2025 Journal of Engineering, Technology and Social Sciences
Jurnal Kejuruteraan, Teknologi dan Sains Sosial
Volume 11 Special Issue: ICETISM
International Conference on Emerging Technologies, Information Science and Mathematics
e-ISSN: 27166848

ENHANCING DUST MANAGEMENT IN MINING OPERATIONS: EXPLORING INNOVATIONS IN PREDICTIVE MODELLING AND MACHINE LEARNING

Mumini Babatunde Saka[1,2] and Mohd Hazizan Mohd Hashim[1*]

1 School of Materials and Mineral Resources Engineering, Universiti Sains Malaysia, Nibong Tebal, Pulau Pinang 14300, Malaysia,

2 Department of Minerals and Petroleum Resources Engineering, Institute of Technology, Kwara State Polytechnic, Ilorin, Nigeria.

[*] mohd_hazizan@usm.my

ARTICLE INFO

Article history:
Received
14 July 2025
Received in revised form
18 Sept 2025
Accepted
3 Oct 2025
Published online
15 Oct 2025

Keywords*:*
Dust Monitoring;
Machine Learning;
Predictive Model

ABSTRACT

The ecological issue posed by particulate matter (PM) in surface mines is substantial. Precise quantification of particulate matter (PM) concentration is necessary to formulate efficacious approaches for mitigating and regulating PM. This study aims to present a viable strategy for the proactive control of dust concentration by leveraging predictive modelling and machine learning techniques. The study used the Thermo Scientific Air Quality Dust Monitor (ADR1200) to assess real-time data collected from the FYS granite quarry in Malaysia. The acquired datasets were analyzed using three machine learning algorithms: decision trees, random forests, and linear regression. The algorithms accurately predicted and anticipated the optimal dust concentration in the quarry. The analysis yields a linear regression model that exhibits highly favourable outcomes. Precisely, the minimal mean squared error (MSE) is calculated to be 17.6, the coefficient of determination (R-squared) is determined to be statistically significant at 0.96, and the mean absolute error (MAE) and root mean squared error (RMSE) are relatively low, measuring at 4.195 and 1.81, respectively. With a 96% account for the variability, the model demonstrates a robust capacity to elucidate the dust concentration data and exhibits exceptional predictive accuracy. This strategy aligns with the Sustainable Development Goals (SDGs 3, 9, and 13), which aim to improve health and foster innovation in the industry.

1. Introduction

Effective dust control is a vital component of mining operations, as suspended dust particles provide significant health hazards to miners and nearby people and adversely affect the environment (Monteiro et al., 2019; Wambwa et al., 2023). Traditional methods for regulating dust emissions often fail to effectively predict and manage dust, leading to concerns about health and safety and non-compliance with regulations (Zafra-Pérez et al., 2023). Nevertheless, recent progress in predictive modelling and machine learning presents encouraging methods to tackle these difficulties. Nonetheless, the efficacy of these methods is occasionally limited, and they may not sufficiently address the intricate dynamics of dust generation and dispersion in mining settings.

As opined by Gholami et al. (2021) and Rahmati et al. (2020), Mining businesses can enhance the implementation of preventive measures, optimize their operations, and mitigate the adverse effects of dust on the environment and worker health by taking a proactive approach to dust management. Ensuring efficient dust control is crucial in quarry operations as it directly affects the health and safety of workers, adherence to environmental regulations, and interactions with the surrounding population (Baah-Ennumh et al., 2021; Mvile & Bishoge, 2024). Due to the frequent occurrence of dust emissions at quarry sites, it is crucial to implement suitable management methods to mitigate the associated hazards(Saka & Hashim, 2024). Hamdan et al., (2023) argue that previous methods mainly rely on manual surveillance techniques and reactive measures, which often fail to handle the complex, compelling, and constantly evolving challenges dust presents in these environments.

In recent years, the mining industry has seen a significant transformation in its approach to dust control (Agboola et al., 2020; Kondrasheva et al., 2021; Moran et al., 2014). This move revolutionizes dust management operations by incorporating state-of-the-art predictive modelling, machine learning techniques, and modern sensor technology (Bahadur et al., 2023; Yuan et al., 2020). These technologies enable mining businesses to analyze complex data sets on dust emissions, including weather conditions, geological properties, equipment operation parameters, and historical dust readings. Predictive models can utilize methods such as neural networks, support vector machines, and decision trees to accurately anticipate rates of dust formation, identify regions with a high risk of dust, and optimize tactics for controlling dust. The Thermo Scientific Air Quality Dust Monitor ADR1200 is a key tool in these developments. This state-of-the-art device allows for the continuous and instantaneous monitoring of particulate matter levels in the air near quarry sites.

The device is a crucial instrument for dust control initiatives due to its precision, resilience, and ability to collect data across a broad spectrum of particle sizes. The combination of machine learning and predictive modelling technologies with the ADR1200 has significantly transformed the perspectives of quarry operators regarding dust management. By harnessing the extensive data gathered by the ADR1200, along with meteorological measurements and operational characteristics, it is possible to create prediction models that can accurately foresee dust-level changes. The ability to anticipate enables quarry operators to engage in proactive decision-making, allowing them to promptly and specifically address dust-related concerns before they worsen.

Sarker, (2021) states that machine learning algorithms enhance the effectiveness of predictive models by continuously evaluating large amounts of data to identify complex connections and patterns. By employing iterative learning methodologies, these algorithms can progressively improve prediction models, adjust to dynamic environmental conditions, and optimize strategies for managing dust. The outcome is a highly responsive and adaptable system that effectively tackles the present and future concerns regarding dust emissions in quarry settings.

Using sophisticated sensor technologies, predictive modelling, and machine learning not only enhances operational efficiency and adherence to regulatory requirements but also demonstrates a commitment to environmental conservation and community welfare. Proactive management of dust emissions by quarry owners can contribute to conserving adjacent ecosystems, mitigating the health impacts on workers, and fostering healthy relationships with the local population.

Agboola et al., (2020) argue that efficient dust control techniques can enhance quarry operation durability and financial viability, ensuring sustained prosperity within a progressively competitive industrial landscape. This study investigated the impact of predictive modelling and machine learning on transforming dust management methods within mining operations. Integrating machine learning and predictive modelling with the Thermo Scientific Air Quality Dust Monitor ADR1200 represents an innovative strategy for managing dust in quarry operations. By utilizing this sophisticated technology, quarry operators can address the complex challenges of dust emissions, resulting in enhanced accuracy, productivity, and environmental accountability. Operators can use this to establish new benchmarks for the industry's most efficient techniques and contribute to a more promising future in mining operations.

2. Materials and Methods

The dataset used in this study was acquired from the FYS Marketing SDN BHD quarry, a representative example of a substantial surface mining operation located in Bukit Mertajam, Penang, Malaysia (100^{0}28′53.30″E, 5^{0}22′44.46″N). The study area is illustrated in Figure 1. This study provides an overview of advancements in predictive modelling and machine learning methodologies for efficiently managing dust in mining operations.

Figure 1. Aerial view of the study area (Google Earth, 2023).

Dust concentration data for (TSP, PM10, PM2.5) and meteorological conditions, such as temperature, humidity, wind speed, wind direction, rain, and air pressure, were collected through on-site monitoring every minute from January 31st to February 7th 2024, using the Thermo Scientific Air Quality Dust Monitor ADR1200 as depicted in Plate 1. The collected datasets were first refined through feature engineering, where environmental and dust concentration data were transformed into structured inputs that better capture the factors influencing PM2.5 emissions in the quarry. These processed datasets were then subjected to three different machine learning algorithms (MLAs): the decision tree model (DTM), linear regression model (LRM), and random forest model (RFM), to accurately predict and forecast PM2.5 concentrations, following the adopted process shown in Figure 2.

Plate 1. Thermo Scientific air quality dust monitor (ADR 1200)

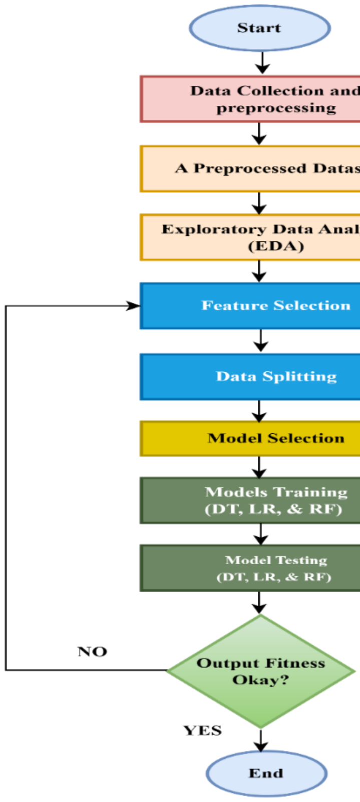

Figure 2. Schematic flowchart of the used algorithms

3. Results

Figure 3 presents the time series plot of particulate matter concentrations (PM10, PM2.5, and TSP) measured during the study period. This figure reflects the raw monitoring dataset rather than the model outputs and is included to illustrate the temporal variation and interrelationship among the particulate fractions in the quarry environment. Specifically, it highlights the persistence of PM2.5 concentrations relative to PM10 and TSP, which provides the empirical basis for subjecting the dataset to machine learning analysis.

The discussion of the machine learning models was subjected to a hypothesis:

- Null hypothesis (H_0): There is no discernible difference between the PM2.5 test and predicted values, respectively, by the models.
- Alternative hypothesis (H_1): The models predict a distinction between the PM2.5 test and predicted values.

Thus, Figure 3 serves as the baseline representation of the dataset, while the subsequent sections of the manuscript present the model training, prediction, and hypothesis testing outcomes.

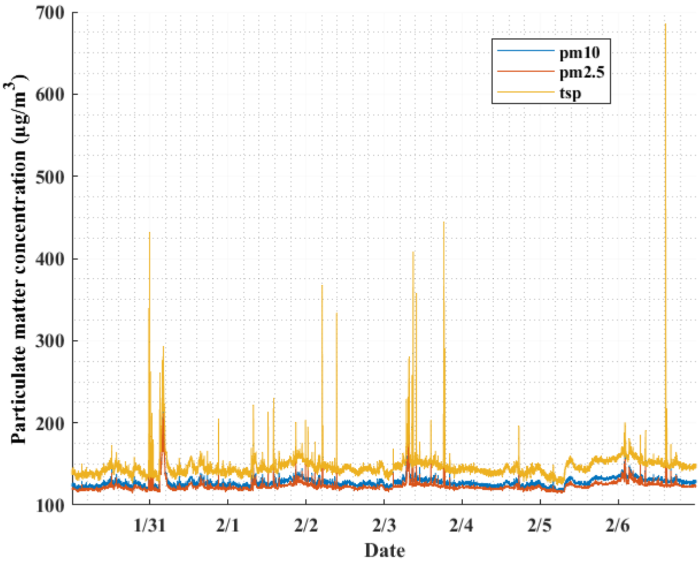

Figure 3. Time series plot

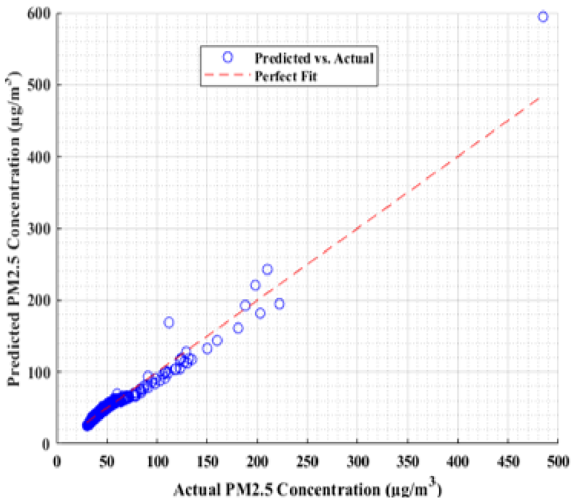

Fig. 4. Linear regression model performance.

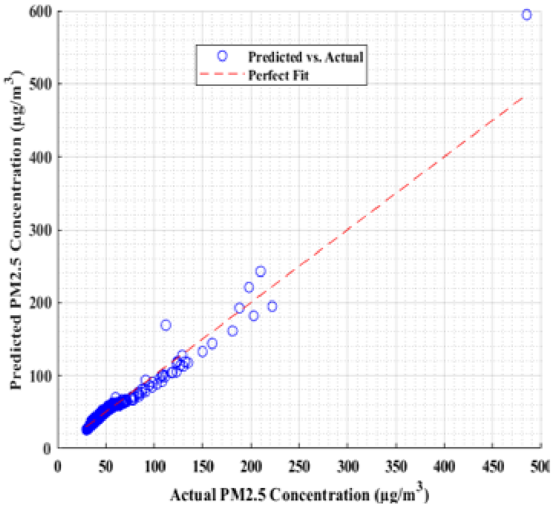

Fig. 5. Decision trees model performance

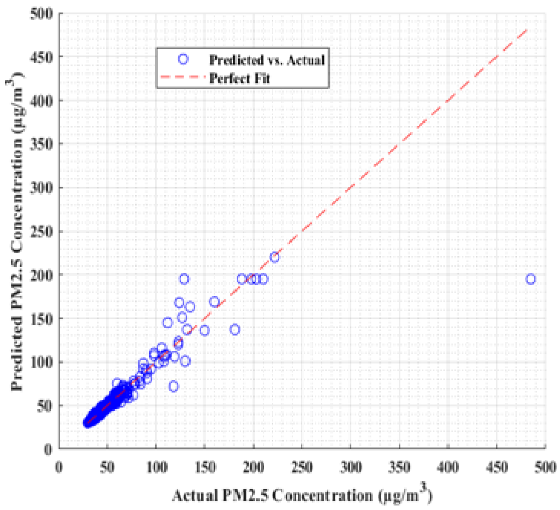

Fig. 6. Random forest model performance.

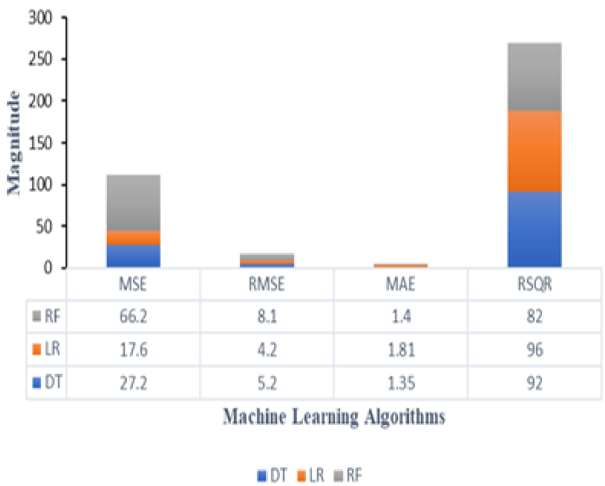

Fig. 7. Assessment criterion of the machine learning models

2025 Journal of Engineering, Technology and Social Sciences
Jurnal Kejuruteraan, Teknologi dan Sains Sosial
Volume 11 Special Issue: ICETISM
International Conference on Emerging Technologies, Information Science and Mathematics
e-ISSN: 27166848

4. Discussion

Table 1. Interpretation of LR, DT, and RF.

Machine Learning Model	Statistical Analysis	Results Interpretations
Linear Regression	Parametric	The study's findings underscore the importance of exploring diverse methodologies and algorithms in predictive modelling for dust reduction in mining operations. There are noticeable differences in the expected amounts of dust between the PM2.5 experimental dataset and the projected cohorts. The PM2.5 test value generally produces lower outcomes, indicating a more prudent approach, whereas the PM2.5 forecast value tends to provide higher values, implying a more sanguine prediction. By analyzing and comparing various techniques, one can acquire valuable insights to improve the effectiveness of dust management measures in mining operations. Furthermore, the Levene test results analysis reveals no statistically significant difference in variance across the datasets under investigation. The obtained p-value of 0.382, which exceeds the commonly accepted significance level of 0.05, provides evidence to reject the null hypothesis, implying significant differences among the observed values. Hence, it is unlikely that any disparities detected among the results can be attributed to fluctuations in variance, thereby enhancing the reliability of the subsequent statistical examinations or data interpretations. It highlights the ongoing need to improve predictive modelling methodologies to adequately address the complexities associated with dust control in the mining industry.
	Non-parametric	In the non-parametric analysis conducted, it is evident that there is a slight difference between the PM2.5 test values and the predicted PM2.5 values of the dependent variable. The median values of the PM2.5 test data and the predicted PM2.5 value show a slight increase, with medians of 44 and 43.94, respectively. Suggests a tendency for somewhat higher measurements in the PM2.5 experimental dataset, while the difference is minimal. The findings are further supported by subsequent analysis employing the Mann-Whitney U-test, which reveals no statistically significant difference between the PM2.5 test and PM2.5 forecasted values of the dependent variable. The obtained U-test statistic of 1258627, along with a p-value of 0.449, provides inadequate evidence to reject the null hypothesis. Implies that no significant difference was observed in the distributions of scores among the datasets. Therefore, it may be asserted with confidence that there is no statistically significant disparity between the PM2.5 test and PM2.5 projected values to the dependent variable, thereby accepting the null hypothesis without any opposition. Additionally, the linear regression analysis reveals that the projected PM2.5 levels effectively explain the fluctuations observed in the PM2.5 test outcomes, as indicated by an R-squared coefficient of determination of 0.96. Furthermore, the analysis of variance (ANOVA) yields strong

		evidence against the null hypothesis, as indicated by a statistically significant F-statistic of 37141.18 and a p-value of less than 0.001. This research suggests that the anticipated PM2.5 levels substantially impact the variability of PM2.5 test outcomes, reinforcing the conclusions on the association between these two variables.
Decision Trees	Parametric	According to the descriptive statistics, the average PM2.5 test values slightly exceed the projected PM2.5 values for the dependent variable. The PM2.5 test exhibits an average (M) of 46.7 and a standard deviation (SD) of 18.52. In contrast, the PM2.5 projected values demonstrate an average of 46.61 and a standard deviation of 16.82. While the average results of both values are comparable, the PM2.5 test demonstrates a somewhat higher mean and a slightly more significant standard deviation, indicating a relatively higher degree of variability. Additionally, Levene’s test findings suggest that there is no statistically significant variation in variance among the evaluated data. The p-value obtained, which is 0.923, surpasses the commonly accepted significance limit of 0.05. Consequently, the null hypothesis is accepted. This observation implies that the groups exhibit equal variances, suggesting that any observed differences in readings are unlikely to be attributed to variations in variance. As a result, the Levene test reveals that the samples exhibit the same degree of variance, with similar amounts of variability within each outcome.
	Non-parametric	Based on the descriptive statistics, the PM2.5 test values slightly exceed the projected PM2.5 values for the dependent variable, on average. The median (Mdn) values for both variables are 44 and 43.5, respectively. It indicates little difference in the average results, although the PM2.5 examination shows slightly higher levels. Nevertheless, the PM2.5 test reveals marginally elevated values for the dependent variable, with the anticipated PM2.5 levels, albeit the disparity is relatively small. Additional examination employing the Mann-Whitney U test indicates no statistically significant difference between the obtained PM2.5 test results and the expected PM2.5 values of the dependent variable. Insufficient evidence exists to reject the null hypothesis based on the U-statistic of 1270322.5 and a p-value of 0.757, which exceeds the significance level of 0.05. It suggests that there is no significant difference between the distributions of scores in the two sets of values. Furthermore, the regression model provides evidence that the anticipated PM2.5 contributes significantly to the variability observed in the PM2.5 test outcomes, as indicated by an R-squared coefficient of determination of 0.92. The analysis of variance (ANOVA), as shown in Table 13, provides strong support for the alternative hypothesis. The F-statistic of 19267.54 and the p-value of less than 001 demonstrate a significant relationship between PM2.5 and the PM2.5 test, indicating a statistically significant anticipated impact.
Random Forest	Parametric	The descriptive statistics reveal that, on average, the PM2.5 test values marginally exceed the PM2.5 concentration predicted values for the dependent variable. Specifically, the PM2.5 test has a mean (M) of 46.7

	and a standard deviation (SD) of 18.52. At the same time, the PM2.5 predicted values have a mean of 46.61 and a standard deviation of 15.56 for the dependent variable. Although the mean values are comparable, the PM2.5 test exhibits a slightly higher average and standard deviation, indicating a somewhat greater level of variability. Moreover, Levene's test indicates no statistically significant variation in variance among the analyzed data, with a p-value of .58 exceeding the conventional significance level of 0.05. Thus, the null hypothesis, asserting equal variances among the datasets, is supported. The results indicate no statistically significant difference in variation across the sets, suggesting that each value exhibits the same level of variability.
Non-parametric	Based on the available data, it can be observed that both the PM2.5 test and PM2.5 predicted values demonstrate a median value of 44. The results indicate that the dependent variable in both datasets has an equivalent mean value. This observation highlights similar central tendencies in the distributions of the PM2.5 test and predicted PM2.5 values, indicating no disparity in the median values of the dependent variable. The observation arises from the fact that both variables exhibit the same magnitude for the dependent variable. The Mann-Whitney U test further supports the findings, which reveal no statistically significant difference between the PM2.5 test and PM2.5 forecast datasets with the dependent variable. The estimated U-statistic is 1260958.5, accompanied by a p-value of 0.504, which is above the predetermined significance limit of 0.05. Therefore, based on the available evidence, it is not possible to reject the null hypothesis, indicating that there is no significant difference in the distributions of scores between the two sets of values. Furthermore, the regression analysis demonstrates that the projected dust concentration (PM2.5) effectively explains a substantial percentage of the variance observed in the dust concentration (PM2.5) test. The strong relationship is evidenced by an R-squared (R^2) value of 0.82, which indicates a robust correlation of 81.51%. The ANOVA statistical analysis yields strong evidence about the influence of PM2.5 on the PM2.5 test, as indicated by a highly significant F-statistic of 7038.05 and a p-value of less than 0.001. Therefore, it is possible to reject the null hypothesis, indicating a significant association between the expected PM2.5 concentration and the measured PM2.5 concentration.

Table 2. Model Evaluation Metrics

Model	Set	RSME	MSE	MAE	R^2
Linear Regression	Train	4.1	17.6	1.81	0.96
Linear Regression	Train	4.1	17.62	1.81	0.96
Decision Tree	Train	5.1	27.24	1.35	0.92
Decision Tree	Train	5.2	27.26	1.35	0.92
Random Forest	Train	8.01	66.22	1.4	0.82
Random Forest	Train	8.01	66.24	1.4	0.82

5. Conclusion

It is crucial to prioritize the development of predictive modelling and machine learning approaches in the field of dust control within mining operations to safeguard the well-being of miners, maintain the safety of production operations, and limit environmental pollution. The green mining initiatives in Malaysia are primarily dependent on these aspects, which align with the Sustainable Development Goals (SDGs) on good health and well-being (SDG 3), industry, innovation, and infrastructure (SDG 9), and climate action for environmental sustainability (SDG 13). Various machine learning algorithms, including decision trees, linear regression, and random forest, exhibit distinct performance characteristics when assessed for their ability to predict and forecast dust concentration (PM2.5) in the FYS quarry environment. This study presents a linear regression model that demonstrates highly favourable results, including a minimal mean squared error (MSE) of 17.6, a statistically significant coefficient of determination (R-squared) of 0.96, and relatively modest values for root mean squared error (RMSE) and mean absolute error (MAE), precisely 4.195 and 1.81, respectively. The model has outstanding predictive accuracy and a strong capacity to elucidate PM2.5 concentration data, encompassing 96% of the spectrum. Based on the findings, it can be concluded that the linear regression model offers highly reliable performance and is well-suited for predicting PM2.5 concentrations in this context. The model demonstrates the lowest errors across all parameters and achieves the best R-squared value, indicating a highly reliable model in predicting and explaining. Linear regression models are known for their simplicity and efficient computational usage, making them well-suited for practical application in real-world situations.

Acknowledgements

The authors are grateful to the management and staff of FYS Marketing Sdn Bhd for giving us access to their quarry site, and we extend our gratitude to the Mineral Research Centre, Minerals and Geoscience Department, Malaysia

References

Agboola, O., Babatunde, D. E., Isaac Fayomi, O. S., Sadiku, E. R., Popoola, P., Moropeng, L., Yahaya, A., & Mamudu, O. A. (2020). A review on the impact of mining operation: Monitoring, assessment and management. *Results in Engineering*, *8*. https://doi.org/10.1016/j.rineng.2020.100181

Baah-Ennumh, T. Y., Yeboah, A. S., & Akularemi, A. E. J. (2021). Contextualizing the effects of stone quarrying: insights from the Wenchi municipality in Ghana. *GeoJournal*, *86*(1), 489–505. https://doi.org/10.1007/s10708-019-10080-8

Bahadur, F. T., Shah, S. R., & Nidamanuri, R. R. (2023). Applications of remote sensing vis-à-vis machine learning in air quality monitoring and modelling: a review. *Environmental Monitoring and Assessment*, *195*(12). https://doi.org/10.1007/s10661-023-12001-2

Gholami, H., Mohammadifar, A., Malakooti, H., Esmaeilpour, Y., Golzari, S., Mohammadi, F., Li, Y., Song, Y., Kaskaoutis, D. G., Fitzsimmons, K. E., & Collins, A. L. (2021). Integrated modelling for mapping spatial sources of dust in central Asia - An important dust source in the global atmospheric system. *Atmospheric Pollution Research*, *12*(9). https://doi.org/10.1016/j.apr.2021.101173

Hamdan, M. Q., Lee, H., Triantafyllopoulou, D., Borralho, R., Kose, A., Amiri, E., Mulvey, D., Yu, W., Zitouni, R., Pozza, R., Hunt, B., Bagheri, H., Foh, C. H., Heliot, F., Chen, G., Xiao, P., Wang, N., & Tafazolli, R. (2023). Recent Advances in Machine Learning for Network Automation in the O-RAN. *Sensors (Basel, Switzerland)*, *23*(21). https://doi.org/10.3390/s23218792

Kondrasheva, N. K., Kireeva, E. V., & Zyryanova, O. V. (2021). Development of new compositions for dust control in the mining and mineral transportation industry. *Journal of Mining Institute*, *248*(2), 272–280. https://doi.org/10.31897/PMI.2021.2.11

Monteiro, N. B. R., da Silva, E. A., & Moita Neto, J. M. (2019). Sustainable development goals in mining. *Journal of Cleaner Production*, *228*, 509–520. https://doi.org/10.1016/j.jclepro.2019.04.332

Moran, C. J., Lodhia, S., Kunz, N. C., & Huisingh, D. (2014). Sustainability in mining, minerals and energy: New processes, pathways and human interactions for a cautiously optimistic future. *Journal of Cleaner Production*, *84*(1), 1–15. https://doi.org/10.1016/j.jclepro.2014.09.016

Mvile, B. N., & Bishoge, O. K. (2024). Mining and sustainable development goals in Africa. *Resources Policy*, *90*. https://doi.org/10.1016/j.resourpol.2024.104710

Rahmati, O., Mohammadi, F., Ghiasi, S. S., Tiefenbacher, J., Moghaddam, D. D., Coulon, F., Nalivan, O. A., & Tien Bui, D. (2020). Identifying sources of dust aerosol using a new framework based on remote sensing and modelling. *Science of the Total Environment*, *737*. https://doi.org/10.1016/j.scitotenv.2020.139508

Saka, M. B., & Hashim, M. H. bin M. (2024). Critical assessment of the effectiveness of different dust control measures in a granite quarry. *Journal of Public Health Policy*. https://doi.org/10.1057/s41271-024-00481-6

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3). https://doi.org/10.1007/s42979-021-00592-x

Wambwa, D., Mundike, J., & Chirambo, B. (2023). Environmental Protection Fund audits in Zambia: A strategy for ensuring regulatory compliance in Zambia’s mining sector. *Extractive Industries and Society*, *16*. https://doi.org/10.1016/j.exis.2023.101353

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., & Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, *241*. https://doi.org/10.1016/j.rse.2020.111716

Zafra-Pérez, A., Boente, C., García-Díaz, M., Gómez-Galán, J. A., de la Campa, A. S., & de la Rosa, J. D. (2023). Aerial monitoring of atmospheric particulate matter produced by open-pit mining using low-cost airborne sensors. *Science of the Total Environment*, *904*. https://doi.org/10.1016/j.scitotenv.2023.166743