Journal of Mining and Environment (JME) Vol. 15, No. 3, 2024, 961-976

Shahrood University of Technology

Journal of Mining and Environment (JME)

Journal homepage: www.jme.shahroodut.ac.ir

Iranian Society of Mining Engineering (IRSME)

Analysis of Concentration of Ambient Particulate Matter in the Surrounding Area of an Opencast Coal Mine using Machine Learning Techniques

Ravi Kiran Podicheti, and Ram Chandar Karra*

Department of Mining Engineering, National Institute of Technology Karnataka, Surathkal, Mangalore, India

Article Info

Received 17 December 2023

Received in Revised form 15 January 2023

Accepted 2 February 2024

Published online 2 February 2024

DOI: 10.22044/jme.2024.13960.2604

Keywords

Opencast coal mine

PM10

PM2.5

Dispersion

Machine learning

Abstract

Opencast coal mines play a crucial role in meeting the energy demands of a country. However, the operations will result in deterioration of ambient air quality, particularly due to particulate emissions. The dispersion of particulate matter will vary based on the mining parameters and local meteorological conditions. There is a need to establish a suitable model for predicting the concentration of particulate matter on a regional basis. Though a number of dispersion models exist for prediction of dust concentration due to opencast mining, machine learning offers several advantages over traditional modeling techniques in terms of data driven insights, non-linearity, flexibility, handling complex interactions, anomaly detection, etc. An attempt has been made to assess the dispersion of particulate matter using machine learning techniques by considering the mining and meteorological parameters. Historical data comprising of mine working parameters, meteorological conditions, and particulate matter pertaining to one of the operating opencast coal mines in southern India has been utilized for the study. The data has been analyzed using different machine learning techniques like bagging, random forest, and decision tree. The performance metrics of test data are compared for different models in order to find the best fit model among the three techniques. It is found that for PM10, many of the times bagging technique gave a better accuracy, and for PM2.5, decision tree technique gave a better accuracy. Integration of mine working parameters with meteorological conditions and historical data of particulate matter in developing the model using machine learning techniques has helped in making more accurate predictions.

1. Introduction

Opencast coal mines are contributing to major portion of coal supply to thermal power plants in India, and are expected to persist for few more decades until there is a significant transition towards renewable sources of energy. The main air pollutants associated with opencast coal mining are particulate matter and gaseous emissions. The particulate matter includes PM_{10} & $PM_{2.5}$, and gaseous emissions consist of CO, CO_2, SO_2, and NO_x. Fine particulate emissions resulting from opencast mining will have significant impact on the health of the nearby habitation. Hence, there is a need to assess the dust concentration due to opencast mining, and take pro-active mitigation measures to safeguard the well-being of surrounding environment through implementation of sustainable mining practices.

Many studies have been conducted on dispersion of particulate matter based on mathematical models but very few have evaluated with testing at mine sites [3, 29]. A number of models were developed such as Box model, Gaussian model, Eulerian model, and Lagrangian model, which were reported to be applied for air quality prediction in the mining industry [26]. However, many of the studies have been focusing on a single activity. For example, the particulate dispersion due to blasting, using computational

Corresponding author: krc_karra@yahoo.com (R. Ch. Karra)

fluid dynamics model was studied [29]. The paths of the mineral particles were modeled using Lagrangian particle tracking. The study concluded that 30-60% emissions will be retained in the quarry, but the dispersion of dust concentrations to different directions was not estimated. The techniques used for isolated buildings and wind tunnels are suitable for analyzing quarry area or building recirculation [7]. Similarly, studies were carried out for prediction of particulate matter due to stock piles in opencast mines [2].

The dispersion equations developed within the pit boundary provide a reasonable accurate estimate of PM_{10} dispersion within the near field of the deep OC coal mines [29, 31]. Some studies were conducted to estimate the particulate matter level inside the mine and understanding its dispersion as they travel from source to the surface. An empirical relationship between the particulate matter concentration and depth of the mine was also proposed [10].

However, the factors being implemented are required to adapt to the local conditions of each mine [18]. The same empirical formulae for determination of emission rates cannot be considered for all Indian mines [5]. Development of region-specific dispersion techniques with regard to changing mine parameters and local meteorological conditions for PM_{10} and $PM_{2.5}$ emanating from opencast coal mines is necessary to allow accurate prediction of PM dispersion to make decisions [8, 27].

Some studies were conducted on the effect of meteorological parameters on fine and coarse particulate matter. It was concluded that the meteorological parameters have a major effect on the monitored $PM_{2.5}$ and PM_{10} [19]. However, the changes in mining parameters have not been taken into account for their impact on particulate matter concentration.

Instead of manually driven rules and build models from analyzing large amount of data, machine learning offers a more efficient alternative for complex data analysis. Studies were carried out on improved machine learning approach for optimizing dust estimation in open-pit mines [20]. Studies were conducted based on the meteorological data and dust concentration at a single point. The study proposed a machine learning model to estimate dust concentrations in opencast coal mines. However, the dust concentration at different locations in and around the mine was not taken into account.

Machine Learning (ML) has evolved as a sub-field of artificial intelligence that involves the development of self-learning algorithms in order to make more accurate predictions. It helps in making prediction based on complex interaction between various factors, allowing mines to proactively manage and mitigate particulate emissions while optimizing their operations.

Machine learning can handle a wide variety of data types such as numerical, categorical, and time-series data. In the context of particulate emissions, one can incorporate data like production rates, equipment specifications, weather conditions, and historical emissions records. ML models can seamlessly integrate and analyze this diverse data, capturing complex relationships. Mining operations are dynamic, which changes in equipment, processes and environmental conditions. Machine learning models can adapt to these shifts. They can learn from new data and evolving trends, ensuring that their predictions remain relevant and effective over time.

Beyond predictions, ML models can offer recommendations for optimization and control. For instance, if the model anticipates high emissions due to adverse weather conditions, it can suggest adjustments to operations like reducing production during that period or deploying dust control measures. This proactive approach helps in emissions reduction and cost savings. In essence, machine learning enhances air quality modeling by leveraging the power of data and algorithms to make more accurate and timely predictions, helping to manage and mitigate the effects of air pollution on public health and the environment. It is a valuable tool to improve air quality, and reduce the impact of pollution on the communities, thereby, accomplishing the objective of green mining practices.

Out of the various machine learning techniques available for different data sets, three techniques viz., bagging, random forest, and decision tree have been used for analyzing the data for arriving at a suitable prediction model for application in the opencast mines.

This study, therefore, incorporates the mining parameters, meteorological data, and estimating the particulate matter concentration at different locations using different machine learning techniques, and also suggesting better technique when compared to each other. Further, the performance metrics of test data are compared for different models in order to select the best model among the three techniques.

The study also included the identification of significant parameters responsible for dust dispersion, ranked in order of preference for each specific location. Recursive Feature Elimination (RFE) technique is used to know the rank order of the independent variables in terms of importance in order to aid the mine management to take mitigation measures accordingly.

2. Methodology

One of the opencast coal mines located in South India has been considered for evaluating the impact of coal mining parameters and meteorological conditions on ambient particulate matter dispersion in the core and buffer zones of the project. Core zone is the project area or the area lying within the project boundary. Buffer zone is the area lying within 10 km radius from the project. It is crucial to monitor the dispersion of particulate matter in the core zone to determine the concentration of dust exposure of working personnel, enabling the implementation of necessary measures in the event of concentrations exceeding the prescribed limit. Similarly, evaluating particulate matter within the buffer zone is equally important for implementing preventive measures aimed at reducing pollution and averting potential health threats to the general public. The mine is in operation since 1993 with a production capacity of 4Mtpa in an area of 900 Ha. Mining operations in the opencast mine involve drilling and blasting, while loading and transportation of overburden and coal are carried out by shovel-dumper combination. A view of the opencast mine is shown in Figure 1.

Figure 1. View of the opencast mine under study.

The study involves field investigations, encompassing the collection of data related to mine configuration, meteorological parameters, and particulate matter concentrations in both the core and buffer zones of the studied area. Further, the data is analyzed using various machine learning techniques.

2.1. Field investigations

Field investigations involved collection of mine working parameters, and monitoring of meteorological data and particulate matter for the last two years and compilation of previous 8 years data. A total of 10 years data, spanning from 2012 to 2022, has been taken into account for the analysis. Processing of the data include mine working plans for determining the effective area contributing to pollution, quarry depth, overburden dump height, distance of material transportation and production. Figure 2 illustrates a representation of the working plan from, which the data has been processed.

2.1.1. Mine working parameters

The mine working parameters include effective area contributing for pollution, quarry depth, overburden dump height, distance of transportation of material from extraction point to dumping point, and coal extracted and overburden removed.

i. Effective area contributing to pollution

Total project area of an opencast coal mine may not be the source of pollution. Certain areas may be highly contributing to pollution, while others may have a moderate impact and some may not contribute significantly or not at all. Accordingly, the total area of the project is classified into different types like quarry area, dumping area, infrastructure area consisting of office buildings, coal handling plant, workshop, roads, etc., and other areas like undisturbed area, reclaimed area, etc. Among these areas, the dumping area, infrastructure area, undisturbed area, and reclaimed areas have fixed locations, with consistent distances from the monitoring points. However, in the quarry area, where progress occurs over time, the concentration of dust at monitoring points is minimally affected, as significant portion of the dust is contained within the quarry itself. The quarry area, dumping area, and infrastructure area are the sources of air pollution due to mining activities, whereas undisturbed and reclaimed areas are not considered as sources of pollution. The undisturbed area is where mining operations are yet to start, which is treated as virgin area. The reclaimed areas are those areas where technical and biological reclamation of overburden dumps have been carried out. Technical reclamation involves leveling, sloping, stabilizing and compaction of dumps, and spreading of top soil, whereas biological reclamation involves planting of saplings on the spread top soil. As such, the net effective area contributing to pollution has been arrived at by considering quarry area, dump area & infrastructure area, whereas undisturbed & reclaimed areas are exempted. Figure 3 presents the variation in different areas from the year 2012 to 2021.

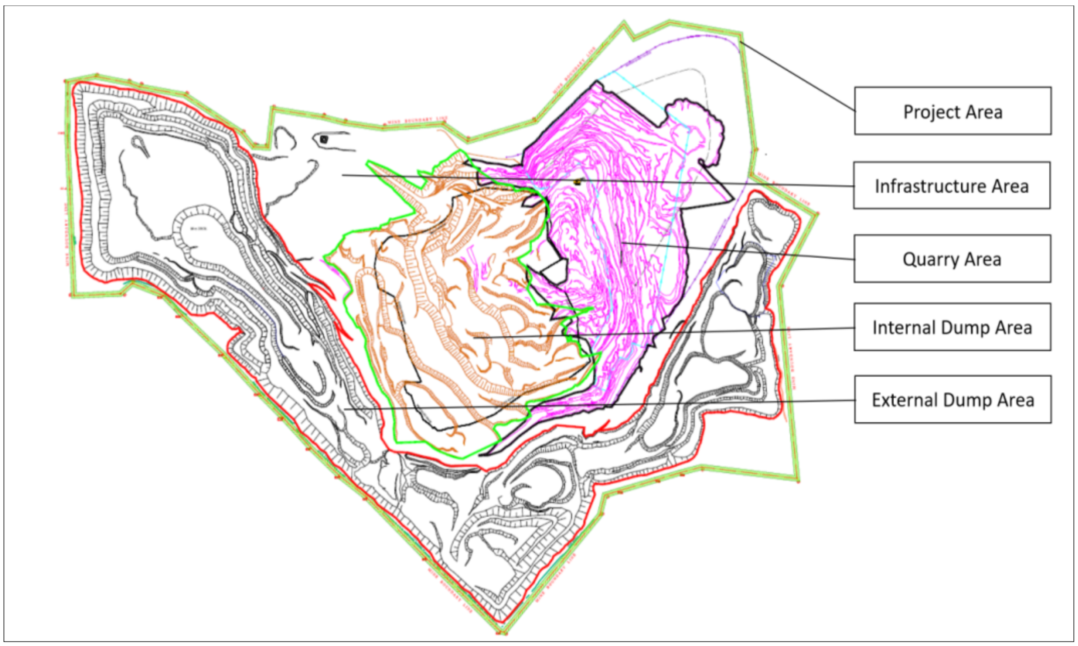

Figure 2. Working plan of opencast mine.

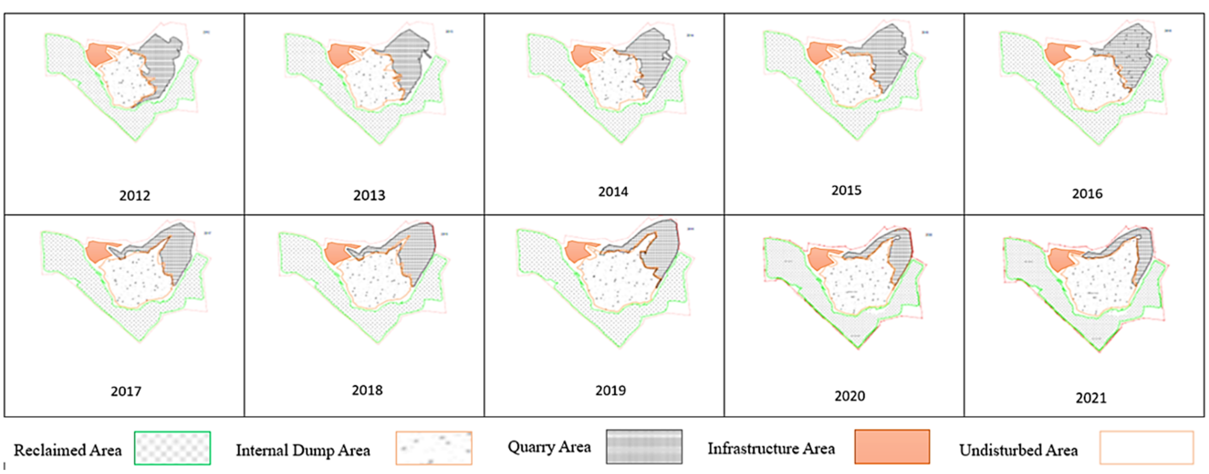

Figure 3. Different areas during the years 2012 to 2021.

ii. Quarry depth

The dispersion of the pollutants from the quarry varies with depth of working, and hence, the depth of the quarry has been taken from the year-wise working plans on the dates of monitoring, for a period of 10 years.

iii. Overburden dump height

The varying dump heights, dependent on the progress of the mine workings, have also been collected from the year-wise plans.

iv.Transportation distance of coal and overburden from the point of extraction to dumping area.

The transportation of coal was done by 35 t trucks, whereas that of overburden was done using 50 t trucks. The average distance of coal and overburden transportation has been recorded for the specified dates.

v. Quantity of coal produced and overburden removed

The quantity of coal extraction and overburden removal was collected for the particular days during which air quality monitoring was carried out.

The minimum and maximum values of the mining parameters processed from the year 2012 to 2022 are given in Table 1.

2.1.2. Meteorological data

The mine area experiences hot summers reaching 48^0C in peak summer. The winter season is from November to early March followed by summer season from mid of March to early June. The period from mid of June to September receives rains due to southwest monsoon, while the period from October to November forms the post-monsoon season with occasional rains from northeast monsoon.

A meteorological station was functioning at the project site for continuous monitoring of meteorological parameters, and also to aid in establishing a network of air quality monitoring stations in and around the project based on predominant wind direction. The meteorological station had sensors for recording wind speed, wind direction, temperature, relative humidity, and rainfall. Data pertaining to meteorological conditions existing in the surrounding area has been collected for the corresponding dates of air quality monitoring from a standard WM271 system, as shown in Figure 4.

Table 1. Minimum and maximum values of mining parameters from 2012 to 2022.

Effective area contributing to pollution (ha)		Quarry depth (m)		Overburden dump height (m)		Lead (km)		Quantity of coal produced and overburden removed (t)	
Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
121.83	340.03	182	235	70	80	0.2	5	8,390	148,682

The meteorological data of daily average temperature, humidity, rainfall, predominant wind direction, and wind speed on the dates of particulate matter monitoring has been collected from the year 2012 to 2022.

2.1.3. Particulate matter (PM) data

The ambient air quality monitoring was carried out in the opencast project as per the guidelines of Central Pollution Control Board (CPCB) and Ministry of Environment, Forest & Climate Change (MoEF&CC). The air quality monitoring was carried out at a frequency of once in a fortnight at 2 locations in the core/work zone environment and 4 locations in the surrounding villages. Particulate matter of two critical sizes viz., PM_{10} and $PM_{2.5}$ using the United States Environmental Protection Agency (USEPA) approved Respirable Dust Sampler and Fine Particulate Sampler, respectively.

The samplers were placed at a height of 3.0m above the ground level and at least 2.0m apart, while monitoring for negating the effects of wind-blown ground dust. The samplers were placed at open space free from any obstruction including trees and vegetation which otherwise act as a sink of pollutants resulting in lower levels of concentration. Two monitoring locations in the core zone were identified near Coal Handling Plant (CHP) and Base Workshop (BWS). Similarly, 4monitoring locations in the buffer zone are identified near 4 different villages in the vicinity of the opencast project covering all four directions.

Fortnightly air quality monitoring data of 24 hours sampling duration for PM_{10} and $PM_{2.5}$ has been collected for a period of 10 years from the year 2012 to 2022. The monitoring points at CHP, BWS, and other locations in the buffer zone remained constant and these points have not been relocated during this period. Figure 5 shows the Google image of the opencast mine and air quality monitoring locations identified in the core and buffer zone of the project.

Figure 4. Micro-meteorological station.

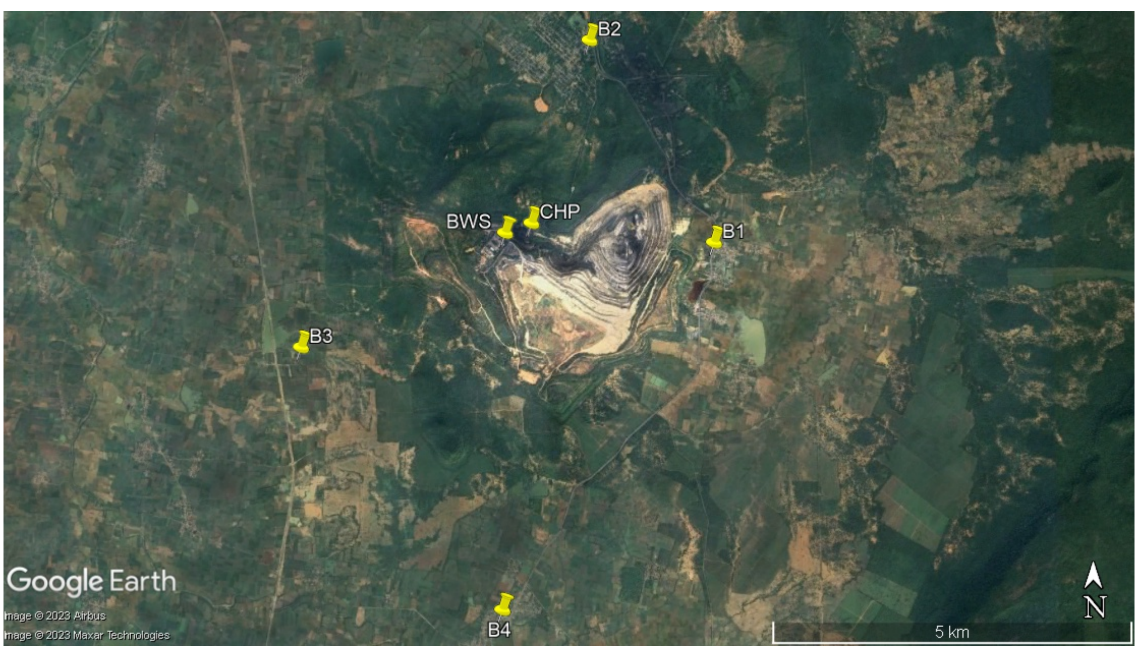

Figure 5. Location of mine and monitored points.

The distance of air quality monitoring locations established at the villages from the mine site is given in Table 2.

Table 2. Directions and distances of villages from the mine boundary.

Station code	Direction	Distance from the mine (m)
B1	East	1000
B2	North	1500
B3	West	2000
B4	South	2500

2.2. Data analysis using machine learning techniques

An approach is made in this study to analyze the concentration of particulate matter at different locations in and around the opencast coal mine vis-a-vis coal produced, overburden extracted, area contributing to pollution, meteorological conditions using different machine learning techniques. Machine learning is a branch of Artificial Intelligence (AI) and computer science, which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy by itself. Machine learning algorithms are classified into 3 types:

1. Supervised learning: This algorithm consists of a target / outcome / dependent variable, which is to be predicted from a given set of predictors / independent variables. Using these sets of variables, a function has been generated that maps inputs to desired output. The training process continues until the model achieves a desired level of accuracy on the training data.
2. Unsupervised learning: In this algorithm, we will not have any target or outcome variable to predict / estimate. It is used for clustering population into different groups, which is widely used for segmenting customers in different groups for specific intervention.
3. Reinforcement learning: Using this algorithm, the machine is trained to make specific decisions as the machine is exposed to an environment where it trains itself continually using trial and error.

Since the data has target/response variable, supervised learning algorithms are used for choosing the model. Choosing the right algorithm will depend on the type of the problem to be solved, and also depends on the scale of the dependent variable. In case of continuous target variable, regression algorithms are used and in case of categorical target, classification algorithms are preferred.

For the given data, the response variables are CHP, BWS, and 4villages at which PM_{10} and $PM_{2.5}$are monitored. The remaining features like mine parameters and meteorological data are independent variables. Since the data is related to time and the response variables are continuous, many of the methods have high time complexity and the methods, which take less time to execute when compared to others, are random forest, decision tree, and bagging algorithms. Hence, this analysis attempts to apply these techniques to the data and compare their performance metrics and choose the best method for prediction. In order to validate the models, the entire data is divided in to two parts viz., training data and testing data. For 80-20 ratio, over 192 observations came under training data with 48 cases for test data.

Decision tree involves dividing the dataset based on relationships between explanatory and outcome/response variable, i.e. the tree building starts by finding the variable / feature for the best split. Finding such variable is done by criterions like Entropy, Information Gain, Gini Index, Chi square test etc. (which are purely mathematical concepts).

Random forest generates the result, which is the average value/result/prediction of several decision trees and these decision trees are formed by taking different training and test data samples each time randomly. For example, training and test datasets, which form the decision tree, need not be same for the rest of the decision trees. In this method, only few independent variables (most probably important variables) are considered for analysis.

Bagging works same as random forest, but the major difference is that it considers all the independent variables for finding the ultimate solution.

Performance: For continuous variables, there are 3 performance metrics viz., Mean Square Error (MSE), Root Mean Square Error (RMSE), and co-efficient of determination (R^2), which can be useful for knowing the best model for the particular dataset. The performance metrics of test data are compared for different models in order to find the best model among the three.

3. Results and Discussion

The details of mine configuration like combined coal and overburden, area contributing to pollution, quarry depth, overburden dump height, and distance of transportation (lead distance) for 10 years from 2012-22 is given in Figure 6.

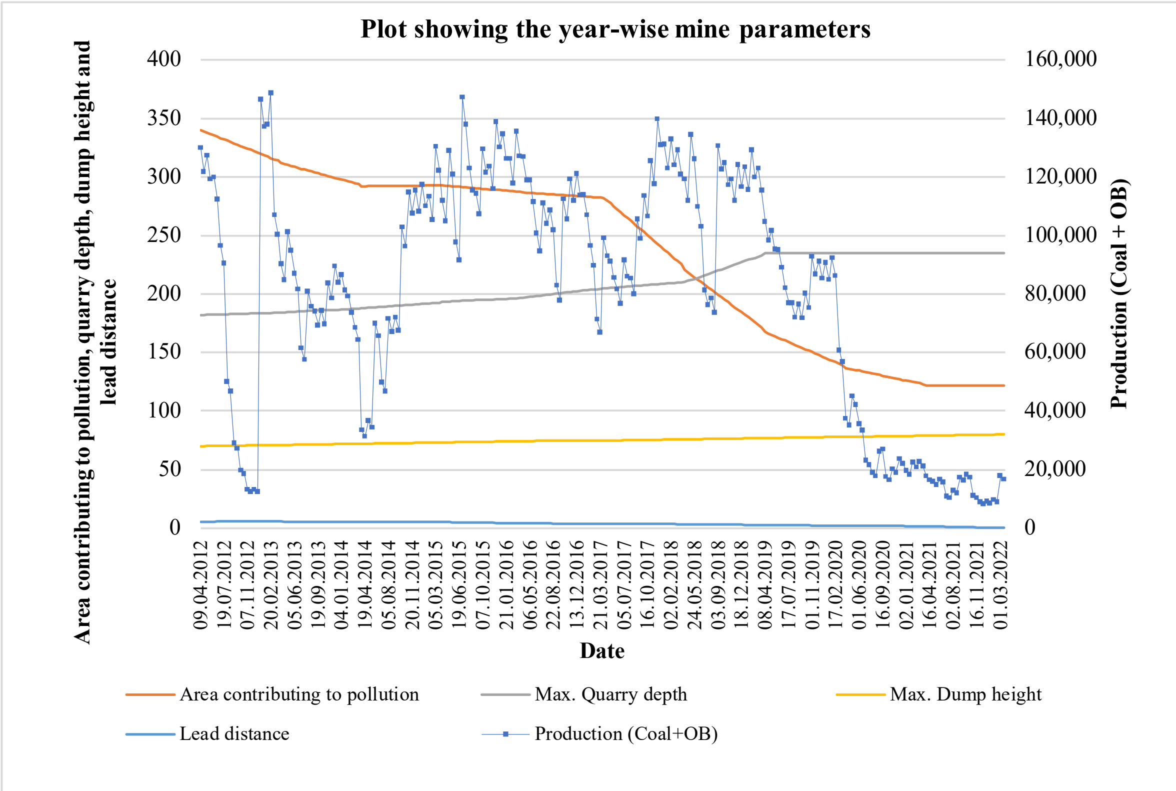

Figure 6. Details of mine parameters during dust monitoring.

The combined coal production and overburden removal ranged from 8,390 t per day to 1,48,682 t per day. The net effective area contributing to pollution was reduced from 340.03 Ha to 121.83 Ha. The depth of the quarry was ranging from 182m to 235m, whereas the height of the overburden dump was ranging from 70m to 80m.Similarly, the distance of transportation reduced from 5.0 km to 0.2 km due to back-filling operations.

Figure 7 shows the meteorological data viz., predominant wind direction, average humidity, average wind speed, average temperature, and average rainfall from the year 2012-22.

The daily average temperature varied from 17.2^0 to 45.8^0C, predominant wind direction from 1 to 360 degrees, humidity from 25.4% to 99%, rainfall from 0 to 74.2mm, and average wind speed from 0 to 31.8m/s.

The plot of PM_{10} and $PM_{2.5}$ in the core and buffer zones are given in Figure 8 and Figure 9, respectively, for the last 10 years from 2012-22.

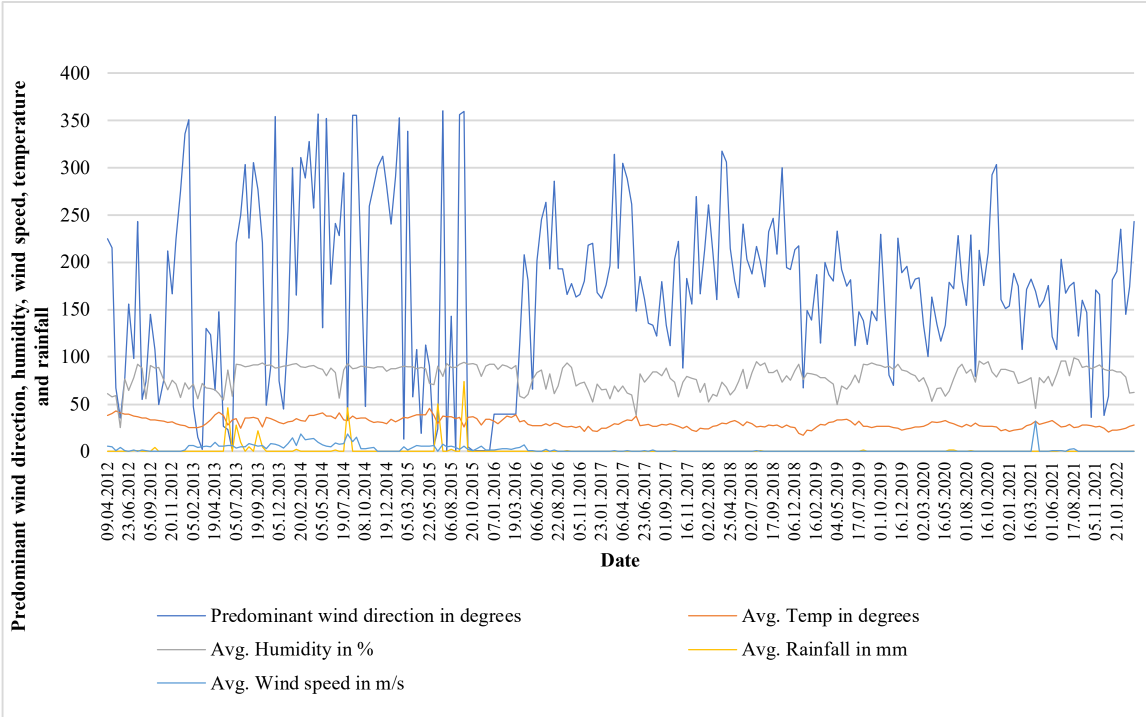

Figure 7. Fortnightly meteorological data from 2012-22.

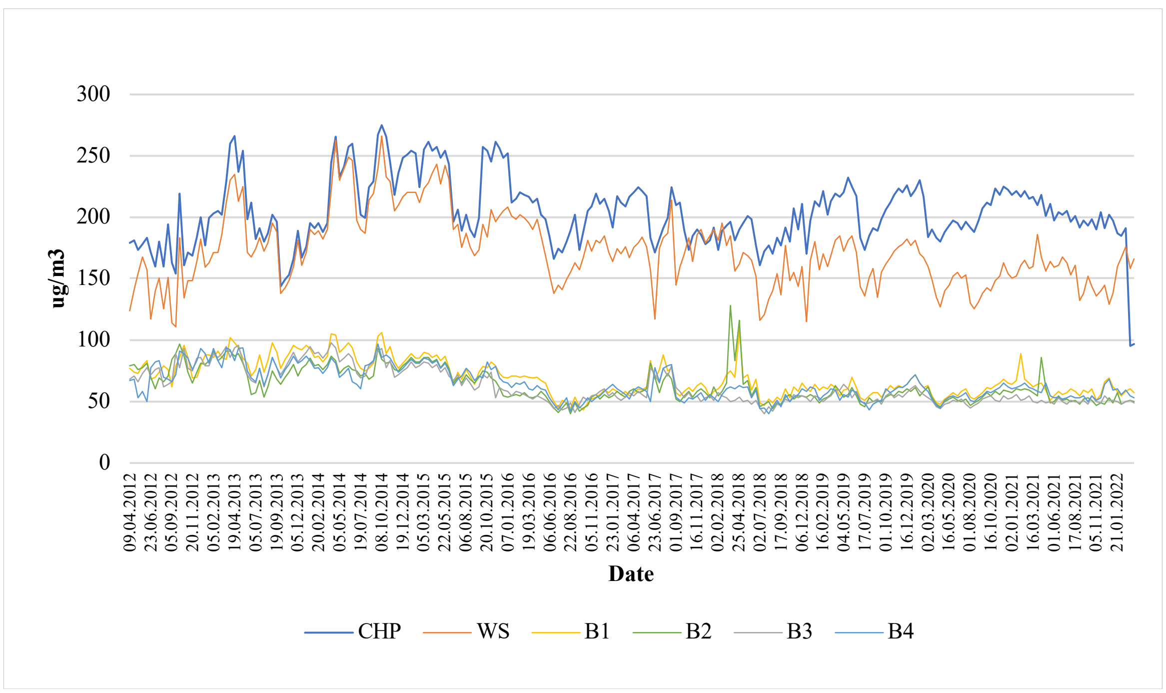

Figure 8. PM_{10} values in the core and buffer zones.

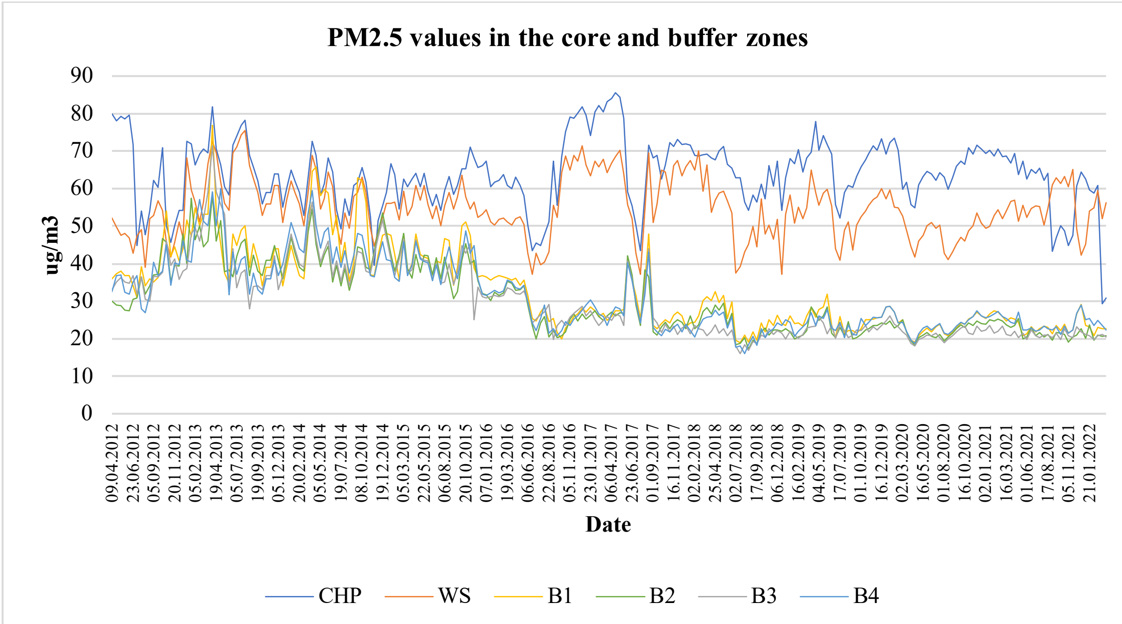

Figure 9. $PM_{2.5}$ values in the core and buffer zones.

The daily average PM_{10} in the CHP varied from 95 to 275 ug/m^3, BWS from 57 to 266 ug/m^3, B3 from 40 to 98 ug/m^3, B1 from 43 to 108 ug/m^3, B4 from 40 to 97 ug/m^3, and B2 from 40 to 128 ug/m^3.

The daily average $PM_{2.5}$ in the CHP varied from 29.4 to 85.6 ug/m^3, BWS from 37.1 to 75.5 ug/m^3, B3 from 16.1 to 71.4 ug/m^3, B1 from 18.9 to 76.9 ug/m^3, B4 from 16.1 to 60 ug/m^3, and B2 from 17.3 to 58.3 ug/m^3.

3.1. Performance of machine learning techniques at different locations

The fortnightly coal production and overburden removal from the mine from 2012-22 and the corresponding configuration of the mine parameters with the meteorological conditions have been considered for analysis. Among the total data analyzed, 80% of the data is used for training, and the rest 20% is used for validation at different locations.

The following tables (Tables 3 to 8) show the performance of the machine learning techniques for the PM_{10}response variable in the core and buffer zones of the mine.

Table 3. Performance at CHP (PM_{10}).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	1.9486	4.262	11.6161	188.7941	0.8307	21.6497
Root mean square error	1.3959	18.169	3.4082	13.7402	0.9114	4.6529
R^2 (in %)	99.73	98.32	98.4	82.5	99.89	97.99

Table 4. Performance at BWS (PM_{10}).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	2.8952	0.5918	13.4509	50.1785	0.4761	2.1153
Root mean square error	1.7015	0.7692	3.6675	7.0836	0.6902	1.4544
R^2 (in %)	99.7	99.93	98.6	93.93	99.95	99.74

Table 5. Performance at B1 (PM_{10}).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	0.0402	0.1168	1.146	2.602	0.077	0.2609
Root mean square error	0.2006	0.341	1.0707	1.613	0.2787	0.5108
R^2 (in %)	99.98	99.95	99.43	98.85	99.96	99.98

Table 6. Performance at B2 (PM_{10}).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	0.0427	0.1079	1.265	7.8659	0	0.5208
Root mean square error	0.2067	0.3285	1.125	2.8046	0	0.721
R^2 (in %)	99.97	99.94	99.2	95.9	100	99.73

Table 7. Performance at B3 (PM_{10}).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	0.0513	0.6459	1.405	32.35	0	0.5417
Root mean square error	0.2264	0.836	1.204	5.688	0	0.7359
R^2 (in %)	99.97	99.73	99.26	86.68	100	99.78

Table 8. Performance at B4 (PM_{10}).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	1.3891	1.5996	3.0836	37.2282	0.0591	3.5843
Root mean square error	1.1785	1.264	1.7560	6.1014	0.2430	1.8932
R^2 (in %)	99.23	99.35	98.29	84.95	99.97	98.55

As per performance metrics, bagging technique shows the highest accuracy of 98.32% and 99.93% for predicting new data in the core zone stations at CHP and BWS, respectively.

However, bagging shows the highest accuracy of 99.94% and 99.35% at two locations B2 and B4 in buffer zone. In case of other two locations, bagging and decision tree have shown same level of accuracy, i.e. at B1 (99.95% and 99.98%) and B3 (99.73% and 99.78%). Hence, any of the methods can be used for future predictions. The following tables (Tables9 to 14) show the performance of the machine learning techniques for the $PM_{2.5}$ response variable in the core and buffer zones of the mine.

Table 9. Performance at CHP ($PM_{2.5}$).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	0.1538	0.736	1.1173	7.3087	0	0.5613
Root mean square error	0.3921	0.8578	1.057	2.7034	0.005	0.7492
R^2 (in %)	99.82	99.15	98.71	91.55	100	99.35

Table 10. Performance at BWS ($PM_{2.5}$).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	0.0229	0.0694	1.6432	5.4896	0.0001	0.088
Root mean square error	0.1508	0.2633	1.2818	2.3429	0.0114	0.2979
R^2 (in %)	99.96	99.87	97.44	89.74	100	99.83

Table 11. Performance at B1 ($PM_{2.5}$).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	0.0388	7.3324	0.6151	12.9164	0.0454	5.4726
Root mean square error	0.1970	2.7078	0.78	3.5939	0.2130	2.339
R^2 (in %)	99.96	94.87	0.9931	90.97	99.95	96.17

Table 12. Performance at B2 ($PM_{2.5}$).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	0.6074	2.074	0.0383	0.281	0.0575	0.3206
Root mean square error	0.7793	1.4401	0.1957	0.079	0.2396	0.5662
R^2 (in %)	99.32	97.83	99.96	99.92	99.94	99.66

Table 13. Performance at B3 ($PM_{2.5}$).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	0.0337	4.6675	0.7005	17.884	0.1054	4.2253
Root mean square error	0.1835	2.16	0.8369	4.228	0.324	2.055
R^2 (in %)	99.97	96.81	99.38	87.78	99.91	97.11

Table 14. Performance at B4 ($PM_{2.5}$).

	Bagging		Random forest		Decision tree	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Mean square error	0.0905	0.7425	0.4801	3.6092	0.0129	0.1354
Root mean square error	0.3007	0.5514	0.6910	1.8997	0.1130	0.3679
R^2 (in %)	99.9	99.43	99.45	96.29	99.99	99.86

Bagging and decision tree accuracies are very close to each other in the core zone locations, CHP (99.15%, 99.35%) and BWS (99.87%, 99.83%). Hence, any of the methods can be used for future predictions. In case of buffer zone, decision tree has shown higher level of accuracy at all the locations, B1 (96.17), B2 (99.66), B3 (97.11), and B4 (99.86).

3.2. Parameters influencing response

From the performance metrics, the significant parameters, which are responsible for dust dispersion, have to be known in the order of preference for each location. Recursive Feature Elimination (RFE) technique is used to know the rank order of the independent variables in terms of importance. RFE is a valuable technique in machine learning for optimizing model performance, reducing over-fitting and enhancing interpretability by selecting the most important features, while discarding less relevant ones. From the best fit algorithm identified for each location, the parameters/features have been eliminated one after the other thus determining the performance of each parameter. This accuracy has been compared with the actual accuracy, and thus ranking has been given based on the difference in accuracies. If the difference is less, the significance of the parameter is more and vice-versa. Thus, lead distance played a significant role, whereas the rainfall played the minimum role in dust concentrations at different locations.

The rankings based on the influence of different parameters for PM_{10} and $PM_{2.5}$are given in Tables15 and 16, respectively.

Table 15. Ranking of parameters with respect to PM_{10}.

Parameter	CHP	BWS	B1	B2	B3	B4
Lead distance (m)	1	1	3	1	3	3
Max. dump height (m)	2	2	1	2	4	1
Area contributing to pollution (ha)	3	3	2	4	2	2
Silt content in OB (%)	6	4	4	3	1	4
Max. quarry depth (m)	4	5	5	5	5	5
Moisture content in OB (%)	7	6	6	9	6	6
Avg. temp. (^{0}C)	5	9	7	10	10	7
Production (Coal+OB) (t)	11	8	10	7	8	8
Avg. wind speed (m/s)	9	7	9	11	7	10
Avg. humidity (%)	10	10	8	8	9	9
Predominant wind dir. (^{0})	8	11	11	6	11	11
Avg. rainfall (mm)	12	12	12	12	12	12

Table 16. Ranking of parameters with respect to $PM_{2.5}$.

Parameter	CHP	BWS	B1	B2	B3	B4
Lead distance (m)	2	1	3	2	5	2
Max. dump height (m)	1	2	4	5	1	3
Area contributing to pollution (ha)	6	4	1	4	4	4
Max. quarry depth (m)	11	10	2	3	1	1
Silt content in OB (%)	10	8	5	1	3	5
Avg. temp. (^{0}C)	5	3	7	7	11	7
Production (Coal+OB) (t)	3	5	6	6	9	11
Avg. humidity (%)	4	6	9	9	7	10
Avg. wind speed (m/s)	8	7	8	11	10	6
Moisture content in OB (%)	7	11	11	8	6	8
Predominant wind dir. (0)	9	9	10	10	12	12
Avg. rainfall (mm)	12	12	12	12	8	9

50% of the Total Suspended Particulates is emitted during the transportation of material over unpaved roads in the mine and studies also concluded that the largest emission source for TSP is transportation over unpaved roads [18]. This result agrees that material handling and wind erosion are the other two main source of TSP [6]. The present study also revealed that transportation and dump height are playing a major role. Other studies have considered either single parameter like depth of the mine or only meteorological parameters, whereas the present study has considered the mining parameters as well as meteorological parameters for analysis and integrated both the parameters in assessing the concentration of particulate matter at varying levels of mining operations.

4. Conclusions

Opencast mining plays a lead role in supply of large quantities of coal to thermal power plants in India. However, opencast mining results in higher degree of air pollution when compared to underground mining. At this juncture, ascertaining pollution levels at different locations in the opencast mines and surrounding residential areas is of paramount importance for implementing necessary mitigation measures. Machine learning model generates predictions based on real time data and recommend actions to optimize operations and control measures. Earlier studies using machine learning techniques have mainly dealt with the predictions based on meteorological data and its impact on particulate matter. In contrast, the present study involves integration of mining parameters in to the model, which are dynamic in nature, in order to achieve more realistic predictions.

In the present study, in case of PM_{10}, bagging has shown the highest accuracy for predicting new data in the core zone stations. However, bagging has shown the highest accuracy at two locations in the buffer zone, while at the other two locations, bagging and decision tree have exhibited similar levels of accuracy.

In case of $PM_{2.5}$, the accuracies of bagging and decision tree are closely aligned in the core zone locations, while decision tree has shown high level of accuracy at all the locations in the buffer zone.

The study also involved usage of recursive feature elimination algorithm to investigate significant parameters contributing to dust pollution in different locations in and around the opencast coal mine. The different parameters pertaining to mining, meteorological, and particulate matter taken once in a fortnight at the core and buffer zones of the mine for a period of 10 years are considered. Analysis has been done based on various machine learning techniques and ranking for each feature has been given based on the significance of contribution to pollution at that particular location.

The key factors contributing to particulate concentrations have been established through the model are summarized below:

1. The CHP, BWS, and B2 are affected by PM_{10}, primarily due to the lead distance followed by dump height as these three locations are near to the opencast mine and due to predominant wind direction, whereas, B4 and B1 being on the upwind side has lesser impact due to their presence in up wind direction and hence dump height and silt content played a prominent role.
2. With regard to $PM_{2.5}$, CHP, B3, and BWS are affected due to dump height as CHP is located within the mine area, and B3 is located in downwind direction. BWS is affected due to lead distance as it is lying adjacent to the haul road.
3. In case of PM_{10}, lead distance is playing a significant role in contributing to dust pollution followed by dump height, effective area contributing to pollution, silt content, quarry depth. It is also evident that rainfall is contributing to the least.
4. In case of $PM_{2.5}$ also, lead distance is playing a major role in contributing to pollution followed by dump height followed by effective area contributing to pollution, etc. In this case also, rainfall has the least ranking.

The model can be regularly updated and retrained with new data to adapt to changes in equipment, processes, and environmental conditions in order to further improve the accuracy of predictions over time. The machine learning can assist mine management in optimizing the operations to contain the pollution levels within the stipulated standards thereby ensuring green and climate smart mining operations.

Subsequent research endeavors could centre on employing advanced and potent algorithms like Cuckoo search, particle swarm optimization or extreme learning machine, particularly in contexts involving the study of multiple mines. Integrating additional parameters into the model could enhance the accuracy of predicting dust concentrations, facilitating regular use for regulatory compliance purposes.

References

[1]. Alimoradi, A., Maleki, B., Karimi, A., Sahafzadeh, M., & Abbasi, S. (2020). Integrating geophysical attributes with the new cuckoo search machine-learning algorithm to estimate silver grade values– Case Study: Zarshouran Gold Mine, *Journal of Mining and Environment*, *11*(3), 865-879.

[2]. Bade, T. & Harion, J.L. (2007). Effect of aggregate storage piles configuration on dust emissions. *Atmos. Environ*. *41*(2), 360-368.

[3]. Bhaskar, R. & Ramani, R.V. (1988). Behaviour of dust clouds in mine airways, *Trans. AIME*, *280*, 2051-9.

[4]. Carvacho, O.F., Ashbaugh, L.L., Brown, M.S., & Flocchini, R.G. (2004). Measurement of $PM_{2.5}$ emission potential from soil using the UC Davis resuspension test chamber, *Geomorphology. 59*(1-4), 75-80.

[5]. Chakraborty, M.K. *et al*. (2002). Determination of the emission rate from various opencast mining operations. *Environmental Modelling and Software, 17*, 467–480.

[6]. Chaulya, S.K. (2004). Air quality status of an open pit mining area in India. *Environmental Monitoring and Assessment, 105*, 369–389.

[7]. Chavez, M., Hajra, B., Stathopoulos, T., & Bahloul, A. (2011). Near-field pollutant dispersion in the built environment by CFD and wind tunnel simulations. *Wind Eng. Ind. Aerod, 104*, 509-515.

[8]. Cooper, N., Green, D., & Meissner, K.J. (2017). The Australian National Pollutant Inventory fails to fulfil its legislated goals. *Environ. Res. Public Health, 14*(5), 478.

[9]. Davari, M.A., Senemari, S., & Alimoradi, A. (2024). Permeability prediction from log data using machine learning methods, *Journal of Petroleum Geomechanics*, doi: 10.22107/JPG.2024.426878.1220.

[10]. Gautam, S. & Patra, A.K. (2015). Dispersion of particulate matter generated at higher depths in opencast mines. *Environmental Technology and Innovation, 3*, 11-27.

[11]. Gautam, S. *et al*. (2012). Opencast mines: a subject to major concern for human health. *Geology and Mining, 2*(2), 25-31.

[12]. Ghasemi Tabar, H.R., Alimoradi, A., Hemmati Ahooi, H.R., & Fathi, M. (2023). Intelligent borehole simulation with python programming, *Journal of Mining and Environment*, doi: 10.22044/jme.2023.13610.2527.

[13]. Ghose, M. K. (2002). Air pollution due to opencast coal mining and the characteristics of air-borne dust - An Indian scenario. *International Journal of Environmental Studies, 59*(2), 211–228.

[14]. Ghose, M.K. & Majee, S.R. (2000). Assessment of the impact on the air environment due to opencast coal mining – an Indian case study. *Atmospheric Environment, 34*, 2791–2796.

[15]. Ghose, M.K. & Majee, S.R. (2007). Characteristics of hazardous airborne dust around an Indian surface coal mining area. *Environ Monit Assess, 130*, 17-25.

[16]. Haj Karimian, H., Alimoradi, A., Hemmati Ahoei, H. R., & Salsabili, M. (2022). Comparison between the performance of four metaheuristic algorithms in training a multilayer perceptron machine for gold grade estimation, *International Journal of Mining and Geo-Engineering, 56*(2), 97-105.

[17]. Holmes, N.S. & Morawska, L. (2006). A review of dispersion modelling and its application to the dispersion of particles: an overview of different dispersion models available. *Atmos, 40*(30), 5902–5928.

[18]. Huertas, J.I., Camacho, D.A. & Huertas, M.E. (2012). Standardized emissions inventory methodology for open-pit mining areas. *Environmental Science and Pollution Research, 19*(7), 2784–2794.

[19]. Lokman Hakan Tecer, Pinar Süren, Omar Alagha, Ferhat Karaca & Gürdal Tuncel. (2008). Effect of meteorological parameters on fine and coarse particulate matter mass concentration in a coal-mining area in Zonguldak, Turkey. *Journal of the Air & Waste Management Association, 58*(4), 543-552, doi: 10.3155/1047-3289.58.4.543.

[20]. Luan B, Zhou W, Jiskani I.M, Wang Z. (2023). An improved machine learning approach for optimizing dust concentration estimation in open-pit mines. *International Journal of Environmental Research and Public Health, 20*(2), 1353, https://doi.org/10.3390/ijerph20021353.

[21]. Mandal, K. (2012). Characterization of different road dusts in opencast coal mining areas of India. *Environ Monitoring and Assessment, 184*, 3427-3441.

[22]. Michael Hendryx, Mohammad Saidul Islam, Guang-Hui Dong, & Gunther Paul (2020). Air pollution emissions 2008–2018 from Australian coal mining: Implications for public and occupational health. *Environ. Res. Public Health, 17*, 1570.

[23]. Mukhopadhyay, S. (2010). Ambient air quality in opencast coal mining areas of Bankola area (under Eastern Coal Field ltd.) of Asansol-Raniganj regions. *The Ecoscan, 4*(1), 19-24.

[24]. National pollutant inventory emission estimation technique manual for mining, Version 3.1, January 2012.
www.npi.gov.au/system/files/resources/7e04163a-12ba-6864-d19a/mining.pdf.

[25]. Reed, W.R., Westman, E.C., & Haycocks, C., The introduction of a dynamic component to the ISC 3 model in predicting dust emissions from surface mining operations. 2002, *30th International Symposium on the Application of Computers and Operations Research in the Mineral Industry*, APCOM 2002, Phoenix, AZ (United States), 659–667.

[26]. Claire Richardson, Shannon Rutherford, & Igor Agranovski (2018). Characterization of particulate emissions from Australian open-cut coal mines: Toward improved emission estimates. *Air & Waste Management Association, 68*, 598-607.

[27]. Sammy, G.K. & Canter, L.W. (1983). Environmental Impact Assessment in developing countries: What are the problems? *Impact Assessment, 2,* 29-43.

[28]. Silvester, S. A., Lowndes, I.S., & Hargreaves, D.M. (2009). A computational study of particulate emissions from an open pit quarry under neutral atmospheric conditions. *Atmos. Environ, 43*(40), 6415-6424.

[29]. Srivastava, A., Kumar, A., & Elumalai, S.P. (2021). Evaluating dispersion modeling of inhalable particulates (PM_{10}) emissions in complex terrain of coal mines. *Environmental Modeling & Assessment, 26*(3), 385–403.

[30]. Sumanth Chinthala & Mukesh Khare, Particle dispersion within a deep open cast coal mine, *Air Quality - Models and Applications*, Prof. Dragana Popovic (Ed.), ISBN: 978-953-307-307-1.

[31]. Tartakovsky, D., Broday, D.M., & Stern, E. (2013). Evaluation of AERMOD and CALPUFF for predicting ambient concentrations of total suspended particulate matter (TSP) emissions from a quarry in complex terrain. *Environ. Pollu., 179,* 138–145.

[32]. United States Environment Protection Agency (USEPA), Compilation of air pollutant EFs: Stationary point and area sources, external combustion sources: Bituminous and sub-bituminous coal combustion Final section. AP 42, Fifth Ed. 1, 1998.

[33]. Boyu Luan *et al.* (2023). An improved machine learning approach for optimizing dust concentration estimation in open-pit mines. *Environ. Res. Public Health, 20*(2), 1353.

تجزیه و تحلیل غلظت ذرات معلق محیطی در منطقه اطراف یک معدن زغال سنگ روباز با استفاده از تکنیک‌های یادگیری ماشین

راوی کیران پودیچتی و رام چاندار کاررا*

گروه مهندسی معدن، موسسه ملی فناوری کارناتاکا، سوراتکال، مانگالور، هند

ارسال ۲۰۲۳/۱۲/۱۷، پذیرش ۲۰۲۴/۰۲/۰۲

* نویسنده مسئول مکاتبات: krc_karra@yahoo.com

چکیده:

معادن زغالسنگ روباز نقش مهمی در تامین نیازهای انرژی یک کشور دارند. با این حال، این عملیات منجر به بدتر شدن کیفیت هوای محیط، به ویژه بـه دلیـل انتشار ذرات خواهد شد. پراکندگی ذرات معلق بر اساس پارامترهای معدن و شرایط جوی محلی متفاوت خواهد بود. نیاز به ایجاد یک مدل مناسب برای پیش‌بینی غلظت ذرات معلق به صورت منطقه‌ای وجود دارد. اگرچه تعدادی از مدل‌های پراکندگی برای پیش‌بینی غلظت گرد و غبار به دلیـل اسـتخراج روبـاز وجـود دارد، یادگیری ماشینی مزایای متعددی را نسبت به تکنیک‌های مدل‌سازی سنتی از نظر بینش مبتنی بر داده، غیرخطی بودن، انعطاف‌پذیری، مدیریت برهم‌کنش‌های پیچیده، تشخیص ناهنجاری و غیره ارائه می‌کند. برای ارزیابی پراکندگی ذرات با استفاده از تکنیک‌های یادگیری ماشین با در نظر گـرفتن پارامترهـای معـدن و هواشناسی ساخته شده است. داده‌های تاریخی شامل پارامترهای کار معدن، شرایط هواشناسی، و ذرات معلق مربوط به یکی از معادن زغالسنگ روبـاز در جنـوب هند برای این مطالعه استفاده شده است. داده‌ها با استفاده از تکنیک‌های مختلف یادگیری ماشین مانند بسته‌بندی، جنگل تصـادفی و درخـت تصـمیم تجزیـه و تحلیل شده است. معیارهای عملکرد داده‌های آزمون برای مدل‌های مختلف مقایسه می‌شوند تا بهترین مدل برازش را در بین سه تکنیک پیدا کنند. مشخص شده است که برای PM10، بسیاری از مواقع تکنیک کیسه‌سازی دقت بهتری دارد و برای PM2.5، تکنیک درخت تصمیم دقت بهتری ارائه می‌دهد. ادغام پارامترهای کاری معدن با شرایط هواشناسی و داده های تاریخی ذرات معلق در توسعه مدل با استفاده از تکنیک‌های یادگیری ماشین به پیش‌بینی‌های دقیق‌تر کمک کـرده است.

کلمات کلیدی: معدن زغال سنگ باز، PM10، PM2.5، پراکندگی، یادگیری ماشین.