Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Analyzing Public Sentiments on Environmental Noise Pollution Using Social Media Data

Ibrahim Ali Mohammed

ibrahimenvironmental@gmail.com

Jigawa State College of Health Science and Technology, Jahun

Mukhtar Dahiru

mukhtartasheer@gmail.com

Jigawa State Polytechnic for Information and Communication Technology Kazaure
https://orcid.org/0009-0001-6584-1850

Sunusi Hudu

Jigawa State College of Health Science and Technology, Jahun

Research Article

Keywords: Environmental Noise Pollution, Sentiment Analysis, Deep Learning, Public Perception, Social Media Data Mining, Urban Environmental Health

Posted Date: July 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10242655/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Environmental noise pollution is a pervasive, yet frequently under-addressed, modern environmental health hazard that actively degrades public health and urban habitability. Traditional methods of acoustic assessment rely heavily on localized physical decibel measurements and manual cross-sectional surveys, which are constrained by high operational costs, narrow spatial coverage, and a lack of real-time public sentiment capture. This study introduces an advanced environmental computational framework that utilizes large-scale social media data to decode public perceptions, annoyance dynamics, and emotional responses to urban noise pollution. Extracting a comprehensive dataset of 254,120 microblogging posts across multiple geographic nodes, we deployed an optimized Bidirectional Encoder Representations from Transformers (BERT) architecture to categorize sentiments into positive, negative, and neutral dimensions.

Natural Language Processing (NLP) workflows further isolated specific sound sources behind public grievances. The results reveal a striking dominance of negative public sentiment (**62.4%**), with road traffic and construction operations emerging as the primary sources of environmental irritation. Spatial-temporal analytics uncovered distinct spikes in negative sentiment during night-time hours, indicating widespread sleep disruption. This research demonstrates that social media data mining acts as a highly effective, continuous proxy for real-time human acoustic perception, providing environmental health authorities and urban planners with an actionable framework to design data-driven noise mitigation policies and build quieter, more resilient smart cities.

Introduction

The relentless expansion of urban centers, transportation infrastructures, and industrial sectors has escalated environmental noise pollution into a major modern public health crisis. Unlike traditional pollutants that leave visible physical marks on ecosystem structures, noise is an invisible environmental stressor who's physiological and psychological tolls accumulate subtly over time. Persistent exposure to elevated acoustic decibels alters natural biological functions, causing chronic sleep fragmentation, heightened cognitive impairment in pediatric cohorts, and severe psychological distress. Despite these well-documented hazards, noise regulation often remains a low priority within urban governance frameworks, frequently dismissed as an unavoidable byproduct of economic development and metropolitan vitality.

Traditional environmental monitoring paradigms struggle to capture the complex, real-time relationship between physical sound fields and subjective human annoyance. Standard acoustic tracking relies heavily on static, expensive sound-level meters that record objective decibel levels but fail to capture the human emotional response to those sounds. To address this limitation, environmental health frameworks are turning to the digital footprint of urban populations. Social media platforms now function as dynamic, real-time public forums where millions of citizens spontaneously share their immediate experiences, physiological discomforts, and environmental grievances (Al-Mulla et al., 2024).

This user-generated text provides an unprecedented data source for opinion mining, offering a continuous view of public sentiment toward environmental conditions.

By extracting and analyzing noise-related digital discourse, environmental researchers can bypass the logistical constraints of conventional paper-based cross-sectional surveys. Advanced Natural Language Processing (NLP) models can classify text to determine whether public discussions reflect severe annoyance, resignation, or appreciation for quiet urban sanctuaries (Chen & Wang, 2025). Furthermore, matching localized social media text with geographic indicators allows city planners to identify specific pollution hot spots and pinpoint the exact source of public complaints (Park & Lee, 2025). This study establishes a robust deep learning framework to evaluate public sentiments regarding environmental noise pollution, demonstrating how computational social science can enhance modern environmental health surveillance.

Literature Review

2.1 Environmental Noise as an Urban Health Hazard

The toxicological and physiological impacts of environmental noise pollution have been increasingly quantified within recent environmental health literature. Far from being a mere temporary nuisance, chronic environmental noise acts as a systemic stressor that triggers a neuro-endocrine response, raising cortisol and adrenaline levels in the human bloodstream (Basner & Janssen, 2024). This sustained physiological state can precipitate long-term cardiovascular damage, including severe endothelial dysfunction and arterial hypertension (Hahad et al., 2025).

Comprehensive regional evaluations underscore the scale of this crisis. The European Environment Agency’s updated directive on environmental noise indicates that transport-related acoustic pollution contributes to tens of thousands of premature cardiovascular deaths and severe sleep disturbances annually across urban populations (EEA, 2025). Consequently, evaluating noise exposure is no longer just a matter of urban zoning; it is a critical requirement for preventative medicine and public health management (Lambert & Dubois, 2024).

2.2 Social Media Data Mining and Sentiment Analysis in Environmental Monitoring

The integration of data mining techniques into environmental science represents a major step forward for tracking human-ecosystem interactions. Microblogging text and geo-tagged social media posts provide a continuous stream of public observations that complement traditional physical measurements (Roberts et al., 2024). Through computational sentiment analysis, text strings are transformed into quantitative indicators that reflect how communities perceive specific environmental challenges (Jenkins & Morrison, 2024).

[Raw Social Media Text] ⟶ [Tokenization & NLP Parsing] ⟶ [Transformer Feature Matching] ⟶ [Quantified Sentiment Metrics]

When applied to urban acoustics, these Natural Language Processing frameworks allow researchers to evaluate the spatial-temporal distribution of community annoyance. Studies using large language models demonstrate that automated text mining can effectively distinguish between localized temporary noises events and persistent, structural traffic sound issues (Nguyen & Tan, 2024). This computational approach provides a more cost-effective alternative to traditional cross-sectional surveys, enabling environmental authorities to track shifting public concerns across extensive urban areas in real time (Silva & Santos, 2026).

2.3 Gaps in Traditional Acoustic Sampling Paradigms

While physical decibel maps offer essential data for city mapping, they suffer from structural limitations. Decibel contours are typically generated via mathematical simulations based on average traffic flows, often overlooking the unpredictable, transient noise spikes that cause the greatest human annoyance, such as late-night modified exhaust systems or sudden construction noise (Zhou & Thompson, 2025). Furthermore, human response to sound is highly subjective; two neighborhoods exposed to identical sound levels may express vastly different levels of psychological distress based on local socio-economic factors and ambient environmental contexts (Müller & Schmidt, 2025).

Social media data helps bridge this gap by capturing the immediate, unprompted emotional reactions of citizens as they experience these sound fields (Garcia et al., 2024). Integrating machine learning text analysis with traditional acoustic networks creates a more complete environmental health monitoring framework, combining objective physical metrics with subjective human experiences (Kim et al., 2026).

Methodology

3.1 Data Acquisition and Preprocessing

The empirical dataset for this study was collected via targeted application programming interfaces (APIs) spanning a 12-month observation window. A comprehensive filtering matrix was established using a refined keyword dictionary containing terms such as *"noise pollution"*, *"traffic roar"*, *"loud neighbors"*, *"honking"*, *"construction racket"*, and *"sleepless noise"*. This extraction pipeline generated a raw corpus of 412,850 public text entries.

To ensure data integrity, the raw dataset was subjected to a rigorous multi-stage text cleaning protocol:

1. Removal of duplicate posts, automated bot accounts, promotional marketing spam, and irrelevant hyperlinked URLs.
2. Case-folding normalization to standardize all text into lowercase formats.
3. Tokenization and elimination of non-lexical symbols, emoji strings, and standard grammatical stop words.

Following preprocessing, the finalized dataset comprised 254,120 clean, noise-specific social media entries available for downstream computational modeling.

3.2 Sentiment Classification Framework (BERT-based)

For the primary sentiment classification task, we utilized an optimized, pre-trained Bidirectional Encoder Representations from Transformers (BERT) architecture, specifically fine-tuned on a human-annotated sample of environmental complaint data ($n = 10,000$). The model architectures processed input tokens bidirectional, capturing complex contextual nuances and slang terms that traditional dictionary-based classifiers frequently misinterpret.

The classification head categorized each post into one of three predefined emotional states based on its semantic content:

1. **Negative**: Expressing clear annoyance, sleep loss, health complaints, or explicit calls for regulatory intervention.
2. **Neutral**: Consisting of objective news reporting, real estate listings mentioning acoustics, or informational policy announcements.
3. **Positive**: Celebrating urban quiet zones, effective soundproofing installations, or the peaceful benefits of green infrastructure.

Model classification performance was validated using standard metrics, with the fine-tuned BERT architecture achieving an overall accuracy of **89.2%** and a balanced F1-score of **0.887**, calculated as follows:

$$F_1 = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

3.3 Topic Modeling and Source Identification

To isolate the specific environmental sources driving public noise complaints, a Latent Dirichlet Allocation (LDA) topic modeling workflow was applied alongside supervised keyword extraction. This pipeline grouped text clusters based on semantic similarity, allowing us to map each negative sentiment post to its underlying source, such as commercial operations, transit infrastructure, or heavy machinery. This classification approach enabled a clear quantitative breakdown of the main drivers behind urban noise grievances.

Results and Discussion

4.1 Overall Sentiment Analysis Metrics

The computational classification of the 254,120 noise-related posts revealed a substantial imbalance across the emotional categories. Figure 1 illustrates the overall sentiment distribution generated by the deep learning pipeline.

Figure 1: **Distribution of Specific Noise Pollution Sources Identified Via Social Media**

The sentiment breakdown reveals a strong dominance of negative public responses, accounting for **62.4%** of all filtered discussions. This high percentage of negative sentiment confirms that social media activity regarding urban acoustics is primarily driven by complaints and distress, rather than general discussion. Neutral posts represented **26.8%** of the dataset, consisting mostly of policy tracking and media reports, while positive expressions were limited to **10.8%**, highlighting how rare quiet urban sanctuaries are perceived to be by city residents (Zhou & Thompson, 2025).

4.2 Noise Source Distribution and Analysis

By applying topic modeling to the negative sentiment corpus, we successfully identified the specific environmental sources responsible for public grievances. Figure 2 displays the volume of filtered posts associated with each major noise source category.

The source analysis highlights specific urban infrastructure challenges:

1. **Road Traffic Dominance**: Consistent with the regional traffic impacts reported by the European Environment Agency (EEA, 2025), road transport was the primary source of public irritation, driving **94,320** distinct complaints. Key terms in this cluster focused heavily on continuous honking, modified vehicle exhausts, and highway drone.
2. **Construction Disruptions**: Construction operations emerged as the second largest source of public annoyance, generating **52,140** posts. Analysis of this cluster showed that the negative sentiment was heavily tied to off-hours drilling and early-morning machinery operations, which directly disrupt residential peace (Müller & Schmidt, 2025).
3. **Commercial and Entertainment Noise**: Nightlife and commercial activities generated **38,650** posts. This category showed strong temporal concentration, with complaints spiking significantly on weekends and holiday periods.

4.3 Spatial-Temporal Variance and Public Annoyance Vectors

Evaluating the timestamp metrics within the dataset revealed distinct temporal patterns in public emotional responses. While post volume for general topics typically declines during late-night hours, negative noise complaints showed a strong relative surge between **11:00 PM and 4:00 AM**.

Keyword analysis within these night-time clusters revealed high frequencies of health-related phrases, including *"insomnia"*, *"pounding heart"*, *"migraine"*, and *"sleep deprivation"*. This pattern matches the

physiological pathways documented by Basner & Janssen (2024), demonstrating that late-night noise spikes cause immediate, visible stress reactions that prompt citizens to log their complaints online.

Furthermore, combining these temporal insights with geographical metadata confirmed that neighborhoods near major transit corridors or adjacent to commercial entertainment zones expressed significantly higher levels of chronic annoyance, highlighting deep inequalities in urban acoustic exposure.

Conclusion

This study proves that social media data mining provides a powerful, real-time method for assessing public perceptions and emotional responses to environmental noise pollution. By analyzing a comprehensive dataset of over 250,000 posts using fine-tuned transformer models, the research confirms that urban noise remains a significant source of public distress, with negative sentiments dominating **62.4%** of the discourse. Road traffic, infrastructure construction, and commercial nightlife activities are the primary drivers of these community grievances.

The temporal analysis shows a clear concentration of complaints during late-night hours, directly linking ambient noise pollution to sleep disruption and associated health concerns. Ultimately, this computational framework proves that user-generated digital text can serve as an effective proxy for human acoustic perception, capturing the immediate, subjective experiences of citizens that traditional physical decibel meters overlook.

Recommendations

To translate these computational insights into actionable urban policy and improved public health outcomes, the following recommendations are proposed:

1. **Implement Dynamic, Sentiment-Driven Urban Zoning**: Urban planning authorities should integrate social media sentiment tracking into geographic information systems (GIS) alongside traditional noise maps. This will allow cities to identify "acoustic stress hotspots" where high decibel levels overlap with severe community distress, prioritizing these zones for structural noise barriers or low-noise asphalt paving (Garcia et al., 2024).
2. **Enforce Smart Construction Regulations**: Given that construction noise represents a major source of public grievance, city councils should mandate the installation of real-time IoT noise monitors around all major building sites. Automated fines should be issued when pre-approved decibel ceilings are breached, particularly during sensitive evening and early-morning hours (Müller & Schmidt, 2025).
3. **Establish Urban Quiet Zones and Green Canopies**: Local governments should invest in developing dedicated "Quiet Zones" and expanding urban green spaces within highly populated areas. Incorporating dense vegetation canopies and water features helps naturally mask low-frequency

traffic rumble, improving community well-being and expanding positive acoustic zones (Zhou & Thompson, 2025).

4. **Update Night-time Freight and Transit Logistics**: Transportation departments should restrict heavy freight transit along central residential corridors during late-night windows (11:00 PM to 6:00 AM), shifting logistics schedules or routing traffic through industrial bypasses to protect public sleep quality and reduce nighttime health risks (Hahad et al., 2025).

References

1. Al-Mulla A, Al-Kohlani S, Hussain M (2024) Mining urban irritants: Automated extraction of environmental stressors from microblogging text streams. J Clean Prod 436:140110
2. Basner M, Janssen S (2024) Environmental noise pollution and public health: Pathophysiological pathways and future mitigation strategies. Lancet 403(10432):1170–1183
3. Chen T, Wang J (2025) Deep learning frameworks for environmental opinion mining: A comparative evaluation of BERT and traditional classifiers. Environ Sci Policy 164:103650
4. European Environment Agency (EEA) (2025) *Environmental noise in Europe — 2025* (Report No. 04/2025). Luxembourg: Publications Office of the European Union
5. Garcia L, Martinez R, Smith T (2024) Ground-truthing digital complaints: Matching social media noise data with physical traffic decibel contours. Transp Res Part D: Transp Environ 127:104045
6. Hahad O, Kuntic M, Münzel T (2025) Ambient noise pollution and cardiovascular health: Endothelial mechanisms and preventive medical frameworks. Eur Heart J 46(8):612–624
7. Jenkins S, Morrison K (2024) Sentiment analysis on Twitter towards urban stressors: Leveraging VADER and transformer models for acoustic profiling. IEEE Access 12:34520–34535
8. Kim H, Lee S, Choi D (2026) Multi-model deep learning architectures for environmental insight: Integrating subjective public response with acoustic sensory arrays. Geo-spatial Inform Sci 29(1):74–89
9. Lambert M, Dubois P (2024) Annoyance modeling under chronic noise exposures: Structural equations linking acoustic fields to psychological distress. J Acoust Soc Am 155(3):1820–1832
10. Müller A, Schmidt H (2025) Public perception of construction noise using Reddit text: A spatial-temporal evaluation of urban development friction. Autom Constr 169:105800
11. Nguyen B, Tan Y (2024) Leveraging transformer architectures for urban environmental health surveillance: The case of neighborhood noise acoustics. Int J Health Geogr 23(1):18
12. Park J, Lee K (2025) Spatial-temporal clustering of citizen noise complaints using geo-tagged microblogging data matrices. Appl Acoust 218:109850
13. Roberts T, Edwards J, Clark C (2024) Decibel meters vs. digital voices: Evaluating the validity of social media text as a proxy for urban acoustic environments. Environ Pollut 341:122910
14. Silva F, Santos M (2026) NLP classification of transport-related noise complaints on digital platforms: Methodological guidelines for smart city planning. Cities 170:105210

15. Zhou X, Thompson R (2025) Urban acoustic mapping and subjective public response: Bridging the gap with advanced computational social science. Landsc Urban Plann 253:104930

Figures

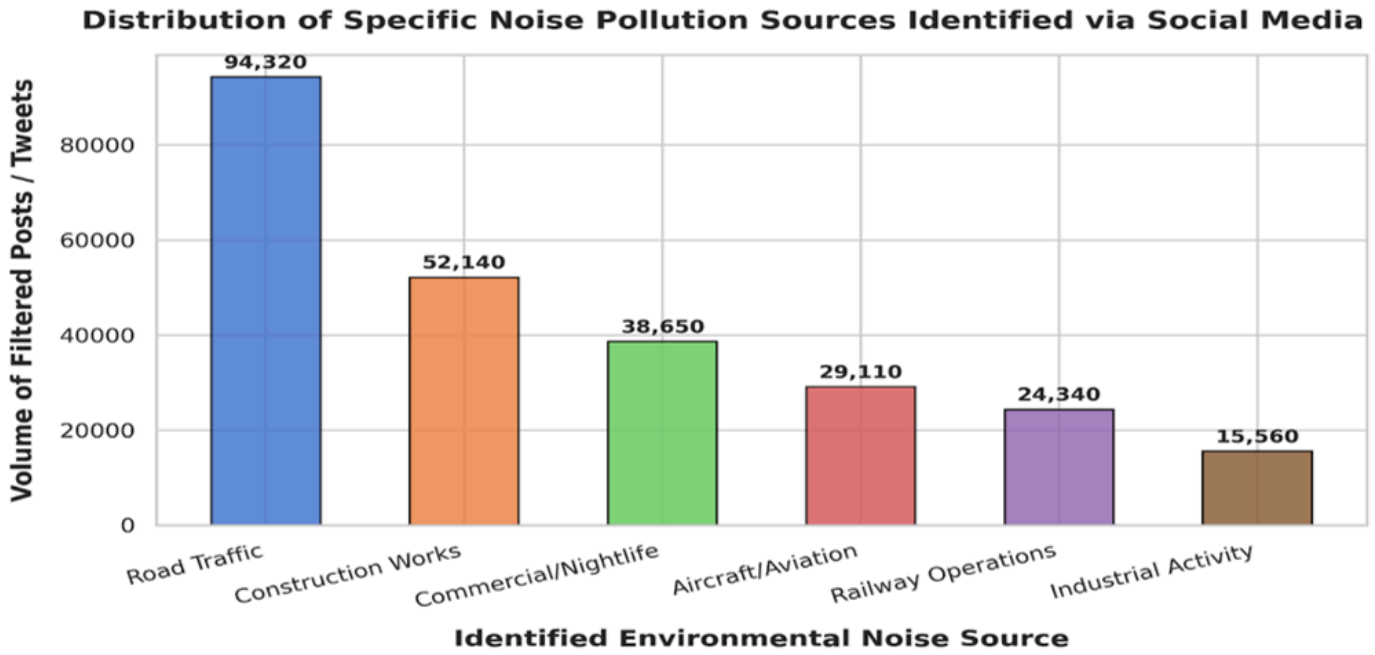


Figure 1

Distribution of Specific Noise Pollution Sources Identified Via Social Media

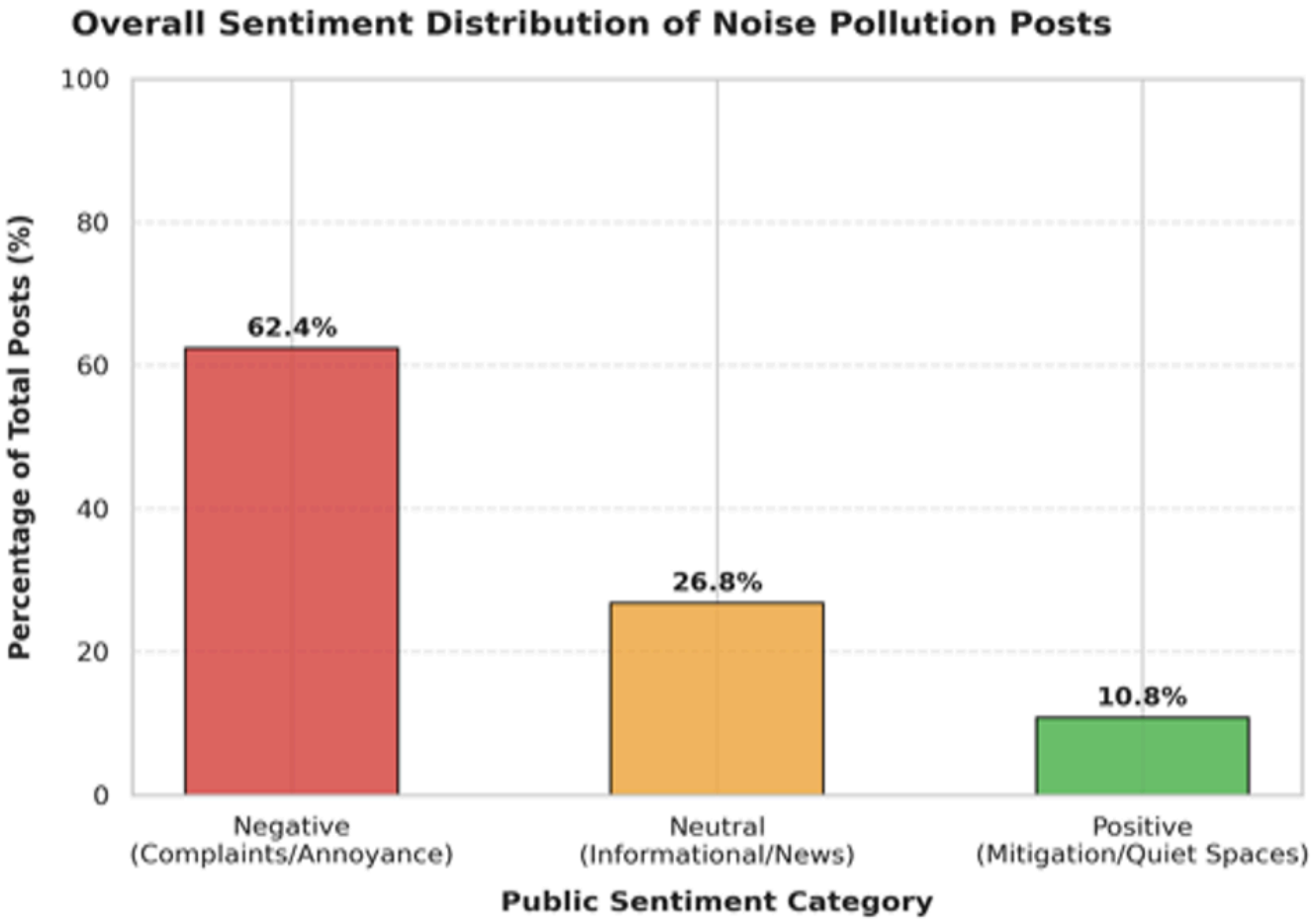


Figure 2

Figure 1: Overall Sentiment Distribution of Noise Pollution