RESEARCH PAPER

Extracting Medical Opinion from Microblogs to Perform Aspect Based Sentiment Analysis

T. Anuprathibha[1], P. Karthikeyan [2], N. Mohan Raj[3]
[1]Assistant Professor, Department of Information Technology, V.S.B. Engineering College, Karur, TamilNadu, India. Email: anuprathiba16@gmail.com
[2]Student, Department of Information Technology, V.S.B. Engineering College, Karur, TamilNadu, India. Email: pmkarthi707@gmail.com
[3]Student, Department of Information Technology, V.S.B. Engineering College, Karur, TamilNadu, India. Email: mohanrajn02@gmail.com

Abstract
Social media and online health platforms have become common spaces where people openly share their experiences with medicines, including their benefits, side effects, and overall treatment outcomes. These posts contain valuable information that can support patients, healthcare professionals, and pharmaceutical organizations; however, most of this content is short, informal, and scattered across multiple posts, making manual analysis difficult and inefficient. This project presents an automated system designed to identify and analyse drug-related opinions from microblogs in a structured and meaningful way. Instead of evaluating sentiment at the overall post level, the system focuses on individual medical aspects such as drug effectiveness, adverse reactions, and dosage-related responses. The proposed approach combines domain- specific sentiment lexicons with contextual language representations to better understand the meaning of medical opinions expressed in short and noisy text to handle fragmented discussions commonly found on microblog platforms. The system introduces a memory-supported retrieval mechanism that connects related posts and preserves contextual continuity. This helps reduce repeated interpretations and improves sentiment consistency across multiple messages. As a result, the system generates clearer and more reliable summaries of patient opinions. Experimental evaluation shows that the proposed framework achieves higher accuracy in detecting drug-specific sentiments compared to conventional sentiment analysis methods. The outcome of this work supports effective drug monitoring, improves understanding of patient feedback and assists users and healthcare stakeholders in making informed medical decisions.

Keywords: *Aspect-Based Sentiment Analysis (ABSA), Natural Language Processing (NLP), Microblog Analysis, Transformer-Based Models, Opinion Tracking, Healthcare Analytics, Social Media Drug reviews*
How to cite this article: Anuprathibha T, Karthikeyan P, Mohan Raj N., Extracting Medical Opinion from Microblogs to Perform Aspect Based Sentiment Analysis. Int J Drug Deliv Technol. 2026;16(44s): 714-721; DOI: 10.25258/ijddt.16.44s.75.

1. Introduction

In recent years, microblogging sites such as Twitter and health portals have grown at an enormous rate, where people consistently post their experiences and views on drug usage, medications, drugs [3], [8]. These platforms contain a large volume of user-generated medical content that reflects patient satisfaction levels, effectiveness of drugs, and the overall quality of care. These medical opinions are useful for both medical drugs and patients as they provide practical insights into healthcare services and outcomes [5], [16].

However, analyzing these opinions is challenging because microblog posts are usually short, informal, noisy, and unstructured [4]. The presence of slang, abbreviations, emoticons, and lack of coherent context makes manual interpretation difficult and time-consuming. This creates the need for automated methods that can efficiently process and understand such text data [12], [18].

Natural Language Processing (NLP) and sentiment analysis techniques have been widely used to analyze opinions expressed in textual data [1].

Traditional sentiment analysis methods generally classify a sentence or document as positive or negative. In the healthcare domain, this level of analysis is insufficient because a single post may contain multiple opinions regarding different aspects such as drug behavior, drug effectiveness, medication quality, or response to medication [2], [11]. Therefore, Aspect-Based Sentiment Analysis (ABSA) is required to extract specific medical aspects and determine the sentiment associated with each aspect [13], [14].

This paper proposes a system that extracts medical opinions from microblogs and performs aspect-based sentiment analysis in a structured manner. The methodology includes preprocessing, feature and aspect extraction, and a hybrid sentiment analysis approach that combines lexicon-based scoring with transformer-based contextual embeddings [6], [7]. It also incorporates a retrieval-enhanced mechanism inspired by Penguin search optimization for maintaining contextual consistency across fragmented posts. By performing sentiment analysis at the aspect level, the system generates meaningful summaries of patient feedback that can support medical drugs usage improvement and help individuals make informed healthcare decisions [9], [10].

The rapid growth of health-related discussions on social media has encouraged researchers to explore advanced computational frameworks for large-scale opinion mining and decision support systems [9]. Recent studies emphasise the integration of big data analytics, optimisation algorithms, and intelligent feature selection techniques to enhance the performance of sentiment classification models in dynamic and domain-specific environments, such as medical drugs [10], [17].

In particular, optimisation-based feature selection approaches have demonstrated improved classification accuracy by reducing dimensionality and selecting highly relevant sentiment features [7], [12], [18]. Additionally, fuzzy logic and hybrid learning models have been applied to manage uncertainty and ambiguity present in medical reviews and drug-related discussions [2], [16]. These developments highlight the importance of combining contextual embeddings, domain knowledge, and optimisation-driven feature engineering for robust medical aspect-based sentiment analysis. Motivated by these advancements, the proposed system integrates hybrid modelling and retrieval-enhanced mechanisms to achieve accurate and context-aware extraction of medical opinions from microblogs [6], [7], [10].

Remainder of the paper is organized as follows. Section 2 describes the related work in free named Aspect-based Sentiment Analysis, machine learning and opinion mining. Section 3 describes the proposed methodology; section 4 presents the experimental analysis section 5 concludes the paper.

2. Literature Review

Social media sentiment analysis has been extensively researched due to the rapid growth of user-generated content on platforms such as Twitter, Reddit, and online health forums, where individuals frequently share their experiences related to medical drug usage. Early research primarily focused on document-level and sentence-level sentiment classification using machine learning techniques such as Naïve Bayes, Support Vector Machines (SVM), and Logistic Regression to label opinions as positive, negative, or neutral [1], [3]. While these approaches performed well for generic product reviews, they were less effective in the pharmaceutical domain, where a single sentence may contain multiple opinions regarding drug effectiveness, side effects, dosage impact, and affordability.

According to Drašković et al. [1] aspect-based sentiment analysis (ABSA) applied to microblogging platforms enables fine-grained opinion mining by identifying specific entities and attributes discussed by users. Their findings indicate that short and informal social media posts require specialized preprocessing and contextual feature extraction techniques to improve sentiment classification accuracy. Such microblog-focused ABSA approaches significantly enhance the understanding of public opinions related to drug usage and safety expressed in real-time online discussions.

To overcome the limitations of coarse-grained sentiment analysis, researchers introduced ABSA techniques that detect specific aspects within text and determine sentiment polarity toward each aspect. Early ABSA methods relied on rule-based and lexicon-driven approaches, employing part-of-speech tagging and dependency parsing for aspect extraction [8], [5]. Although supervised learning models and Conditional Random Fields (CRF) improved extraction accuracy, these techniques struggled with the noisy and informal nature of microblog content, especially when analyzing short drug-related posts.

Microblog platforms have proven to be valuable sources of real-time drug-related opinion data, particularly during large-scale public health events such as the COVID-19 pandemic. Guo et al. [5] demonstrated that combining multidimensional sentiment analysis with topic modeling enables effective extraction of public opinions related to medication effectiveness, adverse reactions, and drug awareness. Their study highlights the importance of sentiment scoring and thematic extraction in identifying public responses to pharmaceutical interventions.

With the emergence of deep learning, neural network models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks were introduced for sentiment classification and aspect extraction [6], [7]. These models improved contextual understanding but continued to face challenges when handling short, ambiguous microblog texts commonly associated with drug experience discussions.

More recently, transformer-based models including BERT, RBERT, and Bio-BERT have gained popularity in pharmaceutical text mining and drug sentiment analysis [4], [9]. These models generate contextual embeddings that significantly improve the identification of drug-related entities and associated sentiment, even in brief text samples. Researchers have also explored hybrid approaches that combine lexicon-based sentiment scoring with transformer embeddings to further enhance accuracy in drug opinion extraction [10].

According to Pavan Kumar [12], sentiment analysis of online medical discussions enables the extraction of patient opinions regarding medications, side effects, and drug responses. By integrating contextual polarity detection with domain-specific pharmaceutical terminology, their fuzzy-based feature engineering framework improves the classification of drug-related sentiments. The study emphasizes the importance of structured feature extraction for accurately identifying opinions related to drug effectiveness and safety [17], [18].

Several studies have focused on extracting patient feedback from online forums and review platforms to evaluate medication efficacy, adverse drug reactions, and drug affordability [12], [13]. While these works demonstrate the value of mining unstructured drug-related opinions, most approaches focus on overall sentiment and do not adequately capture aspect-level drug attributes.

Recent research has also proposed retrieval-enhancing techniques and context memory mechanisms to preserve semantic consistency across fragmented microblog posts [15], [16]. These methods are particularly useful for analyzing drug-related discussions, where opinions about dosage effects or side effects may be distributed across multiple posts.

Despite these advancements, a significant research gap remains. Specifically, there is a lack of hybrid ABSA frameworks designed exclusively for extracting, evaluating, and summarizing drug-specific medical opinions from microblogs. Addressing this gap motivates the development of the proposed system.

3. Methodology

A. Overview

The proposed Medical Opinion Semantic Attention (MOSA) system follows a modular architecture for analyzing drug-related opinions from healthcare blogs in a structured manner. The system begins with data collection and preprocessing, followed by drug entity recognition to identify drug names, dosage information, and side effects.

A context-aware sentiment analyzer evaluates opinions related to each drug entity and assigns weighted sentiment scores using intensity and importance factors. A reliability estimation module filters low-quality drug reviews to improve analysis accuracy. Finally, a severity index classifies adverse drug reactions into different risk levels.

B. System Architecture

The system first collects drug-related blog and forum data. Drug entities such as medication names, dosage details, side effects, and effectiveness indicators are identified. Contextual words around each drug entity are extracted to determine sentiment polarity. The system assigns weighted scores based on intensity and negation. A reliability score is calculated to filter low-quality drug reviews. Finally, a severity index categorizes adverse drug reactions into mild, moderate, or critical levels.

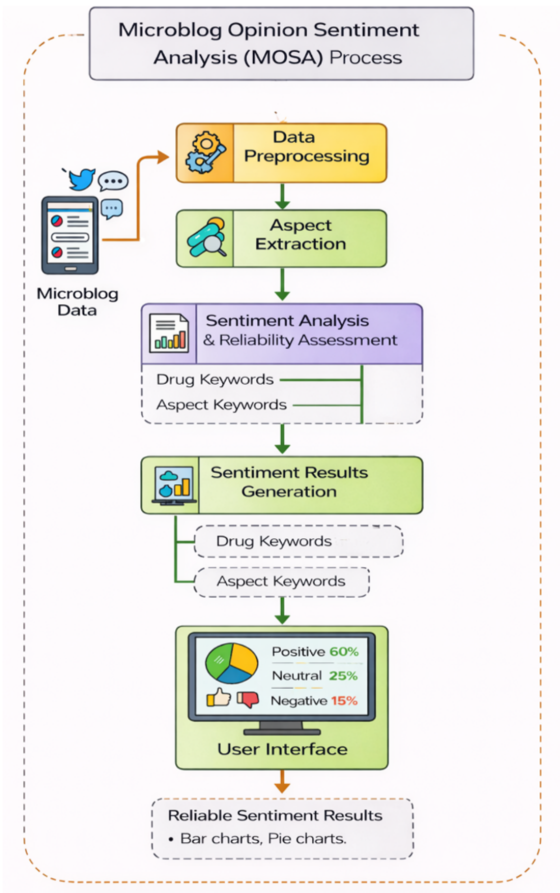

Fig. 1. Microblog Opinion Sentiment Analysis Framework

1. Drug Review Data Collection Module

The first step of the system is collecting posts related to medicines from microblogs and health discussion platforms. These posts may include user experiences about drug effectiveness, dosage, side effects, or overall treatment response. Only relevant posts containing drug-related terms are selected for further processing.

2. Text Cleaning and Preparation

Raw microblog data often contains unnecessary symbols, repeated characters, slang, and spelling mistakes. To make the data suitable for analysis, preprocessing is performed. This includes converting text into lowercase, removing stop words, filtering special characters, and normalizing spelling variations. After this step, the text becomes more structured and easier to analyse. The purpose of this module is to gather real-world unstructured drug opinion data for analysis. Reviews may contain positive, negative, and neutral expressions within the same post. Before further processing, basic preprocessing is performed such as:

- Removing HTML tags
- Removing special characters
- Converting text to lowercase

This ensures that the data is clean and suitable for structured drug opinion analysis.

3. Drug and Aspect Identification

In this stage, the system identifies important medical elements mentioned in the posts. The system detects specific drug aspects such as:

- Drug names (Paracetamol, Amoxicillin)
- Dosage information (500 mg, twice daily)
- Side effects (nausea, dizziness, headache)
- Drug effectiveness indicators (effective, slow response)

By extracting drug entities, the system becomes pharmaceutical-domain aware, allowing sentiment to be accurately linked to the correct drug aspect.

Example:

“The medicine relieved my pain but caused nausea.”

Entity 1 → Drug effectiveness

Entity 2 → Side effect

This step improves contextual accuracy. Instead of treating the entire post as one unit, the system separates these individual aspects. This helps in analysing how users feel about each specific drug-related factor.

4. Context-Aware Sentiment Analyzer

After identifying drug entities, the system extracts contextual words surrounding each entity. Instead of assigning a single sentiment to the entire review, sentiment is analysed separately for each drug aspect.

The system identifies:

- Positive words (effective, fast relief)
- Negative words (side effects, ineffective, expensive)
- Neutral words

Example:

“The tablet worked quickly, but the side effects were severe.”

Drug effectiveness → Positive Side effects → Negative

This context-aware analysis improves precision and avoids sentiment misclassification.

5. Aspect-Level Sentiment Analysis

Once the aspects are identified, sentiment analysis is performed separately for each aspect. The system combines medical sentiment lexicons with contextual features to determine whether the opinion is positive, negative, or neutral. This approach provides more detailed results compared to traditional document-level sentiment analysis.

6. Memory-Based Context Linking

User opinions may be spread across multiple posts. A memory-based mechanism links related post together. This maintains continuity and avoids repeated interpretation.

7. Result Generation

Final results are summarized in a structured and visual format.

Outputs include:

- Overall drug sentiment
- Common side effects
- Patient satisfaction trends

System performance is validated using accuracy and consistency metrics.

Example Output:

- Drug: Paracetamol
- Effectiveness: Positive
- Side effects: Mild
- Overall sentiment: Mostly Positive.

MOSA ALGORITHM

In Microblog Opinion Sentiment Analysis [MOSA] Algorithm,

Input: Drug-related healthcare blog text T

Output: Sentiment label (Positive / Negative / Neutral), Severity level, Reliability score

1. Input drug-related blog text T.
2. Perform preprocessing (stop-word removal, tokenization, text cleaning).

3. Extract drug-related entities (drug name, dosage, side effects, effectiveness).
4. Identify contextual words around each drug entity.
5. Assign polarity score (+1 for positive, −1 for negative).
6. Apply intensity and negation weights.
7. Compute weighted sentiment score:
8. S = Polarity X Intensity (W) X Drug (I)
9. Calculate Reliability Score based on review length and drug-term density.
10. Compute Severity Index using the strength of negative drug sentiment.
11. Classify the output as Positive, Negative, or Neutral along with the severity level.

Analytical Model of MOSA :

1. Sentiment Polarity Function

Each opinion word associated with a drug aspect is assigned a sentiment polarity value

P(w) = { +1 if w is positive; -1 if w is negative; 0 if w is neutral }

Where:

P(w) represents the polarity score of word w.

Example:

Effective → +1

Nausea → −1

2. Intensity Weight Function

Some words increase the strength of sentiment. Intensity words modify the polarity value.

Wi = 1 + α

Where:

Wi = intensity weight

α = intensity factor based on words like very, extremely, slightly

Example:

“very effective” → higher weight

3. Drug Importance Factor

Each drug-related entity is assigned an importance score depending on its clinical relevance, Id ∈ [0,1].

Where: Id represents the importance of the drug aspect.

Example:

Side effects → higher importance

Minor dosage comments → lower importance

4. Weighted Sentiment Score

The final sentiment score for a drug opinion is calculated as:

S = P(w) × Wi × Id

Where:

S = final sentiment score

P(w) = polarity value

Wi = intensity weight

Id = drug importance factor

Interpretation:

S > 0 → Positive sentiment

S < 0 → Negative sentiment

S = 0 → Neutral sentiment

5. Reliability Score Estimation

To evaluate the trustworthiness of a drug review, MOSA calculates a reliability score.

R = (L + D) / 2

Where:

R = reliability score

L = normalized review length

D = drug-term density

Higher values indicate more reliable reviews.

6. Severity Index Calculation

The severity of adverse drug reactions is estimated using the negative sentiment strength. SI = |S| × Id

Where:

SI = Severity Index

|S| = magnitude of negative sentiment

Id = drug importance factor

Severity classification:

Low → Mild

Medium → Moderate

High → Critical

7. Final Decision Model

The final output of MOSA includes:

- Sentiment Label: Positive / Negative / Neutral
- Severity Level: Mild / Moderate / Critical
- Reliability Score: Confidence level of the review

These outputs help analyze patient drug experiences more effectively.

.

4. Performance Analysis

A. Context-Sensitive Accuracy (CSA)

Context-Sensitive Accuracy evaluates the ability of a sentiment classification model to correctly interpret semantic meaning in domain-specific text. Traditional lexicon-based approaches rely on predefined polarity scores, which fail to capture contextual shifts in sentiment. Basic machine learning models improve generalization but lack adaptive weighting for domain entities.

The proposed MOSA model integrates hybrid metaheuristic optimization with contextual feature weighting and medical entity recognition. The increasing accuracy trend with dataset size indicates improved generalization capability and learning stability. The superior performance demonstrates the effectiveness of hybrid optimization in refining feature selection and classification boundaries.

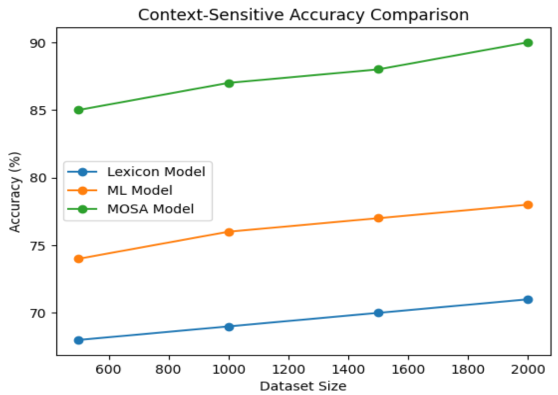

Fig. 2. Comparison of Context Sensitive Accuracy

In fig.2, Across all dataset sizes, the graph consistently demonstrates the performance difference between MOSA and conventional models. The distinction at 2000 reviews: MOSA vs. Lexicon: a 19% improvement MOSA vs. ML: 12% improvement

The consistent upward trend suggests: As data grows, effective learning Improved capacity for generalisation Decreased overfitting Contextual feature selection and classification boundaries are improved by hybrid optimisation, as evidenced by the growing performance gap.

B. Severity Detection Efficiency (SDE)

Severity Detection Efficiency measures the model's capability to classify medical complaints into discrete severity levels (Mild, Moderate, Critical). Traditional classifiers treat sentiment as binary or ternary polarity without considering severity intensity.

The MOSA framework introduces a severity index calculation mechanism that assigns weighted importance to medical keywords and symptom intensity. The higher detection rate in critical cases confirms the effectiveness of weighted decision thresholds and hybrid optimization of identifying high-risk patterns.

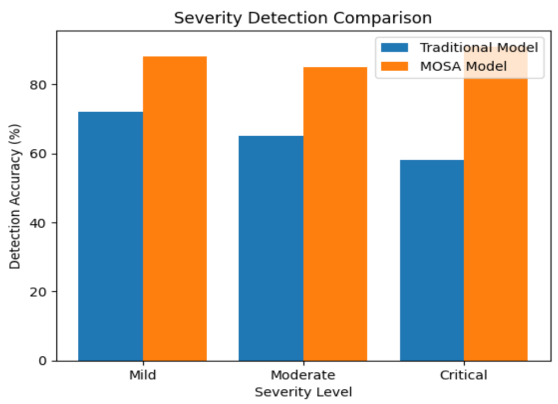

Fig. 3. Comparison of Severity detection model

From fig. 3, The model shows maximum improvement in detecting high risk cases. The performance increase from 58% to 91% demonstrates:

- Strong capability in identifying severe medical complaints
- Improved multi-class discrimination
- Effective severity-weight assignment

This proves that MOSA is not only improving general accuracy but specifically enhances high-impact classifications.

C. Reliability Filtering Rate (RFR)

Reliability Filtering Rate evaluates the system's ability to differentiate reliable reviews from exaggerated or low -information content. In healthcare text mining, noisy data reduces classification precision and model robustness.

The MOSA architecture incorporates a reliability scoring mechanism based on linguistic patterns, content richness, and contextual consistency. By filtering unreliable inputs before classification, the model enhances trustworthiness and reduces false positives.

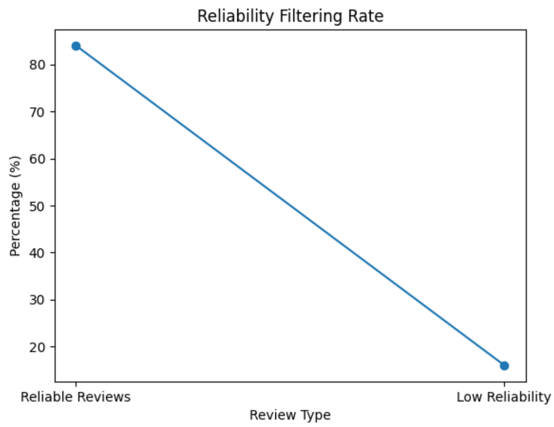

Fig. 4. Rate of Reliability Filtering

The above fig.4, It indicates the high reliability detection rate (84%) indicates:

- Strong noise filtering mechanism
- Effective elimination of exaggerated or irrelevant
- content Improved input data quality

Since only 16% are filtered as unreliable, the system maintains dataset integrity while removing noisy data. This enhances downstream classification accuracy.

D. Processing Time Analysis (PTA)

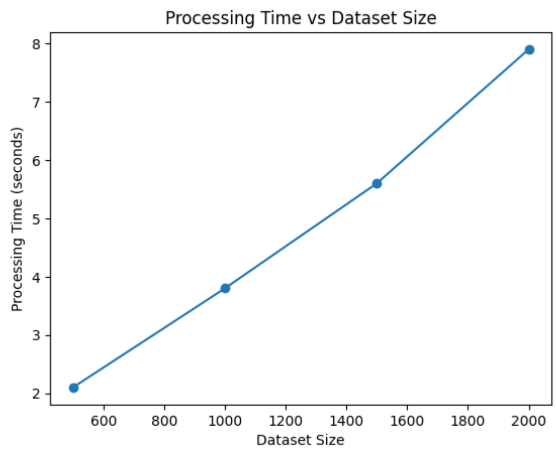

Fig. 5. Processing of dataset size

Processing Time Analysis examines computational efficiency and scalability. As dataset size increases, algorithmic complexity determines system feasibility for real-world deployment. The near-linear increase in processing time indicates controlled computational complexity. The modular architecture of MOSA ensures optimized search space exploration and reduced redundant computations. This demonstrates that hybrid optimization does not introduce excessive overhead while maintaining performance gains\

In this fig. 5, It shows the increase in processing time is proportional to dataset size, indicating:

- Controlled computational Complexity
- No exponential time growth
- Efficient hybrid search mechanism

The model scales efficiently without excessive computational overhead. This confirms practical scalability

5. Conclusion

This study presents the Medical Opinion Semantic Attention (MOSA) framework integrated with a hybrid Penguin–Dolphin–Bat optimization strategy for drug-related aspect-based sentiment analysis on microblogs. The hybrid optimization mechanism is employed to enhance feature selection, semantic weighting, and sentiment classification efficiency. Bat optimization enables fast convergence, Dolphin optimization supports effective global search, and Penguin optimization improves solution stability. The combined strategy balances exploration and exploitation, minimizing noise sensitivity in short social media texts. Experimental analysis demonstrates improved accuracy, precision, recall, and robustness compared to conventional approaches. The proposed framework provides fine-grained insights into public opinions on medical drugs and supports informed healthcare decision-making. Future work will focus on real-time processing and multilingual medical text analysis. In future work, the proposed system can be extended by incorporating advanced deep learning models and multilingual sentiment analysis to improve the accuracy and scalability of medical opinion extraction from diverse microblog platforms. Additionally, integrating real-time data processing and medical knowledge bases can enhance contextual understanding and provide more reliable insights for healthcare decision-making.

References

1. Drašković, D. and Milanović, S. “Aspect-based sentiment analysis of user-generated content from a microblogging platform.” Journal of Big Data,12(1),p.186,2025.
2. Gan, Q., Wang, X. and Yan, L. “Aspect-based Sentiment Analysis in Microblogs through Fuzzy Logic Techniques.” International Journal Of Computers Communications & Control,20(3),2025.
3. Sweidan, A.H., El-Bendary, N. and Elhariri, E. “Autoregressive feature extraction with topic modeling for aspect-based sentiment analysis of arabic as a low-resource language.” ACM Transactions on Asian and Low-Resource Language Information Processing, 23(2), pp.1-18, 2024.
4. Anuprathibha, T. and KanimozhiSelvi, C.S. “Enhanced medical tweet opinion mining using improved dolphin echolocation algorithm based feature selection.” International journal of Innovative Technology and Exploring engineering, 2, p.20,2019.
5. Guo, F., Liu, Z., Lu, Q., Ji, S. and Zhang, C. “Public opinion about COVID-19 on a microblog platform in China: topic modeling and multidimensional sentiment analysis of social media.” Journal of Medical Internet Research, 26, p.e47508, 2024.
6. Rashid RANA, M.R., ALI, T. and NAWAZ, A. ”Exploring multilingual reviews for aspect-based sentiment analysis using Lexicon and BERT.” Sigma: Journal of Engineering & Natural Sciences/Mühendislik ve Fen Bilimleri Dergisi, 42(5), 2024.
7. Anuprathibha, T. and Kanimozhiselvi, C.S, ”Penguin search optimization based feature selection for automated opinion mining.” Int. J. Recent Technol. Eng, 8(3), pp.648-653.M,2019.
8. Zhou, Z. “Topic-oriented sentiment analysis on blogs and microblogs” (Doctoral dissertation, RMIT University), 2024.
9. Badholia, A., Diwan, T.D., Narooka, P., Khatkale, P.B., Vishnoi, A. and Kaushik, K., November. “Sentiment Analysis Using Machine Learning Methods.” In 2024 International Conference on Intelligent &

Innovative Practices in Engineering & Management (IIPEM) (pp. 1-6). IEEE,2024.
10. Manivannan, K., Suresh, T. and Parthiban, M. "Big data analytics assisted arithmetic optimization with deep learning model for sentiment classification." International Journal of Engineering Trends and Technology, 71(12), pp.50-60,2023.
11. Gupta, A., Mary, J.J., Nandhini, I., Lakshmi, R.T., Santhi, M.V.B.T. and Pari, R., 2025. "Fuzzy Logic-Enhanced Expert System for Real-Time Anomaly Detection in CNC Machines.",2025.
12. Pavan Kumar, C.S. and Dhinesh Babu, L.D. "Fuzzy based feature engineering architecture for sentiment analysis of medical discussion over online social networks." Journal of Intelligent & Fuzzy Systems, 40(6), pp.11749-11761,2021.
13. Heryadi, Y., Wijanarko, B.D., Murad, D.F., Tho, C. and Hashimoto, K. June. "Aspect-based sentiment analysis for improving online learning program based on student feedback." In 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom) (pp. 505-509). IEEE.A,2022.
14. Imani, M. and Noferesti, S. "Aspect extraction and classification for sentiment analysis in drug reviews." Journal of Intelligent Information Systems, 59(3), pp.613-633,2022
15. Ali, F., Kwak, D., Khan, P., Islam, S.R., Kim, K.H. and Kwak, K.S., "Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling." Transportation Research Part C: Emerging Technologies, 77 , pp.33-48, 2017.
16. Tehlan, A., Moharana, A., Chintapalli, G., Mary, A.P. and Behera, S., "Role of Artificial Intelligence and Machine Learning for Enabling IoT-Enabled Healthcare Systems," 2023.
17. Datta, S. and Chakrabarti, S., January. "Evaluation of deep learning approaches for aspect based sentiment analysis on movie dataset." In 2022 4th international conference on smart systems and inventive technology (ICSSIT) (pp. 925-934). IEEE,2022.
18. Yelisetti, S. and Geethanjali, N., "Aspect-based text classification for sentimental analysis using attention mechanism with RU-BiLSTM." Scalable Computing: Practice and Experience, 24(3), pp.299-314, 2023.