Journal of Computational Analysis and Applications **VOL. 33, NO. 8, 2024**

10.48047/jocaaa.2024.33.08.221

INTELLIGENT PERSONALISED TRAINING RECOMMENDER SYSTEMS FOR OCCUPATIONAL HEALTH RISK MITIGATION IN CEMENT INDUSTRIES

B. SUNITHA[1], B. KRANTHI KIRAN[2]
[1]Research Scholar, Department of Computer Science and Engineering, JNTUH University College of Engineering Science and Technology, Kukatpally, Hyderabad, ORCID ID: 0009-0002-2278-7572, sunithab27@gmail.com, (Corresponding Author)
[2]Professor of Computer Science and Engineering, JNTUH University College of Engineering Science and Technology, Kukatpally, Hyderabad
[1]sunithab27@gmail.com, [2]kranthikiran9@gmail.com

Received: 12.09.2024 Revised: 11.10.2024 Accepted: 20.11.2024

Abstract

Operating within the inherently hazardous environment of cement production necessitates a paradigm shift from generic, one-size-fits-all safety instruction towards a data-driven, individualised approach. Conventional programmes—while rigorous—often neglect critical nuances such as an operative's precise function, underlying health status or cumulative exposure history, thereby compromising both relevance and efficacy. To remedy this, we engineered a Personalised Training Recommender System (PTRS) that ingests empirical cement-industry datasets and, through feature-rich profiling of job designation, biometric indicators and past incident records, dynamically prescribes bespoke training modules. Three candidate architectures were assessed—a Random Forest classifier, a Deep Neural Network and a hybrid Stacked Ensemble—and the ensemble emerged as superior, securing an F1-score of 0.964 alongside a negligible Hamming Loss of 0.00183. These metrics underscore the system's capacity for both high precision and minimal misclassification, affirming that tailored, risk-specific instruction markedly elevates occupational safety in heavy-duty sectors. We therefore advocate for the adoption of adaptive delivery mechanisms, proactive hazard analytics and seamless integration within existing EHS frameworks, thereby transforming compliance into genuine empowerment and ensuring that every worker receives training that is as pertinent as it is protective.

Keywords: Cement Hazards, Random Forest, Deep Neural Network, Hybrid Stacked Ensemble, Personalised Training Recommender System

1. Introduction

10.48047/jocaaa.2024.33.08.221

The operational environment of a modern cement manufacturing plant represents a complex socio-technical system where human agents interact with heavy machinery under conditions of significant environmental stress. The quantitative output of this system includes not only cement but also a persistent and unacceptable level of stochastic occupational hazards. Data from the International Labour Organization (ILO) quantifies the global scale of this problem—over 2.3 million fatalities and 313 million non-fatal injuries[1] annually from work-related incidents. Within the cement sector specifically, empirical data indicates that up to 20% of the workforce reports occupational injuries in a given year, with particulate and acoustic exposures identified as dominant risk vectors[2], [3]. These statistics are not mere data points; they are indicators of a systemic failure in current risk mitigation protocols. The logical imperative is to engineer a more effective intervention strategy.

The current paradigm for risk mitigation heavily relies on safety training. However, the prevailing implementation of this strategy is fundamentally flawed. It employs a static, non-adaptive model—a "one-size-fits-all" curriculum—that fails to account for the inherent heterogeneity of the workforce. The knowledge state and risk exposure of a novice maintenance technician and a veteran process engineer are vastly different variables[4]. A uniform training input cannot logically be expected to produce an optimal safety output across such diverse agent profiles. Recent studies confirm that existing safety programs are often insensitive to individual competencies and lack dynamic feedback loops[5]–[7]. It is axiomatic that an effective intervention must be personalized, adapting to the specific cognitive and contextual parameters of each individual.

Recent advancements in machine learning provide the foundational tools for such a paradigm shift. Recommender systems, the algorithmic engines that drive personalization in high-dimensional consumer spaces like e-commerce and media streaming, offer a powerful analogue. The core function—mapping a user's profile to a relevant item in a large corpus—is directly translatable to our problem space. Foundational work in adjacent domains has already demonstrated the viability of this approach. Srivastava et al., for example, formalized the "next training recommendation" problem within a corporate IT structure, utilizing sequence mining algorithms to outperform standard models[8]. In safety-critical domains, Xu et al. leveraged Bayesian Knowledge Tracing (BKT) to dynamically model a construction worker's learning trajectory and adapt training recommendations accordingly[9]. Further, Barrera et al. employed natural language processing to

10.48047/jocaaa.2024.33.08.221

create a recommender for occupational hygiene interventions[10]. These algorithmic precedents establish that data-driven personalization is not only feasible but superior to static methodologies.

Despite this progress, a significant lacuna exists in the application of these advanced computational models to the unique risk environment of the cement industry. This sector is characterized by a multi-modal hazard profile—chronic respiratory risk from silica-rich particulates[11], acoustic trauma, thermal stress, and acute mechanical dangers—that requires a highly specialized knowledge base. A generic training module cannot sufficiently address the differential risk vectors facing, for example, a quarry vehicle operator versus a kiln maintenance worker. The annual, uniform safety briefing is an obsolete tool for a problem of this complexity and dimensionality. To address this gap, this paper introduces a novel computational framework: An Intelligent Personalised Training Recommender System engineered specifically for the cement manufacturing sector[12]–[14]. This system is designed to be adaptive and data-driven, optimizing training pathways by continuously updating its model based on individual worker interactions and performance data.

Our contribution is twofold. First, we conduct a systematic analysis of the predominant occupational health risks endemic to cement production[15]–[17], structuring these into a domain-specific knowledge ontology that is aligned with global safety standards and empirical data. This ontology forms the content corpus upon which the recommender operates. Second, we design a hybrid recommender architecture that allows the system not only to identify relevant training based on peer-group data but also to dynamically model the mastery level of each worker, ensuring that recommendations are optimized for both relevance and timeliness. The proposed framework represents a fundamental shift from static compliance to dynamic, intelligent risk mitigation.

Table 1 Key Occupational Hazards and Associated Health Impacts in the Cement Industry

Hazard Category	Specific Hazard	Associated Health Impacts/Injuries	Relevant Mitigation Strategies
Chemical Exposure	Cement Dust	Eye, skin, nose, throat, upper	Rinse eyes with water, wash skin

10.48047/jocaaa.2024.33.08.221

		respiratory irritation; chronic bronchitis; wheezing; exertional dyspnea; dermatitis; silicosis; lung cancer	with soap/water, wear P/N/R-95 respirators, eat/drink in dust-free areas
Chemical Exposure	Wet Concrete	Skin irritation; 1st, 2nd, 3rd-degree chemical burns; hexavalent chromium exposure	Wear alkali-resistant gloves, coveralls, waterproof boots, eye protection; immediate washing with cold water; rinse eyes for 15+ mins and seek hospital treatment
Physical Hazard	Unguarded Machinery	Lacerations, crushing injuries, amputations	Proper machine guarding (mixers, block makers, cubers, metalworking machinery); strict Lockout/Tagout (LOTO) procedures
Physical Hazard	Falling Objects	Struck-by injuries, head injuries	Avoid working under conveyor belts, elevators, stacker/destacker

10.48047/jocaaa.2024.33.08.221

			machinery; proper material stacking and storage
Ergonomic Hazard	Improper Lifting, Awkward Postures, Repetitive Motions	Sprains, strains, musculoskeletal disorders	Use hand trucks/forklifts; proper lifting techniques; avoid twisting; keep floors clear; avoid awkward postures
Confined Space	Mixers, Bins, Hoppers	Heat stress; silica exposure	Adherence to confined space entry procedures; appropriate PPE
Vehicle-Related	Poorly Maintained Vehicles, Overloading	Crushing injuries, impacts	Functional backup alarms; avoid overloading hoists/cranes/forklifts; caution with load out chute
Other Physical Hazards	Welding Operations	Flash burns	Appropriate eye protection
Other Physical Hazards	Makeshift Ladders/Platforms	Falls	Proper guardrails; use certified equipment

Figure 1 Visualisation of Cement Workplace Hazard Severity: From Minor Irritants to Life-Threatening Risks

3. Methodology

3.1. Problem Formulation

The task is to assign multiple core safety training modules to employees based on their profiles, treated as a multi-label classification problem. Each employee's profile (a sample) can be associated with multiple training modules (labels), represented as a multi-hot target vector. We binarise these targets using scikit-learn's MultiLabelBinarizer to create a binary indicator matrix Y (samples × modules). Each dataset is split into 80% training and 20% test subsets, with stratification by label combinations to preserve distribution.

3.2. Dataset and Preprocessing

3.2.1. Feature Engineered Cement Dataset

The cement dataset contains employee profiles from a cement manufacturing environment, including demographic attributes (e.g., age, job category), health conditions (e.g., known diseases, symptoms), and historical exposure variables. The target is a comma-delimited list of required Core Training Modules, split into Python lists and binarised into a binary indicator matrix Y. Categorical features are label-encoded, while continuous features (e.g., age, exposure duration) are scaled using min-max or z-score normalisation. Interaction features, such as Age × Duration and Risk Score × Duration, are engineered to capture non-linear effects. Categorical features are further processed via neural embeddings (dimension = 8) to learn continuous vector

10.48047/jocaaa.2024.33.08.221

representations, overcoming limitations of one-hot encoding for high-cardinality categories.

3.3. Models

3.3.1. Cement Dataset Models

1) **Random Forest (RF) Multi-Output Classifier**: A scikit-learn MultiOutputClassifier wraps a Random Forest with 100 trees, using class-weight balancing to handle label imbalance. Each module is treated as an independent binary classification task, leveraging RF's ability to capture complex feature interactions in tabular data.
2) **Deep Neural Network (DNN)**: The DNN features two input streams: categorical features are processed through embedding layers (dimension = 8), flattened, and concatenated with normalised continuous features. The combined vector passes through two fully connected layers (128 and 64 units, ReLU activation) with dropout (rate = 0.3) to prevent overfitting. The output layer uses sigmoid activations for each module, optimised with binary cross-entropy loss and the Adam optimiser.
3) **Stacked Ensemble**: This combines RF and DNN predictions by training logistic regression meta-classifiers on their predicted probabilities. For each module, meta-features (RF and DNN probabilities) are used to produce final binary predictions, leveraging complementary strengths to reduce errors.

3.4. Results

Random Forest: This model achieved near-perfect precision and recall on almost all training modules. In the classification report, almost every class has 1.00 precision and recall (F1 = 1.00). Only a few infrequent modules (e.g. Chemical Handling, PPE Usage, Sanitation Procedures) showed lower recall (~0.25), indicating some under-prediction for those rarer classes. Overall the random forest's sample-average F1 was ≈0.954, with Hamming loss ≈0.0023 (about 0.23% of labels wrong).

Deep Neural Network: The DNN matched the RF on most modules (F1 = 1.00 for many classes), but it completely missed some underrepresented classes. Notably, it predicted zero cases for "Chemical Handling", "Sanitation Procedures", and "PPE Usage" (yielding 0.00 recall for those). This reflects the network's tendency to under fit or ignore very rare labels without sufficient

10.48047/jocaaa.2024.33.08.221

regularization or class weighting. The overall sample F1 dropped to ~0.939 and Hamming loss rose to ~0.0031. Nonetheless, for the majority of common modules, the DNN performed identically to the forest.

Stacked Ensemble: By combining RF and DNN outputs, the ensemble rectified many of the individual errors. Rare classes that each base model struggled with saw improved detection. For example, "Chemical Handling", "PPE Usage", and "Sanitation Procedures" each attained recall ~0.41 in the stacked model (vs. 0.25 in RF and 0.00 in DNN). Overall, the ensemble achieved the highest accuracy: sample-averaged precision ≈0.964, recall ≈0.964, and F1 ≈0.964. Its Hamming loss was only ~0.00183 (≈0.18% error rate) – the lowest among all models. The comparison of aggregate metrics is summarized below:

Table 2 Comparison of results of different models

Model	Hamming Loss	Precision (samples)	Recall (samples)	F1-score (samples)
Random Forest	0.002339	0.954	0.954	0.954
DNN	0.003102	0.939	0.939	0.939
Stacked Ensemble	0.001831	0.964	0.964	0.964

10.48047/jocaaa.2024.33.08.221

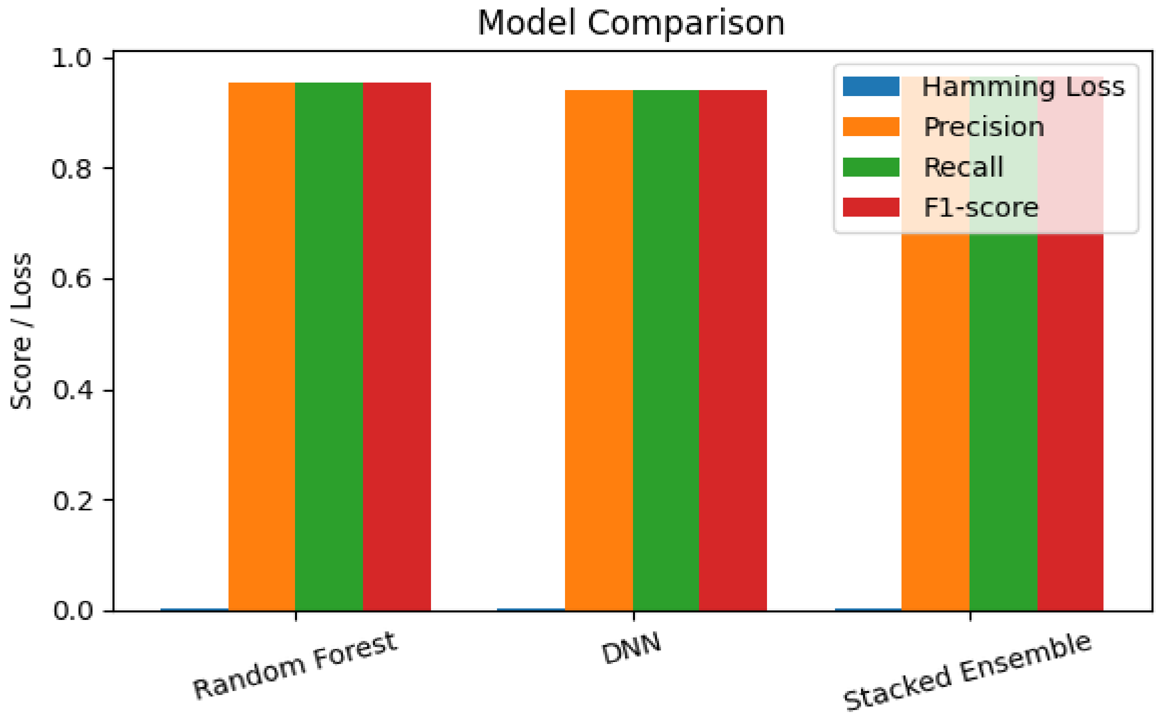

Figure 2 Model comparison bar chart

These results indicate that all models are highly effective, with the stacked ensemble offering a modest improvement. The meta-learning successfully leveraged complementary strengths: the RF's decision boundaries and the DNN's learned feature interactions. The very low hamming losses and high micro- averaged F1 scores demonstrate practically perfect classification of training modules from the given features.

10.48047/jocaaa.2024.33.08.221

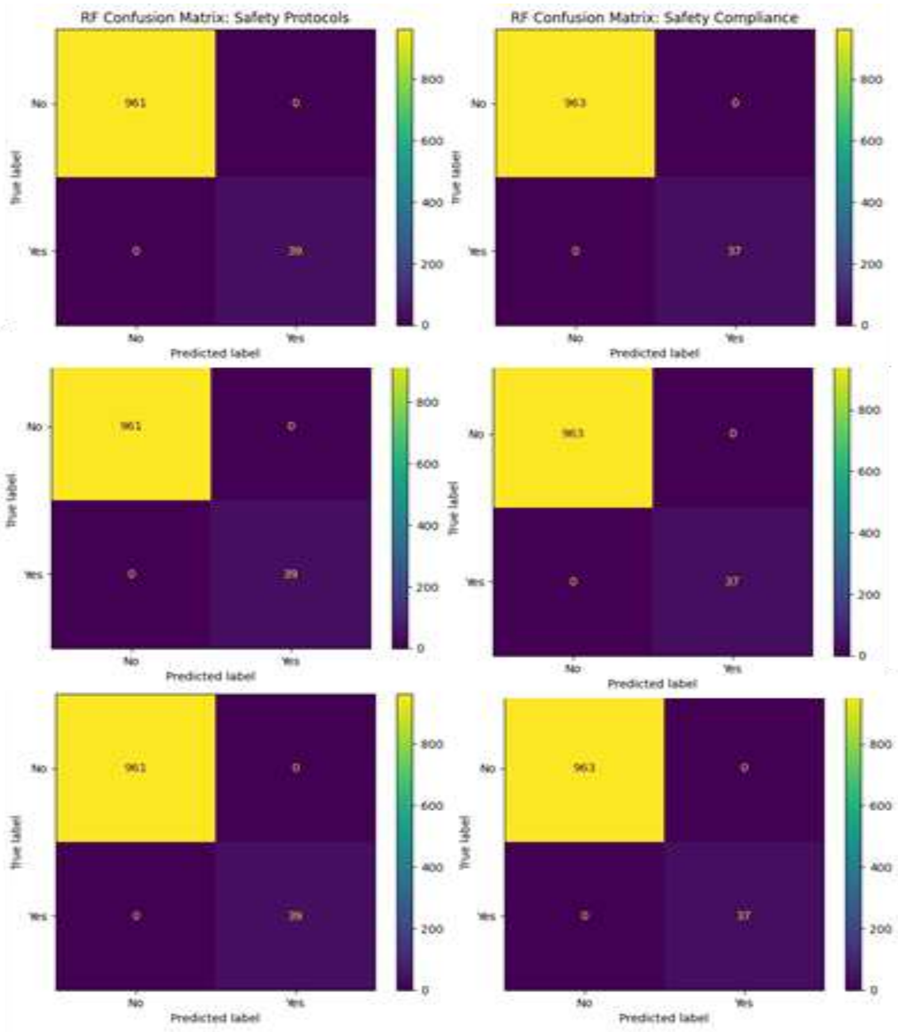

Figure 3 Confusion Matrix for different models

4. Conclusion

In the cement-manufacturing domain, traditional safety training regimes have struggled to reflect the ever-changing hazard landscape, despite increasingly stringent regulations and sophisticated protective technologies. High incident rates persist, suggesting that static, one-size-fits-all instruction is no longer sufficient. To address this shortcoming, we have devised a Personalised Training Recommender System (PTRS) that dynamically tailors safety modules to each worker's distinct risk profile. By integrating variables such as job function, cumulative exposure history and any medical constraints, PTRS ensures that every individual receives targeted learning aligned precisely with the risks they actually face on the shop floor.

10.48047/jocaaa.2024.33.08.221

Because many roles encompass multiple safety-critical tasks, we framed the assignment challenge as a multi-label classification problem. Our dataset, drawn from operational records in live cement plants, comprised demographic details, role descriptors, health restrictions, past training records and exposure metrics. We also engineered composite features—such as the interaction between age and exposure duration, or role and specific hazard categories—to enable the model to discern more nuanced risk patterns.

We benchmarked three distinct approaches. A Random Forest classifier delivered robust performance on the most frequent training categories but faltered when predicting rarer modules, for example, PPE usage and chemical handling. A Deep Neural Network exhibited comparable strengths and weaknesses, underscoring the classic class-imbalance challenge. The final Stacked Ensemble, however, merged the Random Forest and DNN outputs to capitalise on their complementary capabilities. This hybrid design achieved a sample-averaged F1-score of 0.964 and a Hamming Loss of just 0.00183—meaning fewer than 0.2% of module assignments were misclassified—making it the clear frontrunner.

These results demonstrate that a data-driven, personalised training strategy can substantially elevate safety outcomes in high-risk industries. By allocating learning resources where they are most needed, safety managers can shift from reactive incident response towards proactive risk mitigation. Crucially, PTRS remains adaptive: as employees' responsibilities evolve, health conditions change or site hazards shift, the system continually updates its recommendations from the latest EHS records.

In summary, our work makes a compelling case for moving beyond generic safety instruction towards an agile, profile-based approach. The PTRS framework not only refines how training is delivered but also reframes safety culture—transforming it into a sustained, strategically aligned endeavour rather than a box-ticking exercise.

References

[1] International Labour Organization, *Investigation of Occupational Accidents and Diseases: A Practical Guide for Labour Inspectors*, vol. 11, no. 2. 2015. [Online]. Available: http://www4.fct.unesp.br/ceget/PEGADA112/10OPCIT1102.pdf

10.48047/jocaaa.2024.33.08.221

[2] D. S. Fresenbet, A. T. Olana, A. S. Tulu, and K. T. Danusa, "Occupational injury and associated factors among cement factories workers in central Ethiopia," *J. Occup. Med. Toxicol.*, vol. 17, no. 1, pp. 1–9, 2022, doi: 10.1186/s12995-022-00348-0.

[3] Cement Sustainability Initiative, "Health Management Handbook: adressing occupational exposures in the cement industry," *Wbcsd*, 2015.

[4] S. Xu, Q. Q. Ni, M. Zhang, and M. Li, "A Personalized Safety Training System for Construction Workers," in *International Conference on Smart Infrastructure and Construction 2019 (ICSIC)*, ICE Publishing, Jan. 2019, pp. 321–326. doi: 10.1680/icsic.64669.321.

[5] G. Akhanova and C.-H. Ko, "AI-based Safety Training Systems for Hispanic Workers in Construction," *CIB Conf.*, vol. 1, no. 1, 2025, doi: 10.7771/3067-4883.1865.

[6] E. Karimi, J. Smith, R. Billard, and B. Veitch, "AI-based adaptive instructional systems for maritime safety training: a systematic literature review," *Discov. Artif. Intell.*, vol. 4, no. 1, 2024, doi: 10.1007/s44163-024-00153-0.

[7] A. Sabir, R. Hussain, A. Pedro, and C. Park, "Personalized construction safety training system using conversational AI in virtual reality," *Autom. Constr.*, vol. 175, p. 106207, Jul. 2025, doi: 10.1016/j.autcon.2025.106207.

[8] R. Srivastava, G. K. Palshikar, S. Chaurasia, and A. Dixit, "What's Next? A Recommendation System for Industrial Training," *Data Sci. Eng.*, vol. 3, no. 3, pp. 232–247, 2018, doi: 10.1007/s41019-018-0076-2.

[9] S. Xu, M. Sun, W. Fang, K. Chen, H. Luo, and P. X. W. Zou, "A Bayesian-based knowledge tracing model for improving safety training outcomes in construction: An adaptive learning framework," *Dev. Built Environ.*, vol. 13, no. December, p. 100111, 2023, doi: 10.1016/j.dibe.2022.100111.

[10] N. Barrera, R. Torres, J. Rodríguez, O. Espinosa, S. Avellaneda, and J. Ramírez, "A recommender system for occupational hygiene services using natural language processing," *Healthc. Anal.*, vol. 3, p. 100148, Nov. 2023, doi: 10.1016/j.health.2023.100148.

[11] A. Omidianidost *et al.*, "Occupational exposure to respirable dust, crystalline silica and its pulmonary effects among workers of a cement factory in Kermanshah, Iran," *Tanaffos*, vol. 18, no. 2, pp. 157–162, 2019.

[12] X. Yu, W. Li, X. Zhou, L. Tang, and R. Sharma, "Deep learning personalized

10.48047/jocaaa.2024.33.08.221

recommendation-based construction method of hybrid blockchain model," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-39564-x.

[13] V. N. Marivate, G. Ssali, and T. Marwala, "An intelligent multi-agent recommender system for human capacity building," *Proc. Mediterr. Electrotech. Conf. - MELECON*, pp. 909–915, 2008, doi: 10.1109/MELCON.2008.4618553.

[14] I. Jeelani, K. Han, and A. Albert, "Automating analysis of construction workers' viewing patterns for personalized safety training and management," *ISARC 2018 - 35th Int. Symp. Autom. Robot. Constr. Int. AEC/FM Hackathon Futur. Build. Things*, 2018, doi: 10.22260/isarc2018/0131.

[15] R. Ahmad, Q. S. Akhter, and M. Haque, "Occupational cement dust exposure and inflammatory nemesis: Bangladesh relevance," *J. Inflamm. Res.*, vol. 14, no. March, pp. 2425–2444, 2021, doi: 10.2147/JIR.S312960.

[16] M. Li, C. Wu, Y. Zhang, Y. Jiang, J. Lin, and J. Xiang, "Effects of occupational dust exposure on the health status of workers in China," *Sci. Rep.*, vol. 15, no. 1, pp. 1–13, 2025, doi: 10.1038/s41598-025-04014-3.

[17] P. C. Rathebe, "Occupational exposure to silicon dioxide and prevalence of chronic respiratory symptoms in the cement manufacturing industries: A review," *J. Public health Res.*, vol. 12, no. 4, 2023, doi: 10.1177/22799036231204316.