model's decision-making process and demonstrates that the hybrid architecture with custom attention effectively learns domain-relevant patterns from noisy multilingual social media text.

5 Conclusions and Future Work

In conclusion, this research shows a novel framework for detecting opioid overdose discussions on social media platforms, in English and Spanish using the joint multi-lingual and joint translation based. Our proposed approach addresses the issues of evolving language and slang, offering a more accurate and scalable

solution for identifying opioid-related content across diverse linguistic communities. By utilizing the power of various machine learning models, such as AdaBoost, and conducting extensive experiments on multi-lingual datasets, the study validates the efficiency of the proposed framework. According to our experimentation the high classification accuracies attained for English translation dataset=81%, Spanish translation dataset=72%, and joint-multilingual datasets =74% emphasize the potential of this method to help policymakers and healthcare professionals in monitoring drug abuse trends globally, with minimal human intervention.

Future work will explore the integration of advanced deep learning techniques to further enhance the detection and classification of opioid drug consumption patterns. Specifically, leveraging state-of-the-art transfer learning models, such as pretrained transformer-based architectures (e.g., BERT, XLM-R, and RoBERTa), will enable the framework to capture complex linguistic nuances, including evolving slang, abbreviations, and code-mixed language commonly observed on social media platforms.

Author Contributions: Conceptualization, M.A., S.L., and G.A.; methodology, M.A. and I.B.; software, M.A.; validation, M.A., S.L., G.A., and G.S.; formal analysis, M.A., V.B., and N.K.; investigation, M.A., L.C.-H., and S.L.; resources, G.A., R.O., and G.S.; data curation, M.A. and L.C.-H.; writing—original draft preparation, M.A.; writing—review and editing, S.L., G.A., R.O., V.B., N.K., L.C.-H., I.B., and G.S.; visualization, M.A.; supervision, R.O., I.B., and G.S.; project administration, G.A. and G.S.; funding acquisition, R.O., G.S., and I.B. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data are available upon request.

Acknowledgments

The work was done with partial support from grants 20260626 (G.S.), 20260367 (I.B.) by Secretary of Research and Posgraduate Studies (SIP) of Instituto Politécnico Nacional, Mexico.

References

1. **Meier, B. (2003).** Pain Killer: A "Wonder" Drug's Trail of Addiction and Death. Rodale Publishing.
2. **Young, S.D., Lee, S.J., Perez, H. (2020).** Social Media as an Emerging Tool for Reducing Prescription Opioid Misuse Risk Factors. *Heliyon*, Vol. 6, No. 3. DOI: 10.1016/j.heliyon. 2020.e03481.
3. **Kulkarni, M.B., Castelino, M.A., Aher, M.S. (2020).** Painkillers: Necessity or Habit. *International Journal of Indian Psychology*, Vol. 8, No. 4.
4. **Schatman, M.E., Ziegler, S.J. (2017).** Pain Management, Prescription Opioid Mortality, and the CDC: Is the Devil in the Data?. *Journal of Pain Research*, Vol. 10, pp. 2489–2495. DOI: 10.2147/jpr.s145675.
5. **Olsen, Y. (2016).** The CDC Guideline on Opioid Prescribing: Rising to the Challenge. *JAMA*, Vol. 315, No. 15, pp. 1577–1579. DOI: 10.1001/jama. 2016.1464.
6. **Robinson, A., Christensen, A., Bacon, S. (2019).** From the CDC: The Prevention for States Program: Preventing Opioid Overdose through Evidence-Based Intervention and Innovation. *Journal of Safety Research*, Vol. 68, pp. 231–237. DOI: 10.1016/j.jsr.2018.10.005.
7. **Balsamo, D. (2022).** The Opioid Epidemic Through the Lens of Social Media: Mining Digital Traces for Public Health. *Pharmacovigilance, and Rehabilitation*, Doctoral Dissertation.
8. **Carpenter, K.A., Nguyen, A.T., Smith, D.A. (2024).** Which Social Media Platforms Provide the Most Informative Data for Monitoring the Opioid Crisis? *medRxiv*, pp. 1–24. DOI: 10.1101/2024.07.08.24310067.
9. **Peres, M.A., Macpherson, L.M.D., Weyant, R.J. (2019).** Oral Diseases: A Global Public Health Challenge. *The Lancet*, Vol. 394, No. 10194, pp. 249–260. DOI: 10.1016/s0140-6736(19)31146-8.
10. **Carabot, F., Fraile-Martínez, O., Donat-Vargas, C. (2023).** Understanding Public

- Perceptions and Discussions on Opioids Through Twitter: Cross-Sectional Infodemiology Study. *Journal of Medical Internet Research*, Vol. 25, pp. 50013. DOI: 10.2196/50013.
11. **World Health Organization (2009).** Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks, Geneva, Switzerland. World Health Organization.
 12. **Chenworth, M., Perrone, J., Love, J.S. (2021).** Methadone and Suboxone Mentions on Twitter: Thematic and Sentiment Analysis. *Clinical Toxicology*, Vol. 59, No. 11, pp. 982–991. DOI: 10.1080/15563650.2021.1881588.
 13. **Gandy, L.M., Ivanitskaya, L.V., Bacon, L.L. (2025).** Public Health Discussions on Social Media: Evaluating Automated Sentiment Analysis Methods. *JMIR Formative Research*, Vol. 9, pp. e57395. DOI: 10.2196/57395.
 14. **Sarker, A., Gonzalez-Hernandez, G., Ruan, Y. (2019).** Machine Learning and Natural Language Processing for Geolocation-Centric Monitoring and Characterization of Opioid-Related Social Media Chatter. *JAMA Network Open*, Vol. 2, No. 11, pp. e1914672. DOI: 10.1001/jamanetworkopen.2019.14672.
 15. **Ahmad, M., Batyrshin, I., Sidorov, G. (2025).** Sentiment Analysis Using a Large Language Model-Based Approach to Detect Opioids Mixed with Other Substances via Social Media: Method Development and Validation. *JMIR Infodemiology*, Vol. 5, pp. 70525. DOI: 10.2196/70525.
 16. **Gubler, D.J. (2002).** The Global Emergence/Resurgence of Arboviral Diseases as Public Health Problems. *Archives of Medical Research*, Vol. 33, No. 4, pp. 330–342. DOI: 10.1016/s0188-4409(02)00378-8.
 17. **Kim, S.J., Marsch, L.A., Hancock, J.T. (2017).** Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. *Journal of Medical Internet Research*, Vol. 19, No. 10, pp. 353. DOI: 10.2196/jmir.6426.
 18. **Weeks, M.R., Clair, S., Borgatti, S.P. (2002).** Social Networks of Drug Users in High-Risk Sites: Finding the Connections. *AIDS and Behavior*, Vol. 6, No. 2, pp. 193–206. DOI: 10.1023/a:1015481501692.
 19. **Ahmad, M., Batyrshin, I., Sidorov, G. (2025).** Sentiment Analysis Using a Large Language Model-Based Approach to Detect Opioids Mixed with Other Substances via Social Media: Method Development and Validation. *JMIR Infodemiology*, Vol. 5, pp. 70525. DOI: 10.2196/70525.
 20. **Lokala, U., Phukan, O.C., Dastidar, T.G. (2024).** Detecting Substance Use Disorder Using Social Media Data and the Dark Web: Time- and Knowledge-Aware Study. *JMIRx Med*, Vol. 5, pp. e48519. DOI: 10.2196/48519.
 21. **Silwal, A., Dayton, Z.A. (2024).** Understanding the Political Frames of the Opioid Crisis: A Content Analysis of US Senators' Tweets. *Health Communication*, pp. 1–15.
 22. **Holbrook, E., Wiskur, B., Nagykaldi, Z. (2024).** Discovering Opioid Slang on Social Media: A Word2Vec Approach With Reddit Data. *Drug and Alcohol Dependence Reports*, Vol. 13, pp. 100–302. DOI: 10.1016/j.dadr.2024.100302.
 23. **Almeida, A., Patton, T., Conway, M. (2024).** The Use of Natural Language Processing Methods in Reddit to Investigate Opioid Use: Scoping Review. *JMIR Infodemiology*, Vol. 4, pp. 51156. DOI: 10.2196/51156.
 24. **Henson, T.D. (2024).** Fentanyl, Social Media, and Health Policy: A Mixed-Methods Health Policy Analysis of Public Comments on YouTube Videos About Fentanyl, Master's Thesis, Central Michigan University.
 25. **Holbrook, E., Wiskur, B., Nagykaldi, Z. (2024).** Discovering Drug Slang on Social Media: A Word2Vec Approach With Reddit Data. Preprint.
 26. **Ahmad, M., Waqas, M., Hamza, A. (2025).** UA-HSD-2025: Multi-Lingual Hate Speech Detection from Tweets Using Pre-Trained Transformers, *Computers*. Vol. 14, No. 6, pp. 239. DOI: 10.3390/computers14060239.

- 27. Ahmad, M., Sidorov, G., Amjad, M. (2025).** Opioid Crisis Detection in Social Media Discourse Using Deep Learning Approach. *Information*, Vol. 16, No. 7, pp. 545. DOI: 10.3390/info16070545.
- 28. Ahmad, M., Waqas, M., Hamza, A. (2025).** Hope Speech Detection in Code-Mixed Roman Urdu Tweets: A Positive Turn in Natural Language Processing. *arXiv Preprint*. DOI: arXiv:2506.21583.

- 29. Ahmad, M., Ullah, F., Usman, M. (2025).** Leveraging Large Language Models for Multi-Class and Multi-Label Detection of Drug Use and Overdose Symptoms on Social Media, *arXiv Preprint*. DOI: arXiv:2504.12355.

Article received on 19/06/2025; accepted on 10/09/2026.
**Corresponding author is Ildar Batyrshin.*