

ADVERSE DRUG EFFECT MONITOR



PROBLEM STATEMENT

Traditional ADR reporting systems like **FAERS** and clinical trials often miss real-world patient feedback. Social media and online platforms contain valuable signals on **side effects, patient concerns, and drug effectiveness**, but extracting this information is challenging due to **noisy, unstructured data**. Thus, there is a need for an **automated, multi-source ADR detection system** that provides clear insights and risk assessment for drugs.

LITERATURE REVIEW

Paper	Problem	Method	Dataset	Findings	Limitations
Zhang et al. (2020)	ADR detection needs richer linguistic features	Predicate–ADR pairs + semantic + pooling (deep + shallow)	Twitter (5,076), DailyStrength (3,705)	AUC: 94.4% (DS), 89% (Twitter); better than CNN/baselines	Lacks advanced semantic & domain features
Oyebode & Orji (2023)	ADR detection without annotated data	ADR Framework (MetaMap + UMLS; lexicon + NLP rules)	6,797 reviews (AskAPatient, WebMD, Iodine)	Found 2,572 ADRs incl. new ones; compared by gender/age; benchmarked with SIDER	Weak on slang/informal text; limited causality
Elbiach et al. (2021)	Evaluate ML models under imbalance	9 ML models (NB, SVM, RF, etc.) with SMOTE/undersampling	CADEC, ADE corpus, TwiMed	Naïve Bayes best (F1: 94% CADEC; 79% ADE; 65% PubMed)	Small corpora; poor real-world generalization
Spandana & Prakash (2024)	Enhance ADR detection with deep learning	Deep CNN + statistical, sentiment & medical keyword features	ADE-Corpus V2 (13,981), PubMed (32,684)	High accuracy/sensitivity/specificity; outperformed baselines	Black-box; low interpretability; no real-time use

Paper	Problem	Methodology	Dataset	Findings	Limitations
Omar & Harris (2016)	ADR systems under-reporting; testing social media as supplement	Cross-sectional survey of public, HCPs, pharma	100 public, 46 HCPs, 17 pharma (Europe)	Positive attitude: 83% public, 71% pharma supportive; privacy concerns	Europe-only; excludes offline users
Huang (2022)	Social media ADR detection hampered by imbalance & informal text	ML vs. BERT + BAW for imbalance handling	6,842 tweets (737 ADRs)	BERT+BAW best (F1=0.55, AUC=0.87)	Single dataset; no cross-lingual test
Davis (2024)	Extracting & normalizing ADEs to MedDRA from tweets	spaCy NER + RoBERTa for mapping	#SMM4H 2024: 30,949 tweets	Recall 37.4%, effective extraction	Low precision (8%); many false positives
Golder et al. (2023)	No consensus on social media's value in PV	Scoping review protocol across 11 DBs	Planned literature review	Goal: map evidence, ID gaps	Limited to 4 languages; fast-changing field

RESEARCH GAP

❑ Informal & Noisy Language

- ADR mentions in social media often include slang, abbreviations, and misspellings → current systems fail to capture these.

❑ Limited & Imbalanced Datasets

- Few annotated ADR corpora exist; class imbalance (few ADRs vs. many non-ADR posts) weakens model performance.

❑ Lack of Context & Causality Detection

- Existing models detect keywords but miss **implicit meaning, sarcasm, negations**, and fail to confirm drug → reaction causal links.

❑ Weak Real-world Integration

- Most methods are tested in isolation, with **poor cross-domain generalization** and little connection to **official pharmacovigilance systems (FAERS, EMA, SIDER)**.

COMPARISON OF PROPOSED ADR MINING SYSTEM WITH ZHANG ET AL. (2020)

Feature / Criteria	Zhang et al. (2020)	Proposed ADR Mining Project	Why Our Is Better
Objective	Detect ADRs using deep linguistic features	Detect ADRs + Sentiment + Reporting + Visualization	More comprehensive
Model Used	Deep learning pipeline (predicate-ADR pairing + semantic features + pooling)	Biomedical NER (HuggingFace) + RoBERTa Sentiment + ADR Normalization	Uses modern transformer-based NER suitable for biomedical domain
Data Sources	Single-source datasets: Twitter (5,076), DailyStrength (3,705)	Multi-source: Kaggle (250k), WebMD, Drugs.com, Reddit, Web Search	Much larger, diverse, real-world dataset
Dataset Size	~8,781 records total	250,000+ reviews	30× bigger, more representative
Real-World Review Coverage	Limited to short tweets + forum	Covers forums, medical websites, long reviews, Reddit threads	Better ADR extraction due to richer context
ADR Entity Extraction	Pattern-based (predicate-ADR pairs)	Biomedical NER trained on scientific + clinical corpora	Higher accuracy, context-aware extraction
Sentiment Analysis	Not included	Included using transformer model	Adds safety evaluation + patient satisfaction insights
ADR Cleaning & Normalization	Not performed	Dictionary-based normalization + typo correction + noise removal	More accurate ADR grouping
Visual Outputs	No visualization	Dashboard with charts, top ADRs, drug-wise analysis	Useful for pharmacovigilance monitoring
Automated Reports	Not available	Generates per-drug PDF safety reports	Complete end-to-end system
Performance	AUC 89–94% on small datasets	ADR Accuracy ~86% on 250k real-world data	Tested on larger, more diverse dataset
Scalability	Limited due to dataset constraints	Highly scalable: pipeline + dashboard + multi-source ingestion	Suitable for real-world deployment
Explainability	Lower (deep feature model)	Shows extracted ADR entities + sentiment + frequency	Interpretable + transparent
Overall Advantage	Good linguistic modeling but limited context	Complete pharmacovigilance tool with multi-source ADR mining	Broader scope + practical real-world utility

PROPOSED METHODOLOGY

1. Data Sources

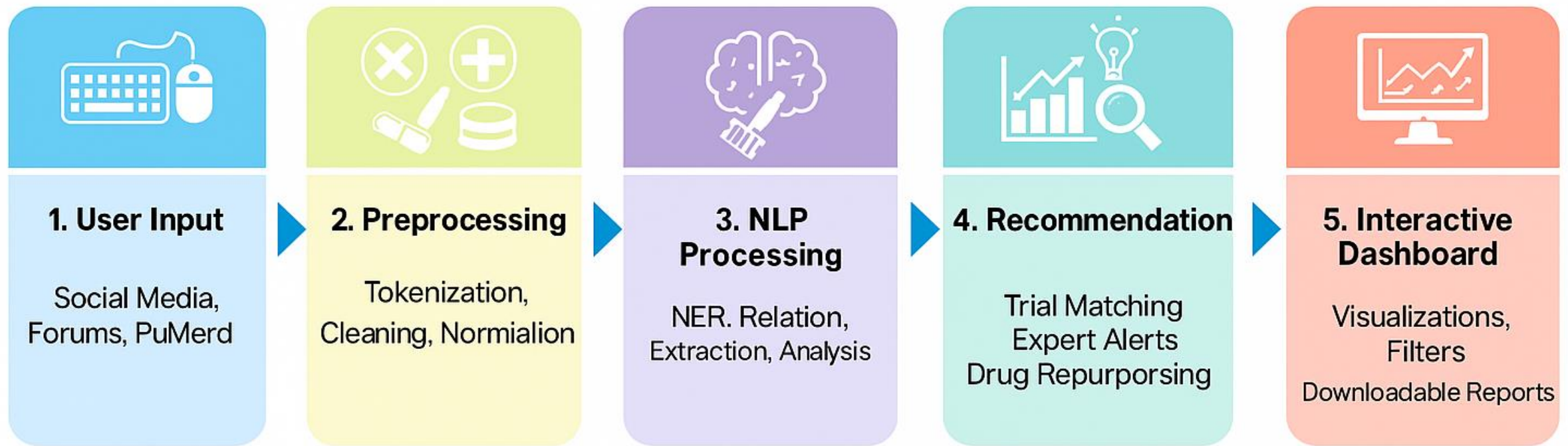
- **Kaggle** – structured drug reviews
- **Reddit API** – real-world patient posts
- **Web Search API** – medical forums & sites

2. Techniques & Tools

- **Python, Streamlit** (dashboard)
- **Hugging Face Transformers**
- **NER** → d4data/biomedical-ner-all
- **Sentiment** → twitter-roberta-base-sentiment
- **ADR Lexicon** → adr_keywords.json
- **Visualization** → Matplotlib, Streamlit charts

3. Recommendation Logic

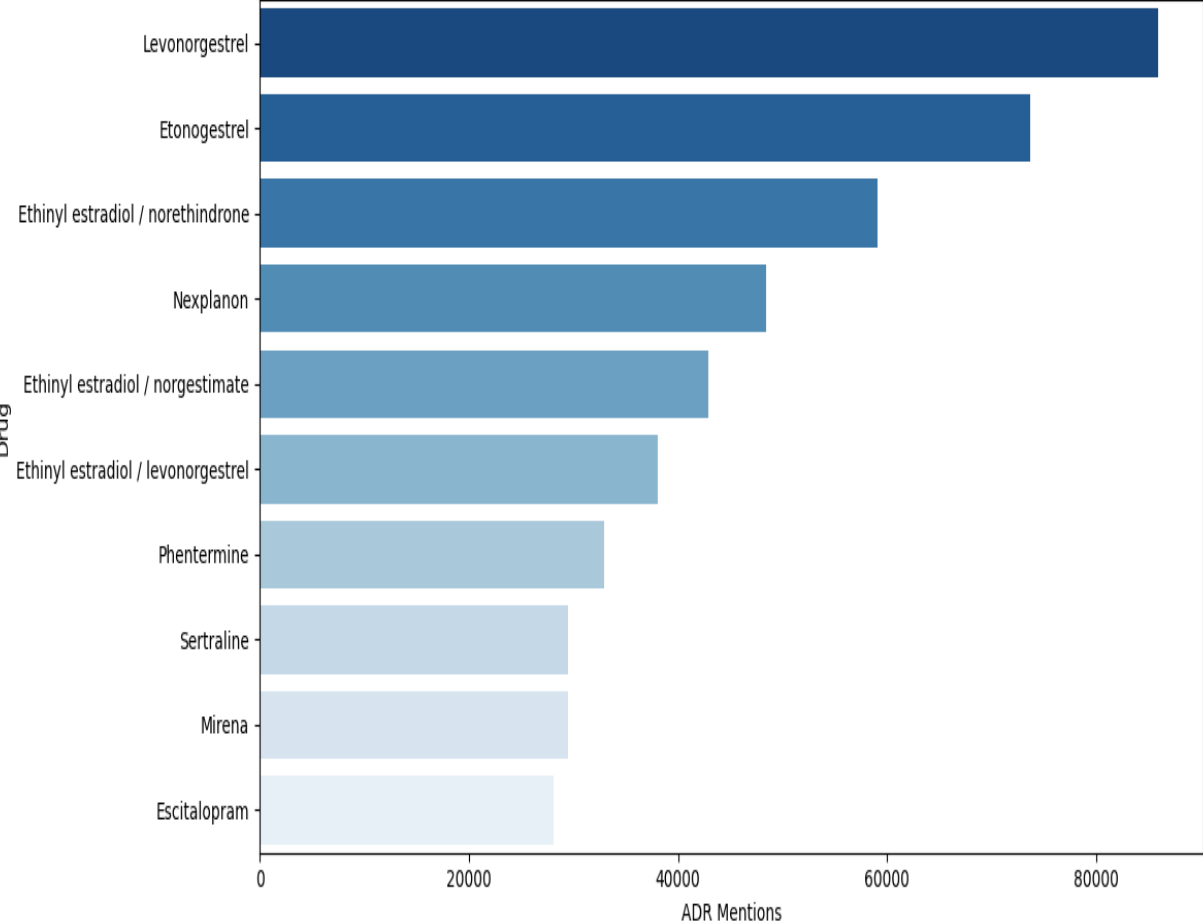
- **Safe** – low ADR mentions
- **Monitor** – $\text{ADR} \leq 80\%$ but rising concerns
- ⚠ **Caution** – $\text{ADR} > 80\%$ with mixed sentiment
- **Risky** – $\text{ADR} > 90\%$ + strong negatives



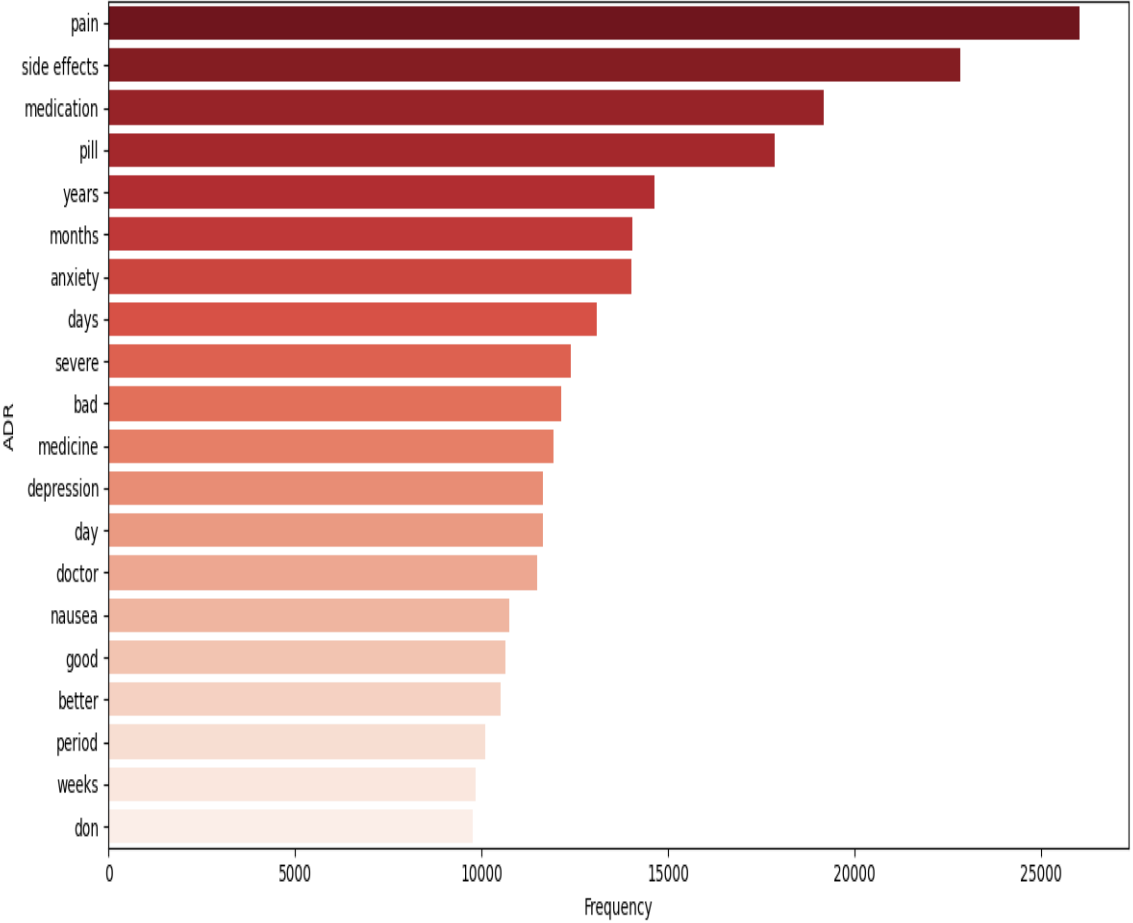
SYSTEM WORKFLOW

RESULTS

Top 10 Drugs with Most ADR Mentions



Top 20 Most Common ADRs



INSIGHT EACH DRUG REPORT

- Use different dataset : 5 lakh rows
- Total each drug reports : 3435

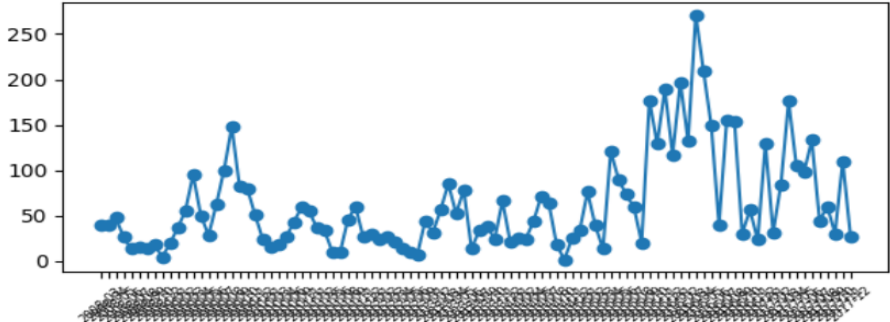
Drug Safety Report: Levofloxacin

Total Reviews	327
ADR Reviews	323
% ADR Mentions	98.78%
Positive %	15.8%
Negative %	59.1%
Mild ADRs	89
Serious ADRs	4

Top ADRs

ADR	Mentions
days	91
infection	84
pain	77
side effects	66
day	66
lev	64
levaquin	63
antibiotic	57
medicine	50
medication	48

Monthly ADR Mentions for Levofloxacin



THANK YOU

Presented By: Group 6

Bhavuk

Jatin Phogat

Daksh

