Modelling particulate matter concentration from loading operations in mineral quarries with a decision tree approach

by Z. Duran[1], B. Erdem[2], T. Dogan[3], M. Genc[4]

Affiliation:
1. Sivas Vocational School of Technical Sciences, Sivas Cumhuriyet University, Türkiye
2. Geophysics Engineering Department, Sivas Cumhuriyet University, Türkiye
3. Industrial Engineering Department, Sivas Cumhuriyet University, Türkiye
4. Hekimhan Mehmet Emin Sungur Vocational School, Malatya Turgut Ozal University, Türkiye

Correspondence to:
Z. Duran

Email:
zduran@cumhuriyet.edu.tr

Dates:
Received: 4 Jul. 2024
Revised: Nov. 2025
Accepted: 27 Feb. 2026
Published: May. 2026

How to cite:
Duran, Z., Erdem, B., Dogan, T., Genc, M. 2026. Modelling particulate matter concentration from loading operations in mineral quarries with a decision tree approach. *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 126, no. 5, pp. 261–278

DOI ID:
https://doi.org/10.17159/2411-9717/3496/2026

ORCiD:
Z. Duran
https://orcid.org/0000-0002-9327-8567
B. Erdem
https://orcid.org/0000-0002-1226-9248
T. Dogan
https://orcid.org/0000-0002-2628-4238
M. Genc
https://orcid.org/0000-0002-9950-0720

Abstract

This study aims to develop a model for particulate matter concentration during the loading process in open pit mining. The researchers conducted simultaneous measurements of particulate matter and meteorological parameters and collected samples to determine the moisture content of the loaded materials. The analysis used 7,895 measurement data points from gypsum and limestone quarries, employing two data analysis programs, SPSS® and Waikato Environment for Knowledge Analysis, to derive equations for predicting particulate matter release. In the modelling, particulate matter measurements were the dependent variables, while meteorological parameters, laboratory measurement results, and loader bucket capacities were the independent variables. The classical regression models did not adequately capture the dependent variable, thus, the researchers explored the decision tree approach for further modelling. The M5P algorithm was used to generate regression equations for the different data sets, and the findings showed that the models had a satisfactory degree of predictive capability. The number of fine particles released during loading is influenced by various weather factors. Temperature, humidity, wind speed, and station pressure can all affect the dispersal of these particles. Additionally, the moisture content of the loaded material and the capacity of the loading equipment contribute to this process. The M5P decision tree algorithm, which is rarely used in particulate matter concentration models, provides an innovative approach to developing local concentration estimates for non-coal surface mining.

Keywords

data mining, decision tree algorithm, M5P, particulate matter concentration, regression model

Introduction

Air pollution is a major environmental health problem that primarily affects residents of low- and middle-income countries. In 2019, particulate matter was responsible for an estimated 4.2 million premature deaths worldwide due to cardiovascular, respiratory, and cancer diseases (WHO, 2023). The term "dust" generally refers to particles smaller than 75 µm that can remain suspended in the air for a certain period (WHO, 1999). Dusts are defined as substances produced from various inorganic and organic materials during processes such as drilling, crushing, transportation, abrasion, grinding, decomposition, and combustion. They can range in size from 1 µm to 100 µm and settle under the influence of gravity (WHO, 1999; Vidinli et al., 2016).

The size of particulate matter (PM) is typically expressed as an aerodynamic diameter, defined as the diameter of a sphere with a density of 1 g/cm^3 that settles in still air under gravity at the same rate as the particle under the prevailing temperature, pressure, and relative humidity. PM size can range from a few nanometers (nm) to several tens of micrometers (µm) (Keskinkılıç et al., 2010). There are various approaches to defining the PM size range. Total suspended particles (TSP) has been used to refer to all

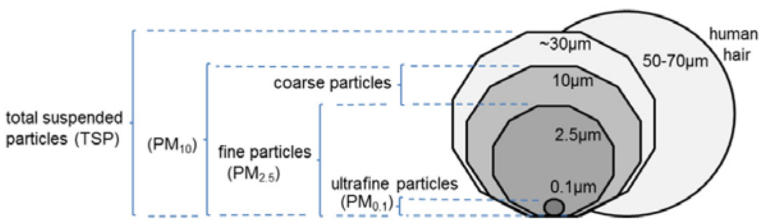


Figure 1—Classification of PM size range, not to scale (Hime et al., 2015)

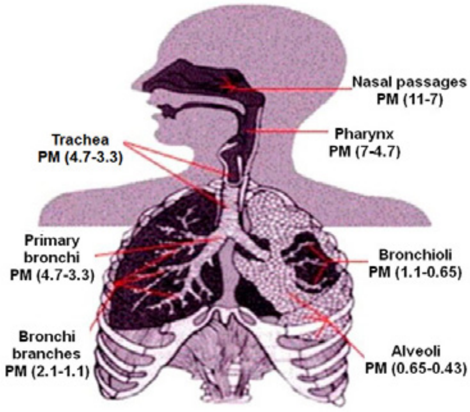


Figure 2—Collection areas of dust particles of different diameters in the human body (Löndahl et al., 2006 in Kim et al., 2015)

suspended particles of 40 µm and smaller (Chow, Watson, 1998; USEPA, 1998b) or 50 µm and smaller (Araújo et al., 2014). TSP has also been defined as particles between 0 µm and approximately 30 µm – 50 µm (Cao et al., 2013) or all suspended particles of 30 µm and smaller (Evyapan et al., 2012, Hime et al., 2015; NSW EPA, 2015; Patra et al., 2016a). Additionally, particles equal to or smaller than 10 µm are designated as PM10, while those equal to or smaller than 2.5 µm are designated as PM2.5 (WHO, 2006; Keskinkılıç et al., 2010; Evyapan et al., 2012; Araújo et al., 2014; Hime et al., 2015; NSW EPA, 2015; EPA, 2020). Figure 1 presents the classification of PM size ranges.

Figure 2 shows the diameters of particles entering the airways, lungs, and bronchi of the human body. The range of particulate matter (PM) that poses a risk to human health is 0.43 µm to 11 µm, while the PM sizes that reach the bronchi and alveoli range from 0.43 µm to 2.1 µm. $PM_{2.5}$, in particular, can remain suspended in the air for long periods and be transported over long distances. Inhaling particulate matter of this size can directly affect the respiratory system by reaching the deeper parts of the lungs. Workers in mines exposed to particulate matter of this size can suffer from diseases such as asthma, chronic lung disease, silicosis, asbestosis, berylliosis, inflammation, bauxite fibrosis, and siderosis (Gautam et al., 2012).

Surface mining releases a considerable amount of dust that can be hazardous to health and harmful to the environment, especially when dust particles are smaller than 100 µm. Trivedi et al. (2010), Gautam et al. (2012), and Arregocés et al. (2021) have reported on the health and environmental hazards associated with dust from open pit mining. Trivedi et al. (2010) specifically noted that TSP and PM_{10} emissions from surface coal mines are significant pollutants. Long before the environmental movement began, coal miners faced various occupational diseases that developed later in their working lives, especially during retirement. Long-term exposure to respirable coal dust causes several lung diseases, collectively known as coal miner's lung. In the United States, approximately 76,000 miners have died from this disease since it was legally recognised in 1969. However, reducing exposure limits to respirable dust is unlikely to be sufficient to reverse the increase in coal miner's lung disease observed among the current generation of miners (Duncan, 2015). Additionally, surface coal mining activities such as drilling, blasting, loading, unloading, storage, and transportation on unpaved roads have been identified as sources of air pollution (USEPA, 1995; Singh et al., 2016; Kumar, Kumar 2017; Rojano et al., 2020; Murzin, Gorlenko, 2021; Trechera et al., 2021).

Dust emissions from surface coal mining vary depending on the type of activity, such as drilling, loading, and transportation. Gautam et al. (2012) argue that modelling is essential for predicting and analysing the movement, distribution, and concentration of particulate matter released during mining activities. They also state that meteorological data is the most important factor determining the dilution of particulate matter. Emission prediction equations vary depending on the physical characteristics of the material, such as moisture and silt content, as well as equipment characteristics, such as bucket or dipper capacity, truck travel speed, work duration, and wind speed. Therefore, further work is needed to develop more consistent prediction equations (Ghose, 2004; Lashgari, Kecojevic, 2016).

Dust in mining is a significant problem affecting the environment, human health, workplace safety, and productivity in mining areas. Air and water pollution are among the most significant environmental problems facing communities in and around the copper mining areas of Zambia (Muma et al., 2020). Studies have shown that dust exposure in various work areas of coal mines can be hazardous (Petavratzi et al., 2005; Ghose, Majee, 2007; Papagiannis et al., 2014; Tripathy et al., 2015). In a recent study at the Lakhanpur opencast coal mine in India, dust exposure was monitored in different work areas (Tripathy et al., 2015). The aim of the study was to analyse dust collected from various sources, calculate personal exposure to dust, and estimate dust concentration in different areas of the mine and its vicinity. The results showed that dust emissions vary between mines and that most workers are exposed to dust concentrations exceeding permissible limits. The main sources of dust in mining were drills and bucket wheel excavators. A comparative analysis estimated dust emissions from electric excavators and wheel loaders in an open coal mine in the USA (Lashgari, Kecojevic, 2016). Three methods of dust estimation were used, including field measurements and laboratory studies: the EPA AP-42 equations to estimate emission factors, the methodology used in developing the AP-42 equations (Type 2 dust emission estimation), and the methodology used in the EPA AERMOD model (Type 3 dust emission estimation). The results showed that dust emissions estimated by the AP-42 method were higher than the area-based emissions determined by the Type 2 and Type 3 methods. In addition, the Type 2 method resulted in higher dust emissions compared to the Type 3 method. Önder and Yiğit (2009) calculated that in an open coal mine in Türkiye, the operator of a drill was exposed to the highest personal dust concentration of 3.08 mg/m^3, while the operator of a coal loader was exposed to the lowest personal dust concentration of 1.3 mg/m^3. In a separate study, researchers measured the amount of dust to which the operators of heavy machinery were exposed while working in three different opencast mines. Although the dust exposure of the operators of drills and crawler excavators was high, the dust concentration values for all operators were below the limit specified in the relevant regulations (Öztürk, 2016).

The primary objective of this study is to model the concentrations of particulate matter (PM) emitted into the atmosphere during loading operations, a key activity in open pit gypsum and limestone mining. While most studies on PM emissions have focused on minerals such as coal and iron ore, this research makes a unique scientific contribution by addressing the less-studied mining of gypsum and limestone. Monitoring fine particulates, particularly PM_{10}, $PM_{2.5}$, and PM_1, which are critical to human health, strengthens the environmental and public health aspects of the study and provides an important basis for air quality assessment. Methodologically, this study differs significantly from existing modelling approaches. While PM emissions are typically expressed in lb/ton, kg/ton, or g/s in the literature, this study uses PM concentration data directly emitted into the atmosphere during loading, and the models are presented in volume-normalised $\mu g/m^3$. This approach allows for a more concrete, on-site, and measurable assessment of the impacts of PM directly emitted into the environment on living organisms and ecosystems.

One of the most notable aspects of the study is the development of separate and independent concentration equations for TSP, PM_{10}, $PM_{2.5}$, and PM_1, which represents a rare level of specificity in modelling within the literature. Additionally, because the PM data measured in the study were widely distributed (TSP and PM10 ranged from 0 to over 6527.9, and $PM_{2.5}$ and PM_1 ranged from 0 to over 652.79), traditional linear regression techniques had low explanatory power. Therefore, the M5P algorithm, a decision tree algorithm, was chosen. The M5P algorithm improves the overall explanatory power of the model by incorporating both raw and transformed datasets into the analysis.

During the modelling process, variables such as meteorological factors (temperature, wind speed—including headwind and crosswind speeds—station pressure, relative humidity, etc.), moisture content of the loaded material, and loader bucket capacity were considered to develop PM concentration equations adaptable to similar mine site conditions. In the decision trees created for this purpose, specific PM classes (e.g., LM1, LM2) were specified at each leaf node, along with the number of samples and intraclass ratios, clearly demonstrating the model's classification success and the accuracy of the decision points. The explainability and visual interpretability provided by the decision tree structure offer significant advantages for developing field-based decision support systems. For these reasons, this study makes an original contribution to the literature in both content and methodology for modelling PM concentrations resulting from loading activities in gypsum and limestone open pit mining. The modelling structure, developed based on the M5P algorithm and enabling multidimensional data integration, aims to provide a scientific basis for sustainable environmental management strategies by systematically analysing the effects of environmental conditions and operational variables on PM concentrations.

Materials and methods

Study area

One gypsum quarry and two limestone quarries where PM concentration measurements were taken are located within the boundaries of the Sivas Basin, an orogenic sedimentary basin bounded by faults running northeast to southwest. The basin, situated in Central Anatolia, developed as a foreland basin after the subduction of the Neotethys Ocean during the Late Cretaceous (Poisson et al., 1996). Mining was conducted in the Oligocene-aged massive gypsum deposit, one of the Maastrichtian-Paleocene-aged limestone blocks, and in early Quaternary travertines in the gypsum, limestone (A), and L

limestone (B) quarries, respectively. Between July 2020 and November 2021, a total of 7,895 PM concentration measurements were taken in the quarries. To comply with the work permits negotiated with the mine administrations, the identities of the companies were not disclosed. The mining companies in the Sivas region did not operate between December and April due to adverse weather conditions.

In Sivas, which has a continental climate, precipitation occurs in winter, spring, and autumn, while summers are typically dry. The average annual rainfall is 420 mm, and the average snow depth is approximately 20 cm. Of the total precipitation, 22% falls in autumn, 36% in spring, 32% in winter, and 10% in summer. The average atmospheric pressure around Sivas is 853.2 millibars. As a low-pressure center, the area is particularly vulnerable to northern sector winds in summer. The northwest wind accounts for 19.3% of all winds in the Sivas region, followed by the north wind (18.1%), the northeast wind (16.8%), and other winds (SÇR, 2023).

Devices and sampling frequency

The portable PM monitor uses light scattering technology to measure the concentration of particles and dust in the air. Using a standard laser nephelometer detector, it simultaneously measures the characteristics of TSP, PM_{10}, $PM_{2.5}$, and PM1. The recorded data can be uploaded to a computer via PC-link for further analysis. The device responds quickly and can detect sources of dust and smoke in the air at concentrations as low as 0.1 $\mu g/m^3$. It can simultaneously measure TSP, PM_{10}, $PM_{2.5}$, and PM1 outdoors, and TSP, PM_{10}, and PM4 indoors. The device measures PM diameters from 0.5 to 20 µm, with particles larger than 20 µm reduced to this size. It accurately measures up to 6000 $\mu g/m^3$ (equivalent to 60 mg/m^3 without size selection) at an airflow rate of 0.6 l/min, with an accuracy of 0.01 $\mu g/m^3$ (Hwang et al., 2014; Dreshaj et al., 2017; Econorm, 2021; Turnkey, 2024). The PM monitor was new and used for the first time in this study (serial number DM12034). It was calibrated by a laboratory accredited by the Turkish Accreditation Agency (TÜRKAK) under registration number AB-0078-K for TS ISO IEC 17025:2017 on 17 February 2020. The relative error in volumetric flow rate (l/min) ranged from 1.8 to 5.3 per mille. In this study, AirQ™ software was used to transfer and process the collected data. The organised data were then uploaded to Excel™ software for statistical analysis.

The portable equipment used for weather monitoring (Kestrel 5500) complies with the Environmental Engineering Considerations and Laboratory Tests (MIL-STD-810G) standard. Air temperature, relative humidity, station pressure, dew point temperature, wind speed, headwind, and crosswind speed were measured and recorded in the device's internal memory. The data in CSV format were transferred to Excel™ via the device's integrated LiNK interface and processed simultaneously with the PM data.

Studies were conducted to optimise the sampling frequency. The initial measurement duration of one minute was insufficient to detect fluctuations in PM concentrations during the loading process, so the duration was kept short. Although both devices can measure continuously, the PM meter's battery is quickly depleted in this scenario. The battery, which has a backup, can measure for about one hour if data are collected every 5 seconds and up to 2.5 hours if data are collected every 10 seconds. The measurement interval was set to 10 seconds to maximise the use of the day shift for PM measurements and to collect as many samples as possible. The weather meter was started at the same time as the PM meter, and recordings were made at 5-second intervals. The data recorded by the PM meter during that period were used for the analyses.

The PM and meteorological parameter measurements were scheduled to last at least 10 minutes and to cover a full duty cycle of the loader. They were performed simultaneously to accurately represent the task. However, due to possible variations in the duration of the full duty cycle during the loading process, the measurement time was adjusted accordingly. The portable PM meter and weather meter device were placed at a safe distance of 2 to 10 metres from the loading area. The PM meter was placed on a three-legged tachymeter stand, while the weather meter was mounted on a separate tripod. The devices were positioned to collect samples at a height of 1.5 to 1.6 metres, corresponding to the breathing height of an adult (Figure 3).

Crawler excavators were used in the quarries. A Hitachi Zaxis 520 LCH and a Hidromek HMK 300 were used in the limestone (A) and gypsum quarries, respectively. A Hitachi Zaxis 490 LCH and a Hidromek 370 LCH operated in the limestone quarry (B). The bucket capacities ranged from 1.50 m^3 to 3.20 m^3.

To determine moisture content, samples were collected every 5 minutes during the measurement cycle. All 1-kg samples delivered to the laboratory on the same day were reduced using the coning and quartering method. Moisture content in limestone quarries was determined according to the TS 1900-1 (2006) standard, and in gypsum quarries according to the TS 4447 (1985) standard. Samples from the limestone quarries were weighed at 100 g in a laboratory sample container and placed in an oven at 105 °C for 2 hours. Samples from the gypsum quarry were weighed at 100 g in the sample container and placed in the oven at 40 °C for 30 minutes (Figure 4).

Data analysis

The study's three-stage methodology is outlined in the following (Duran, 2022). In the first stage, all PM variables (TSP, PM_{10}, $PM_{2.5}$, and PM_1) were measured simultaneously using the PM meter. The weather meter recorded air temperature, dew point temperature, relative humidity, station pressure, wind speed, headwind, and crosswind speed. Samples were also collected to represent material characteristics. In the second stage, the moisture content of typical

Figure 3—Field sampling setup

Figure 4—Oven drying

samples collected during field measurements was tested in the laboratory. In the third stage, both field and laboratory data were used to generate equations to estimate PM concentrations. Two software packages, SPSS® and Waikato Environment for Knowledge Analysis (WEKA®), were used for data analysis, with PM data as dependent variables and meteorological parameters and laboratory measurement results as independent variables (Figure 5).

The linear stepwise regression approach

The suitability of all data groups collected in the quarries and determined in the laboratory for normal distribution was assessed using the Shapiro-Wilk test, and a decision was made regarding the use of parametric or non-parametric test methods. If the values for skewness and kurtosis are between ±1.5 (Tabachnick, Fidell, 2013; Erbay, Beydoğan, 2017) or, according to another approach, between ±2.0 (George, Mallery, 2010), the data are considered normally distributed. The variance inflation factor (VIF) is used to test for multicollinearity, which indicates the correlation between independent variables. A multicollinearity problem is identified when the VIF is ≥ 10 (Albayrak, 2005; Montgomery et al., 2012; Büyükuysal, Öz, 2016; Karabulut, 2019; Ahmad et al., 2021) or the VIF is ≥ 5 (Bükey, Çetin, 2017; Avcı, 2020; Shrestha, 2020).

The raw PM concentration datasets covered a large area, which prevented them from conforming to a normal distribution. Therefore, they were transformed using inversion (1/), division by 1000 (/1000), natural logarithm (ln), and logarithm (log). The suitability of the transformed datasets for a normal distribution was tested using skewness and kurtosis values. The datasets transformed with the natural logarithm and logarithm conformed to a normal

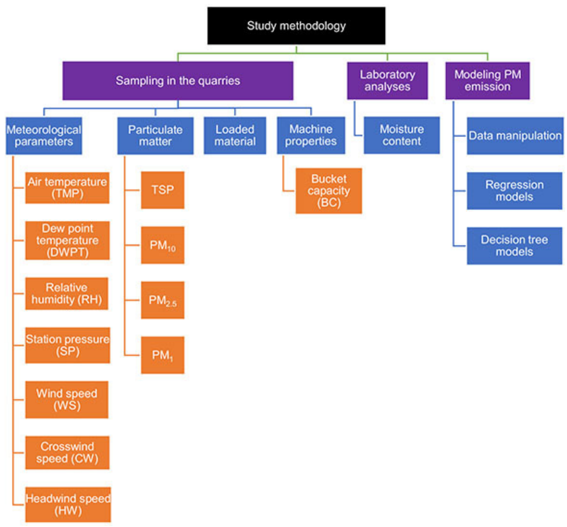


Figure 5—Study methodology

Table 1

Results of normality tests for PM concentration datasets

Dataset type	PM concentration class	Mean ± SD (μg/m³)	Median (min-max)	Skewness	Kurtosis	Sig.
raw	TSP	457.91 ± 967.29	98.40 (1.10 - 6527.90)	3.872	16.928	<0.001
	PM_{10}	308.88 ± 871.79	46.50 (1.00 - 6527.90)	4.862	25.960	<0.001
	$PM_{2.5}$	18.82 ± 39.08	7.12 (0.30 - 540.22)	5.657	43.898	<0.001
	PM_1	3.52 ± 12.64	1.31 (0.01 - 397.47)	18.844	453.374	<0.001
log(raw)	**log(TSP)**	**2.07 ± 0.74**	**1.99 (0.04 - 3.82)**	**0.104**	**-0.303**	**<0.001**
	log(PM_{10})	**1.78 ± 0.72**	**1.67 (0.00 - 3.82)**	**0.572**	**0.180**	**<0.001**
	log($PM_{2.5}$)	**0.89 ± 0.54**	**0.85 (-0.52 - 2.73)**	**0.499**	**0.018**	**<0.001**
	log(PM_1)	**0.19 ± 0.49**	**0.12 (-2.00 - 2.60)**	**0.370**	**1.276**	**<0.001**
1/raw	1/TSP	0.03 ± 0.07	0.01 (0.00 - 0.91)	5.387	37.589	<0.001
	1/PM_{10}	0.05 ± 0.08	0.02 (0.00 – 1.00)	4.366	26.195	<0.001
	1/$PM_{2.5}$	0.24 ± 0.28	0.14 (0.00 - 3.33)	2.583	10.783	<0.001
	1/PM_1	1.23 ± 3.66	0.76 (0.00 – 100.00)	20.654	521.727	<0.001
ln(raw)	**ln(TSP)**	**4.76 ± 1.70**	**4.59 (0.10 - 8.78)**	**0.104**	**-0.303**	**<0.001**
	ln(PM_{10})	**4.11 ± 1.66**	**3.84 (0.00 - 8.78)**	**0.572**	**0.180**	**<0.001**
	ln($PM_{2.5}$)	**2.04 ± 1.24**	**1.96 (-1.20 - 6.29)**	**0.499**	**0.019**	**<0.001**
	ln(PM_1)	**0.44 ± 1.12**	**0.27 (-4.61 - 5.99)**	**0.370**	**1.276**	**<0.001**
raw/1000	TSP/1000	0.46 ± 0.97	0.10 (0.00 - 6.53)	3.872	16.928	<0.001
	PM_{10}/1000	0.31 ± 0.87	0.05 (0.00 - 6.53)	4.862	25.961	<0.001
	$PM_{2.5}$/1000	0.02 ± 0.04	0.01 (0.00 - 0.54)	5.657	43.902	<0.001
	PM_1/1000	0.00 ± 0.01	0.00 (0.00 - 0.40)	18.813	452.124	<0.001

distribution (Table 1).

The decision tree approach

Decision trees are widely used for classifying both numeric and alphanumeric data and are recognised for their high interpretability and understandability (Gargano, Raggad, 1999; Chien, Chen, 2008; Czajkowski, Kretowski, 2010; Onan, 2015; Aydemir et al., 2020). These qualities make them particularly suitable for simplifying complex data structures and improving decision-making processes. Technical advantages such as low cost, fast processing, easy integration into databases, and high reliability have made decision trees among the most preferred classification algorithms (Akçetin, Çelik, 2014; Aydemir et al., 2020). They are widely applied in areas such as knowledge discovery and pattern recognition. Because they produce understandable classification and regression models, decision trees are also extensively used in critical applications such as medical diagnosis and credit risk assessment. Their nonparametric structure enables them to achieve satisfactory accuracy in both classification and regression problems while providing significant interpretability. In regression, decision trees organise independent variables into a hierarchical structure to predict the dependent variable. Starting from the root node, branching is performed through tests on the attributes, and a specific regression equation is presented at each leaf. The estimation process is completed by following the branches based on the results of these attribute tests, starting from the root (Barros et al., 2015; Aksu, 2018).

The M5P algorithm is a Java implementation of the M5 algorithm developed by John Ross Quinlan for the WEKA software (Quinlan, 1992; Witten et al., 2011). This algorithm provides a decision tree-based regression method, generating linear regression models at each leaf node. The tree is constructed inductively, with independent variables split into branches based on the dependent variable (Del Campo-Avila, Luengo, 2011; Njeguš, Štavljanin, 2015). The splitting process stops when low variance is observed among the class values at a node or when the number of samples at the node decreases (Öztürk, 2012; Kara, Şamlı, 2021). Pruning is then performed at each leaf node, and linear regression inequalities are derived from the pruned leaves (Sihag, Kumar, 2021). In the model output, each leaf is presented with two numerical values in parentheses: the first indicates the number of samples in that leaf, and the second shows the ratio of the mean square error of the linear model in the leaf to the global absolute deviation in the entire dataset (Witten et al., 2011; Aydemir, 2018; Altair, 2024; Stackexchange, 2024; WEKA, 2024). This structure combines the explanatory power of a decision tree with the predictive power of linear regression, enabling effective modelling, especially for large and multi-layered datasets. The M5P algorithm provides greater explanatory power and accuracy than classical regression trees, and patterns and correlations in the data can be directly identified through rules and regression equations (Figure 6).

The success of numerical models developed using decision tree algorithms such as DecisionStump, RandomForest, RandomTree, and M5P is evaluated using error criteria including the correlation coefficient (R), coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE) (Gültepe, 2019). Among these criteria, R^2 expresses the model's accuracy and ranges from 0 to 1. When this value approaches 0, it indicates that the model does not fit the data (Çınaroğlu, 2016). Similarly, a high correlation coefficient indicates high predictive success; values less than 0.40 are classified as low correlation, 0.40 – 0.70 as normal correlation, and greater than 0.70 as high correlation, indicating better performance (Aydemir, 2018; Schober et al., 2018; Sabti et al., 2019; Tanni et al., 2020). Values close to zero for MAE and RMSE, the most commonly used error measures, indicate high model performance (Wang, Xu, 2004; Gültepe, 2019), and it is recommended to evaluate these two measures together (Chai,

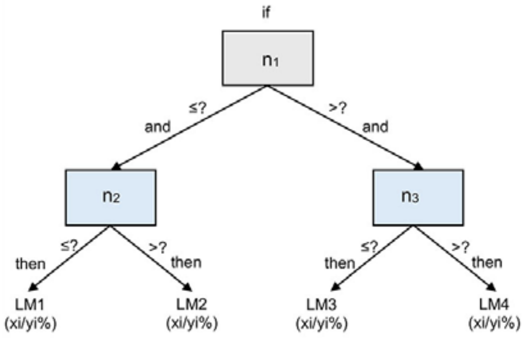


Figure 6—Decision tree algorithm

Draxler, 2014). If MAE is zero, the model is considered to produce excellent results (Aydemir, 2018), and if RMSE is zero, it indicates ideal performance (Çınaroğlu, 2017). MAE is a fundamental accuracy parameter that measures the average magnitude of errors in the prediction results. It averages the differences between actual and predicted values and indicates how close the predictions are to the final results (Usha, Balamurugan, 2016). RMSE, in contrast, tends to be higher than MAE when error magnitudes are large. While MAE generates a linear score by considering all individual differences equally, RMSE calculates the mean error by giving greater weight to large errors. Both metrics range from 0 to ∞ and are used to identify error variation in the predictions (Alsultanny, 2020). In this study, in addition to the model success criteria mentioned in the literature, it was found that having a small number of nodes in the model (< 40) is also crucial for model success.

Three different meta-learning algorithms were applied to examine the relationship between the dependent and independent variables: training set, cross-validation, and percentage split. Both forward addition (FA) and backward elimination (BE) search approaches, as well as standard regression methods, were used for model development. The modelling methods of the decision tree technique are illustrated in Figure 7. In the initial approach, the entire dataset was used for both training and testing. The next experiment used a 5- to 15-fold cross-validation technique, where the data was randomly divided into 5 to 15 subsamples. In each iteration, one subsample was set aside as validation data to evaluate the model, while the remaining k - 1 subsamples were used for training. This cross-validation process was repeated k times, with each subsample serving as the validation data once. For the percentage split method, the dataset was divided into two parts: one for training and one for testing. The training portion used 60% to 70% of the entire dataset, in 1% increments.

Results

Analysis of data recorded from the quarries

During the dry months from July to September, the amount of particulate matter in the air was higher than in October and November. The highest dispersion was recorded in July, and the lowest in November. For all measurements, TSP and PM10 concentrations ranged from 1 μg/m^3 to 6527.9 μg/m^3, $PM_{2.5}$ concentrations from 0.3 μg/m^3 to 540.22 μg/m^3, and PM_1 concentrations from 0.01 μg/m^3 to 397.47 μg/m^3 (Figure 8). The values for meteorological parameters also varied over the months and years (Figure 9). In months with high PM concentrations, relative humidity was low, while air temperature and wind speed were high. The moisture content of the loaded material also differed between the dry and rainy seasons. This variation results from the mineralogical inhomogeneity of the loaded rock, the ability of intervening clay bands to store water, and the effect of water seeping into the subsoil due to the cracked structure of the rock layers. The moisture content of the loaded materials ranged from 0.02% to 21.21%.

Modelling PM concentration using the linear stepwise regression method

The correlation matrix between the normally distributed logarithmic and natural logarithmic dependent variables and the independent variables is shown in Figure 10. The correlation between the dependent and independent variables is low (R ≤ 28%). This indicates that it is very difficult to explain the change in a dependent variable with a single independent variable. There is a low negative correlation (R ≤ 28%) between the moisture of the loaded material and the normally distributed logarithmic and natural logarithmic TSP and PM_{10} concentrations. There is a high positive correlation (R ≥ 70%) between wind speed, crosswind speed, and headwind speed. There is a positive normal correlation

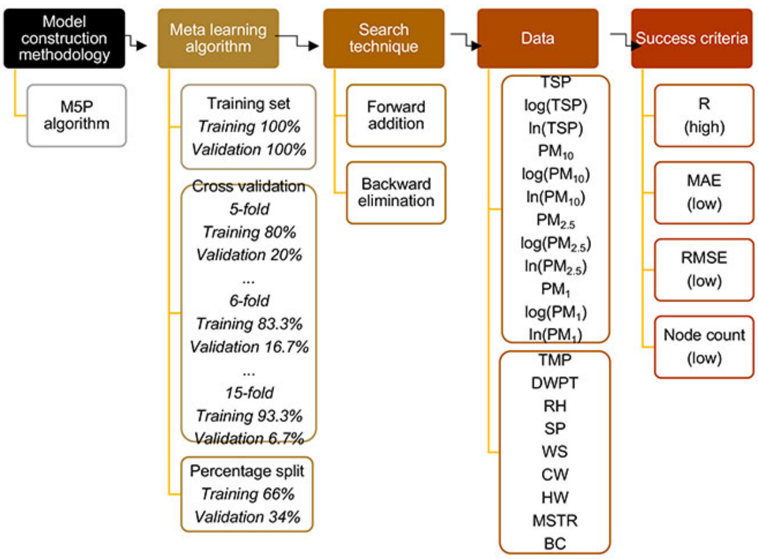


Figure 7—Modelling methodology used in the study

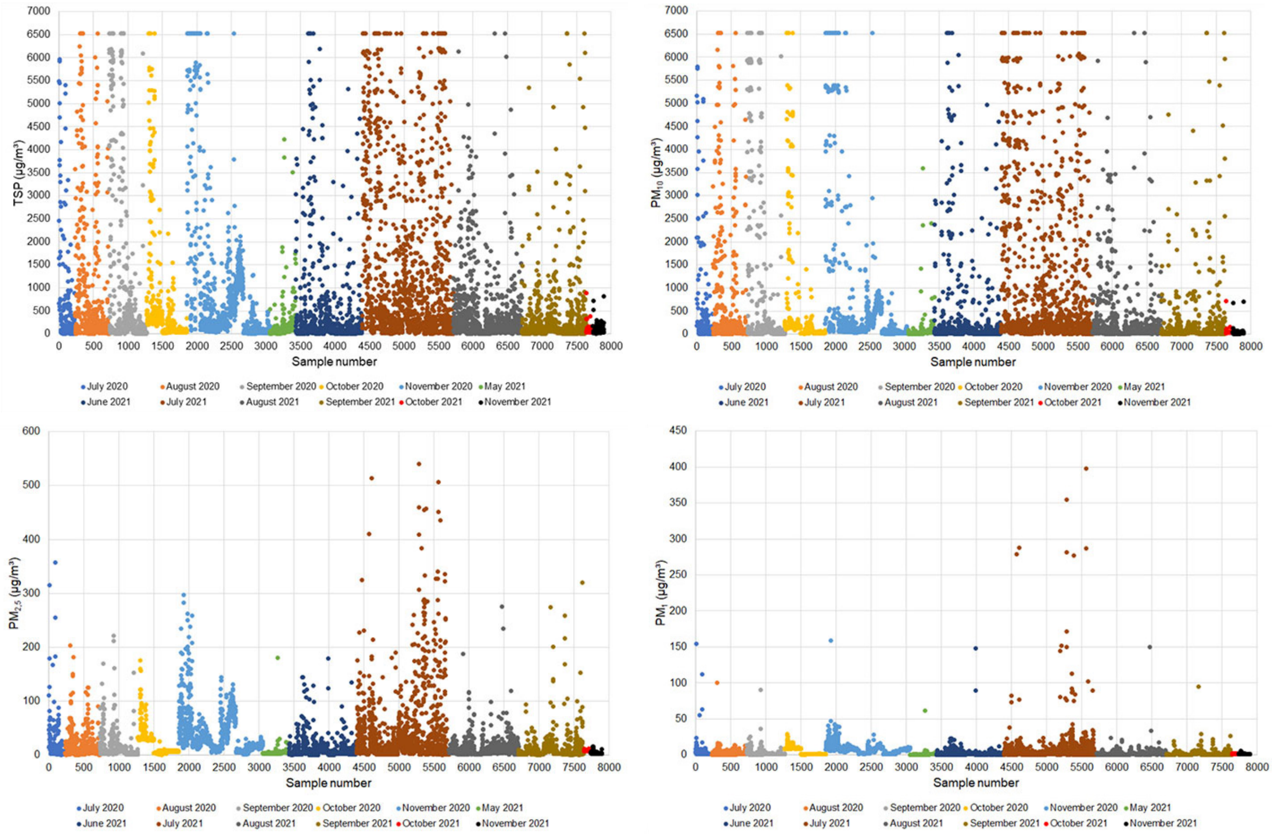


Figure 8—PM concentration by month of sampling

with air temperature and dew point temperature (R < 70%), and a negative normal correlation with relative humidity.

The normally distributed datasets were analysed using stepwise regression with the SPSS® statistical analysis package (George, Mallery, 2010). Several variable selection techniques, including enter, stepwise, and forward selection, were applied during standard regression analyses. However, the coefficient of determination was low in all models ($R^2 < 15\%$) indicating that they could not accurately capture changes in the dependent variable. Additionally, models developed for similar quarries used different independent variables. This indicates that estimating PM concentrations with standard regression models is statistically unreliable. The main reason for this inadequacy is the wide distribution of the dependent variable data. Nevertheless, strong partial correlations were observed in the developed models, particularly with variables such as the moisture content of the loaded material, wind speed (including headwind), and loader bucket capacity. In contrast, the effects of variables such as air temperature, station pressure, and relative humidity were relatively minor.

Table 2 presents the dominant parameters, the explanatory power of the predictive models, and the partially correlated parameters. The regression models for the logarithmic and natural logarithmic PM datasets were significant ($p \leq 0.001$). The independent variables in the regression equations for modelling TSP concentrations are moisture of the loaded material, loader bucket capacity, relative humidity, air temperature, and crosswind and headwind speed. However, these models explain only a small portion of the variation in PM concentrations (adj. $R^2 \leq 12.2\%$). For partial correlation, the most influential variables in the regression models are moisture of the loaded material and headwind speed, while air temperature is the least influential variable.

The PM_{10} regression model is explained by air temperature, relative humidity, moisture of the loaded material, headwind speed, and crosswind speed. The model's significance for explaining changes in particulate matter concentrations is weak (adj. $R^2 \leq 11\%$). The most effective variables in the model are headwind speed and moisture of the loaded material.

The PM2.5 regression model is explained by loader bucket capacity, air temperature, relative humidity, station pressure, wind speed (including headwind), and moisture of the loaded material. The model's significance for explaining changes in particulate matter concentrations is also weak (adj. $R^2 \leq 7.1\%$). The most effective variables in the model are loader bucket capacity and headwind speed.

The regression analysis showed that the release of PM1 depends on the moisture of the loaded material, air temperature, relative humidity, wind speed (including headwind), and the loader's bucket capacity. However, the independent variables were not sufficient to explain PM release (adj. $R^2 < 13\%$). Wind speed and loader bucket capacity had the highest partial correlation with PM concentration, while relative humidity had the lowest correlation.

Modelling PM concentration using decision tree approach

As the explanatory power of traditional linear regression models was low, the decision tree technique was used for further modelling. In addition to the original data, transformed datasets were also used in the regression analyses conducted with the M5P decision tree method. The models using raw datasets provided coefficients for TSP and PM_{10}, while the models with logarithmic datasets yielded coefficients for $PM_{2.5}$ and PM_1. Regardless of the meta-learning algorithm used, the decision tree equations that best predicted the concentration for each PM size shared the same

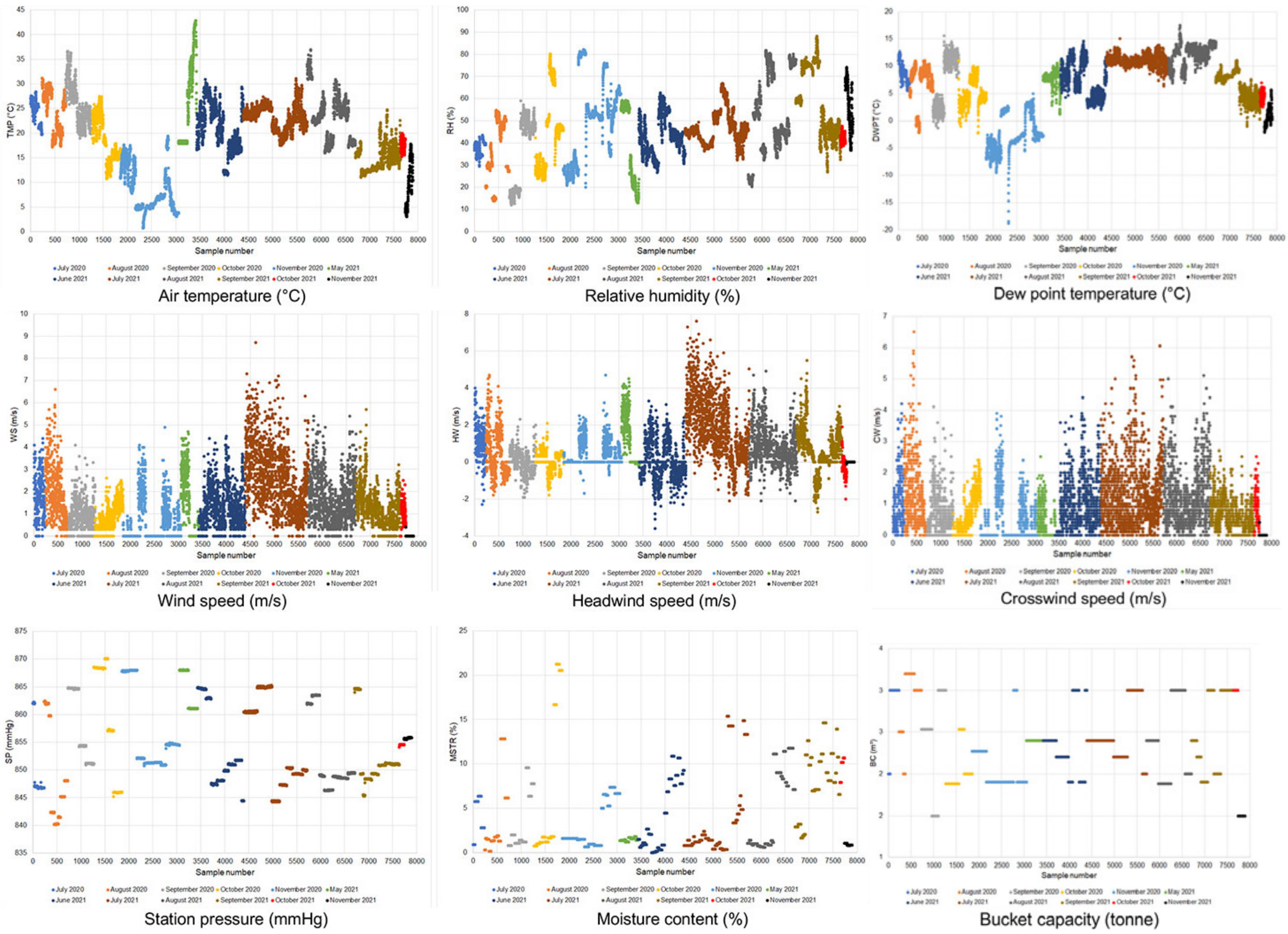


Figure 9—Meteorological and material parameters by month of sampling

characteristics: 5dew point temperature, wind speed (including headwind and crosswind), loader bucket capacity, and moisture content of the loaded material. Decision trees organised variables by importance, starting from the root node, and generated regression equations for data subsets that met specific conditions in each branch. This approach enabled flexible and reliable predictions of PM concentrations by considering different variable combinations. Additionally, regression equations based on unique parameter combinations for each subset allowed detailed analysis of interactions between variables. The statistical significance of all

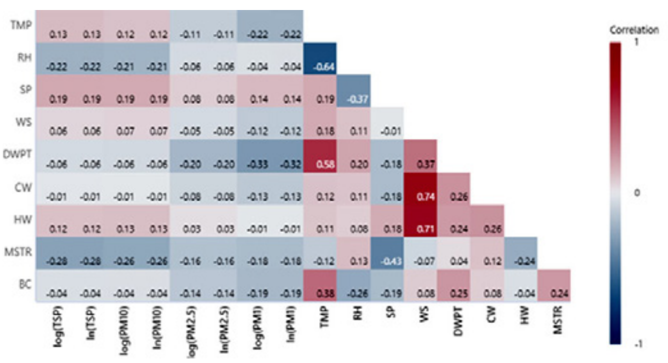


Figure 10—Correlation matrix of transformed dependent and independent variables

independent variables improved the overall predictive performance of the models.

The M5P model tree developed for predicting TSP concentration and its associated equations is shown in Figure 11. The model has a top-down tree structure with 22 terminal nodes, each containing a regression equation. The first split in the tree is based on station pressure, the most important variable (SP ≤ 858.5 and SP > 858.5), followed by secondary variables such as air temperature and relative humidity. At lower levels, the moisture content of the loaded material and the bucket capacity are included. Each regression equation in the model is calculated using different independent variables. For example, to use the LM22 regression equation, SP must be greater than 858.5 and RH must be greater than 47.95.

The M5P model tree and prediction equations for PM_{10} concentration are shown in Figure 12. This model has 21 terminal nodes, each with a regression equation. Similar to the TSP model, the first split is based on station pressure (SP ≤ 858.5 and SP > 858.5), followed by air temperature and relative humidity. Variables such as the moisture content of the loaded material and bucket capacity play important roles in the lower branches.

The $PM_{2.5}$ concentration prediction model has 25 terminal nodes, each represented by its own regression equation (Figure 13). Unlike other models, the first split in the decision tree is based on dew point temperature (DWPT ≤ 2.75 and DWPT > 2.75). Subsequent levels include station pressure and relative humidity,

Table 2

Independent variables with the highest and lowest correlations with dependent variables in the stepwise linear regression models

Dependent variable	Stepwise linear regression			
	Independent variables having a major impact on the model (p <0.05; VIF ≤5)	Adjusted R^2 (%)	Partial correlation	
			Highest	Lowest
ln(TSP) / log(TSP)	HW, MSTR, BC, CW, RH, TMP	12.2	HW, MSTR	TMP
ln(PM_{10}) / log(PM_{10})	HW, MSTR, CW, RH, TMP	11.0		
ln($PM_{2.5}$) / log($PM_{2.5}$)	BC, HW, WS, TMP, MSTR, RH, SP	7.1	BC, HW	SP
ln(PM_1) / log(PM_1)	WS, BC, HW, TMP, MSTR, RH	12.7	WS, BC	RH

Table 3

Regression models developed with the decision tree models

Algorithm	Criteria	TSP	PM_{10}	log($PM_{2.5}$)	log(PM_1)
Training set	Correlation coefficient	0.5958	0.5711	0.6523	0.7109
	MAE	705.0799	624.0791	0.3126	0.1908
	RMSE	1206.4865	1176.8853	0.4141	0.2879
	RAE (%)	72.9091	73.0528	69.6594	56.0129
	RRSE (%)	80.3352	82.1016	75.819	70.3418
	Number of nodes	22	21	25	23
	Leaf size	350	390	390	350
	Search technique	FA	FA	BE	BE
Cross validation	Correlation coefficient	**0.5848**	**0.5524**	**0.6311**	**0.7027**
	MAE	712.2343	633.0651	0.3191	0.1941
	RMSE	1218.4169	1194.9823	0.4237	0.2913
	RAE (%)	73.6248	74.0844	71.0847	56.9541
	RRSE (%)	81.106	83.3403	77.5596	71.1416
	Number of nodes	22	21	25	23
	Leaf size	350	390	390	350
	Search technique	FA	FA	BE	BE
Percentage split	Correlation coefficient	0.5666	0.5370	0.6252	0.6977
	MAE	746.7993	657.9597	0.3265	0.2016
	RMSE	1246.8336	1219.7655	0.4308	0.2966
	RAE (%)	77.4732	76.6544	71.97	58.7606
	RRSE (%)	82.4524	84.3925	78.2057	71.7399
	Number of nodes	22	21	25	23
	Leaf size	350	390	390	350
	Search technique	FA	FA	BE	BE

while lower branches consider headwind speed, moisture content of the loaded material, and bucket capacity.

The M5P model tree developed for PM_1 concentration estimation has 23 terminal nodes, each with a unique regression equation (Figure 14). Unlike other models, the initial split in this tree is based on bucket capacity (BC ≤ 2.335 and BC > 2.335). Subsequent variables incorporated into the model include station pressure, relative humidity, crosswind speed, and the moisture content of the loaded material.

The equations with the highest number of samples used in the nodes for TSP, PM_{10}, $PM_{2.5}$, and PM_1 concentrations are LM14, LM13, LM21, and LM23, respectively.

These M5P decision tree models allow for the estimation of PM concentrations in similar gypsum and limestone fields. During the estimation process, PM concentrations can be predicted with reasonable accuracy by measuring the independent variables specified in the decision trees and applying the corresponding regression equations according to the tree structure.

The data analysis showed that when the PM concentration data were dispersed, the training set model had a slight advantage of about 5% over the other two approaches. However, when the data followed a similar pattern, the representative capabilities of all techniques were comparable. The final particulate matter concentration models developed, using the three different test methods, had the same number of nodes, the same variable addition methods for regression analysis, and the same minimum sample requirements for the nodes.

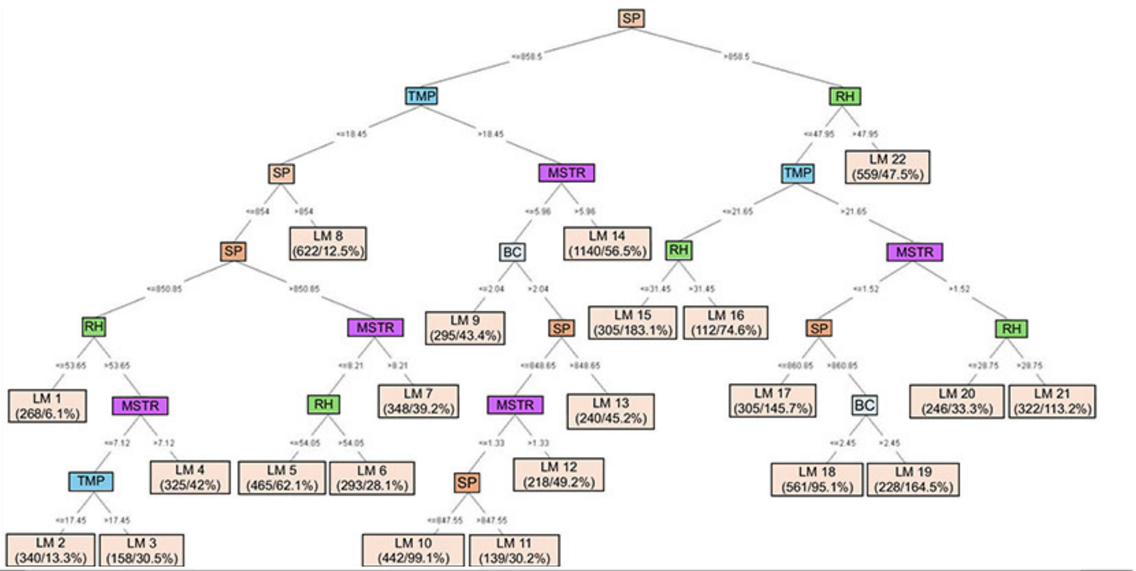
The PM concentration models developed using the decision tree technique did not show significant changes in MAE or RMSE over the course of iterations. Therefore, the number of nodes in the decision tree and the correlation coefficient were identified as crucial factors for accurately modelling PM concentrations. The success of the models was influenced by the meta-learning technique, the search method (regression analysis variable addition method), and the minimum number of samples allowed in a node. The final prediction equations were derived from the data sets with the highest correlation coefficient and the fewest nodes. The correlation coefficients obtained using the training data set were higher than those from the other two test methods.

Discussion

No previous studies have used decision tree algorithms to model PM concentrations resulting from loading, which increases the originality and contribution of this study to the literature. We conducted experiments with various decision tree algorithms and artificial neural networks (ANNs), including M5P, RandomForest, RandomTree, REPTree, and DecisionStump, during the modelling process. It should be noted that RandomForest, RandomTree, and REPTree do not provide reference equations, and artificial neural networks do not generate model equations or allow parameter estimation (Yazıcı et al., 2007; Duran, 2022).

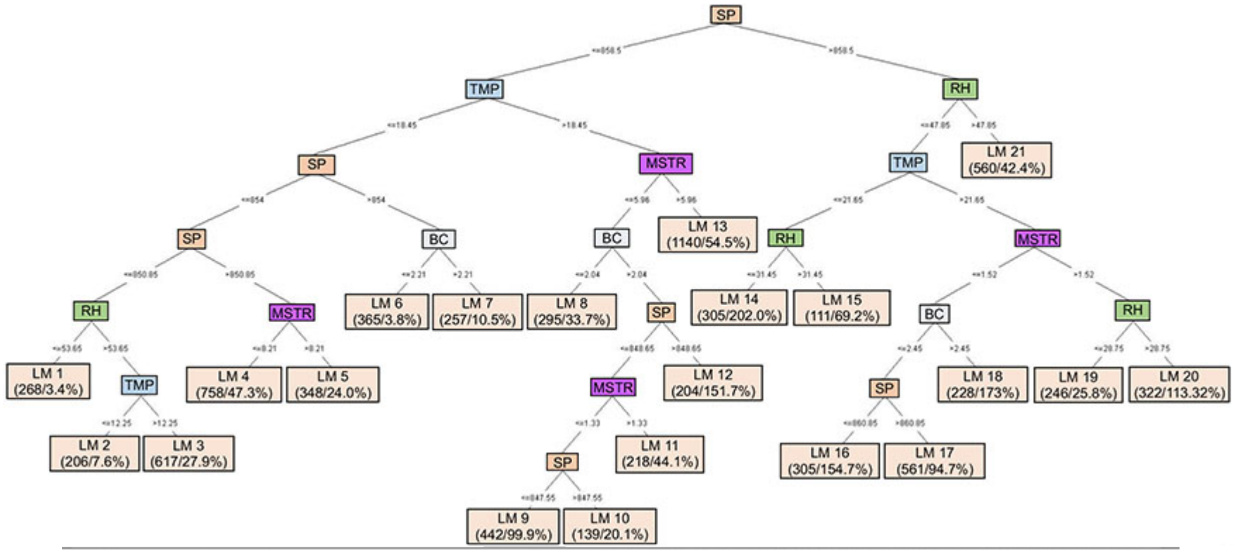
Artificial neural networks (ANNs) have three basic layers: the input layer, where data enter the network; the output layer, which generates the response after processing; and the hidden layer(s), which transform and interpret data during learning (El-Shahat, 2018). Although artificial intelligence, machine learning, and deep learning share logical similarities, they also differ functionally. Artificial intelligence makes decisions based on pre-taught data, while machine learning produces results from provided data and its inferences. Deep learning, a subfield of supervised learning, uses artificial neural network algorithms and is structured based on inspiration from brain function (Copeland, 2016; Akın, Şahin, 2024). They can continuously generate feature values, make inferences based on these values, and iteratively learn complex data structures. These structures, which can mimic the observation, analysis, learning, and decision-making behaviours of living organisms, can extract features from large amounts of data and perform operations such as transformation and classification with high accuracy (Deng, Yu, 2014; Kayaalp, Süzen, 2018; Akın, Şahin, 2024). However, advanced algorithms such as ANN and SVR conceal their prediction equations due to their internal structure (Henedy et al., 2022). Therefore, algorithms other than the M5P algorithm were not evaluated in this study because they did not meet the main objective.

Determining the levels of particulate matter (PM) to which workers in the mining industry are exposed is critical due to the health risks posed by PM. The emission prediction equations and exposure parameters from previous studies on coal and iron mines are presented in Table 4. Previous studies have developed equations to predict TSP release using raw data sets. These studies used lb/t (Axetell and Cowherd, 1984; USEPA, 1998a), kg/t (USEPA, 1991; USEPA, 1995; USEPA, 2006; NPI, 2012), and g/s (Chakraborty et al., 2002; Chaulya, 2006; Lal, Tripathy, 2012) as units of measurement. In this study, however, the prediction equations are expressed in μg/m^3, indicating the PM concentration in the environment during loading activities. Several parameters were considered when estimating TSP released during coal extraction in open coal mines. These included the moisture content of the loaded material (Axetell, Cowherd, 1984; USEPA, 1991; USEPA, 1995; USEPA, 1998a; NPI, 2012), as well as the moisture and silt content of the loaded material, unloading height, wind speed, loading frequency, and bucket capacity (Chakraborty et al., 2002; Lal, Tripathy, 2012). When modelling TSP release during overburden



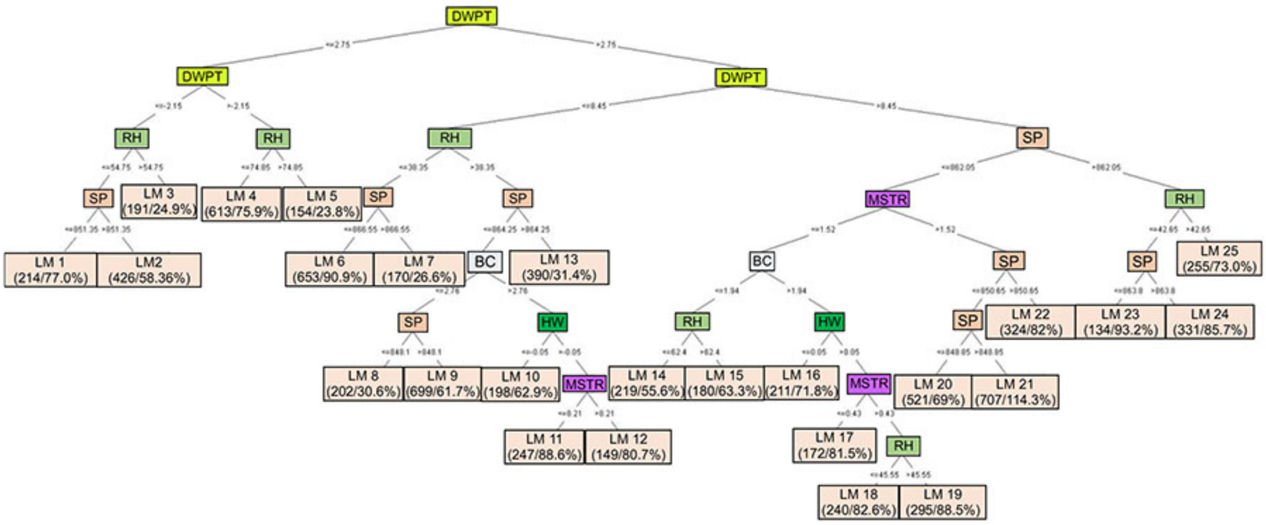
Node	TSP concentration equation	Node	TSP concentration equation
LM1	-0.2868*TMP+0.0621*RH+0.028*SP-0.7710*MSTR+2.0618*BC+68.963	LM12	0.5144*TMP+2.4941*RH-3.5653*SP-4.3414*MSTR+18.4208*BC+3240.207
LM2	0.5872*TMP-0.1082*RH+0.028*SP-0.4316*MSTR+2.0618*BC+111.450	LM13	0.5144*TMP+2.0966*RH+0.8572*SP+6.0348*MSTR+18.4208*BC+930.664
LM3	1.2512*TMP-0.1082*RH+0.028*SP-0.4316*MSTR+2.0618*BC+233.252	**LM14**	**0.5812*TMP+0.1727*RH-0.0509*SP-1.0835*MSTR+4.6335*BC+369.838**
LM4	0.0797*TMP-0.1082*RH+0.028*SP-0.2589*MSTR+2.0618*BC+311.986	LM15	-6.5873*TMP-10.7114*RH-6.7073*SP-17.0412*MSTR+21.2611*BC+9715.175
LM5	-0.2845*TMP+23.1546*RH+0.02508*SP-2.1533*MSTR+2.062*BC-344.408	LM16	-6.5873*TMP-23.6772*RH-6.7073*SP-17.0412*MSTR+21.2611*BC+7636.839
LM6	-0.2845*TMP-1.2473*RH+0.02508*SP-2.1533*MSTR+2.0618*BC+345.703	LM17	-4.4391*TMP-0.9962*RH-10.314*SP-15.6130*MSTR+14.8096*BC+11595.647
LM7	-0.2845*TMP-0.8588*RH+0.02508*SP-3.7188*MSTR+2.0618*BC+203.724	LM18	-27.348*TMP-0.9962*RH-6.560*SP-15.6130*MSTR-1282.5034*BC+10308.776
LM8	-0.45*TMP-0.2048*RH+39.223*SP-0.8486*MSTR+62.665*BC-33536.0615	LM19	-1.8587*TMP-0.9962*RH-6.5600*SP-15.6130*MSTR+44.4593*BC+8219.331
LM9	0.5144*TMP+1.2630*RH+2.9313*SP-4.3330*MSTR+50.8640*BC-2164.785	LM20	-3.9604*TMP+1.3922*RH-5.2313*SP-23.5313*MSTR+22.5761*BC+4908.625
LM10	0.5144*TMP+37.0223*RH-5.7344*SP-1.6048*MSTR+18.4208*BC+4379.66	LM21	-3.9604*TMP+0.8536*RH-5.2313*SP-23.5313*MSTR+22.5761*BC+5675.488
LM11	0.5144*TMP+4.306*RH-15.302*SP-1.6048*MSTR+18.4208*BC+13091.316	LM22	-4.394*TMP-1.8366*RH-126.8015*SP-11.2437*MSTR+0.976*BC+110353.222

Figure 11—Decision tree model and predictive equations for TSP concentrations



de	PM_{10} concentration equation		
l1	0.302*TMP-0.034*RH+0.826*SP+0.552*HW-0.451*MSTR+10.358*BC-672.86	LM11	0.410*TMP+2.246*RH-2.670*SP+1.726*HW-1.757*MSTR+14.25*BC+2367.51
l2	0.704*TMP-0.09*RH+1.118*SP+0.552*HW-0.451*MSTR+15.970*BC-918.70	LM12	0.410*TMP+1.910*RH+0.538*SP+1.726*HW+6.908*MSTR+14.25*BC+863.36
l3	0.266*TMP-0.09*RH+0.620*SP+0.552*HW-0.451*MSTR+138.493*BC-672.93	**LM13**	**0.461*TMP+0.128*RH+0.013*SP+1.881*HW-0.727*MSTR+4.397*BC+211.89**
l4	-0.107*TMP-9.584*RH+0.109*SP+0.552*HW-1.079*MSTR+3.009*BC+796.96	LM14	-6.36*TMP-10.527*RH-5.93*SP+0.574*HW-14.91*MSTR+30.533*BC+8736.13
l5	-0.107*TMP-0.461*RH+0.109*SP+0.552*HW-1.791*MSTR+3.009*BC+29.54	LM15	-6.36*TMP-23.404*RH-5.93*SP+0.574*HW-14.91*MSTR+30.533*BC+6683.64
l6	-0.182*TMP-0.130*RH-0.251*SP+0.552*HW-0.490*MSTR+4.713*BC+262.35	LM16	-9.045*TMP-0.98*RH-17.104*SP+0.574*HW-13.98*MSTR-41.6*BC+17318.40
l7	-0.182*TMP-0.139*RH-0.251*SP+0.552*HW-0.490*MSTR+5.194*BC+304.14	LM17	-30.84*TMP-0.98*RH-11.846*SP+0.574*HW-13.98*MSTR-1090.84*BC+14277.57
l8	0.41*TMP+1.107*RH+1.737*SP+1.726*HW-0.669*MSTR+37.641*BC-1336.85	LM18	-4.606*TMP-0.976*RH-9.975*SP+0.574*HW-13.980*MSTR+58.170*BC+10964.25
l9	0.41*TMP+33.467*RH-4.567*SP+1.726*HW+0.147*MSTR+14.25*BC+3289.97	LM19	-3.681*TMP+1.154*RH-4.456*SP+0.574*HW-21.212*MSTR+31.272*BC+4076.92
l10	0.41*TMP+3.857*RH-12.145*SP+1.726*HW+0.147*MSTR+14.25*BC+10291.04	LM20	-3.681*TMP+0.673*RH-4.456*SP+0.574*HW-21.212*MSTR+31.272*BC+4768.59
		LM21	-4.177*TMP-1.750*RH-86.760*SP+0.574*HW-9.934*MSTR+12.044*BC+75519.72

Figure 12—Decision tree model and predictive equations for PM10 concentrations

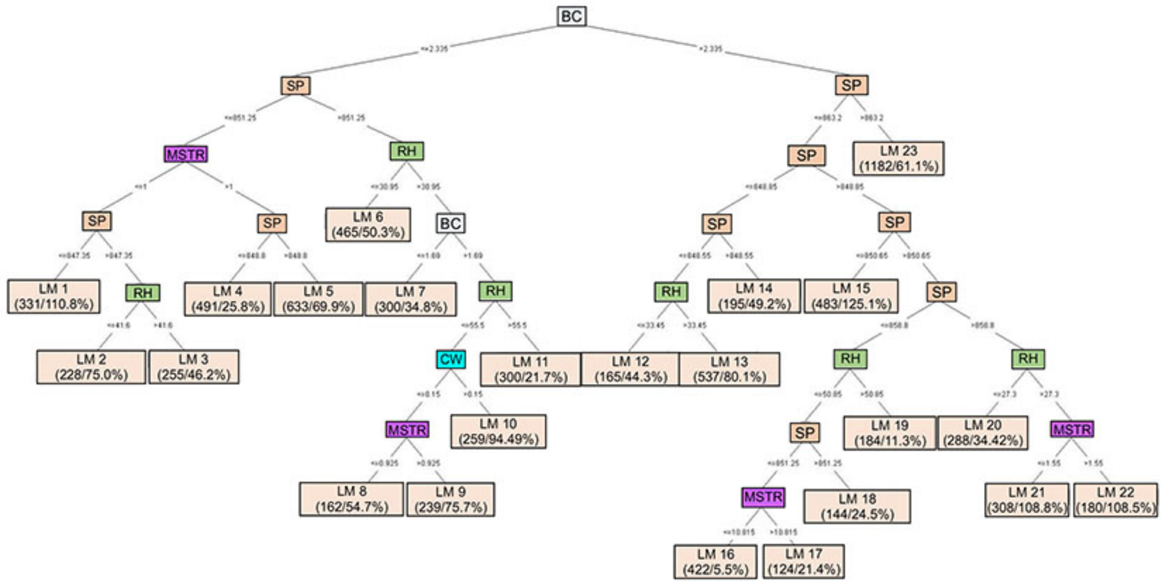


Node	$\log(PM_{2.5})$ concentration equation		
LM1	0.00015*RH+0.0012*SP-0.0002*WS-0.0011*DWPT+0.0003*HW-0.0065*MSTR-0.0008*BC+0.40	LM13	-0.0001*RH+0.04*SP+0.00008*WS+0.00008*DWPT+0.0004*HW+0.0002*MSTR+0.001*BC-34.81
LM2	0.00015*RH+0.0009*SP-0.0002*WS-0.001*DWPT+0.0003*HW-0.11*MSTR-0.0008*BC+1.0	LM14	-0.00115*RH-0.00014*SP+0.0015*WS+0.00005*DWPT+0.001*HW-0.0058*MSTR+0.01*BC+1.16
LM3	0.00015*RH+0.0012*SP-0.0002*WS-0.001*DWPT+0.0003*HW-0.0044*MSTR-0.0008*BC+0.18	LM15	-0.00132*RH-0.00014*SP+0.0015*WS+0.00005*DWPT+0.001*HW-0.0058*MSTR+0.01*BC+0.75
LM4	0.00042*RH+0.03011*SP-0.0002*WS-0.0012*DWPT+0.00034*HW-0.00086*MSTR-0.06*BC-24.86	LM16	0.0003*RH-0.00014*SP+0.0015*WS+0.00005*DWPT+0.0017*HW+0.004*MSTR+0.004*BC+1.15
LM5	0.0011*RH+0.00345*SP-0.0002*WS-0.0012*DWPT+0.00034*HW-0.00086*MSTR+0.0063*BC-1.62	LM17	0.0002*RH-0.00014*SP+0.0015*WS+0.00005*DWPT+0.0008*HW-0.0007*MSTR+0.004*BC+1.57
LM6	-0.0002*RH+0.0006*SP+0.00008*WS+0.00008*DWPT+0.13*HW+0.00005*MSTR+0.12*BC-0.137	LM18	-0.0002*RH-0.00014*SP+0.0015*WS+0.00005*DWPT+0.0008*HW-0.0007*MSTR+0.004*BC+1.48
LM7	-0.0002*RH+0.002*SP+0.00008*WS+0.00008*DWPT+0.0077*HW+0.00005*MSTR-0.017*BC-0.24	LM19	-0.0002*RH-0.00014*SP+0.0015*WS+0.00005*DWPT+0.0008*HW-0.0007*MSTR+0.004*BC+1.24
LM8	-0.0001*RH-0.001*SP+0.00008*WS+0.00008*DWPT+0.0004*HW+0.00017*MSTR+0.0006*BC+1.76	LM20	0.00002*RH+0.0015*SP-0.001*WS+0.00005*DWPT-0.0001*HW+0.0074*MSTR-0.0001*BC-0.48
LM9	-0.0001*RH-0.0092*SP+0.00008*WS+0.00008*DWPT+0.0004*HW-0.0223*MSTR+0.0006*BC+8.66	**LM21**	**-0.0021*RH+0.001*SP-0.16*WS+0.00005*DWPT-0.0001*HW-0.0007*MSTR-0.00011*BC+0.60**
LM10	-0.0001*RH-0.00007*SP+0.00008*WS+0.00008*DWPT-0.003*HW+0.0002*MSTR+0.0012*BC+0.85	LM22	0.00004*RH-0.0004*SP+0.0014*WS+0.00005*DWPT-0.0001*HW-0.0003*MSTR-0.0001*BC+0.9
LM11	-0.0001*RH-0.00007*SP+0.00008*WS+0.00008*DWPT-0.0013*HW-0.0007*MSTR+0.0012*BC+0.82	LM23	-0.0004*RH-0.019*SP+0.002*WS+0.00005*DWPT-0.0003*HW-0.00035*MSTR-0.0005*BC+17.43
LM12	-0.0001*RH-0.00007*SP+0.00008*WS+0.00008*DWPT-0.0013*HW-0.0012*MSTR+0.0012*BC+0.58	LM24	-0.0004*RH-0.01*SP+0.002*WS+0.00005*DWPT-0.0003*HW-0.00035*MSTR-0.0005*BC+9.36
		LM25	-0.0007*RH-0.0055*SP+0.002*WS+0.00005*DWPT-0.0003*HW-0.00035*MSTR-0.0005*BC+5.42

Figure 13—Decision tree model and predictive equations for log(PM2.5) concentrations

removal in coal mines, factors such as the moisture and silt content of the loaded material, unloading height, wind speed, loading frequency, and bucket capacity were included (Chakraborty et al., 2002; Lal, Tripathy, 2012). Moisture content and wind speed were also considered (NPI, 2012).

TSP released during ore extraction and overburden removal in open pit iron mining was modelled using parameters including the moisture and silt content of the loaded material, unloading height, wind speed, loading frequency, and bucket capacity (Chaulya, 2006). Similarly, aggregate loading in quarries has been modelled using the moisture content of the material and wind speed (USEPA, 2006). While previous studies have primarily focused on modelling TSP release from open pit iron and coal mines, this study specifically examines TSP concentrations from gypsum and limestone quarries. The study identifies several effective independent variables for modelling TSP concentrations, including air temperature, dew point temperature, station pressure, relative humidity, wind speed (including headwind and crosswind),



Node	log(PM_1) concentration equation		
LM1	0.00021*RH-0.00076*SP-0.00041*CW-0.0004*MSTR+0.004*BC+1.1318	LM12	0.0004*RH+0.00001*SP+0.00073*CW-0.00004*MSTR-0.0005*BC+0.054
LM2	0.00054*RH-0.00053*SP-0.00041*CW-0.0004*MSTR+0.004*BC+0.631	LM13	0.00021*RH+0.00001*SP+0.00073*CW-0.00004*MSTR-0.0005*BC+0.180
LM3	0.0005*RH-0.00053*SP-0.00041*CW-0.0004*MSTR+0.004*BC+0.787	LM14	0.00026*RH-0.00023*SP+0.00073*CW-0.00004*MSTR-0.0005*BC+0.302
LM4	-0.0034*RH+0.00036*SP-0.00041*CW-0.00033*MSTR+0.004*BC-0.0073	LM15	0.0001*RH-0.00012*SP+0.00318*CW-0.00033*MSTR-0.0005*BC+0.570
LM5	0.00002*RH-0.15761*SP-0.00041*CW-0.00033*MSTR+0.004*BC+134.182	LM16	0.000004*RH+0.00083*SP-0.0284*CW-0.0006*MSTR-0.0005*BC-0.573
LM6	0.0002*RH+0.074*SP-0.00412*CW-0.00124*MSTR+0.042*BC-63.324	LM17	0.000004*RH+0.00083*SP-0.00018*CW-0.00137*MSTR-0.0005*BC-0.656
LM7	0.00073*RH+0.00024*SP-0.0076*CW-0.00165*MSTR+0.079*BC-0.189	LM18	0.000004*RH+0.00181*SP+0.00224*CW-0.00065*MSTR-0.0005*BC-1.386
LM8	0.00057*RH+0.00024*SP-0.0149*CW-0.0065*MSTR+0.038*BC+0.611	LM19	-0.00022*RH+0.00088*SP+0.00224*CW-0.00012*MSTR-0.0005*BC-0.704
LM9	0.00057*RH+0.00024*SP-0.0149*CW-0.0054*MSTR+0.038*BC+0.364	LM20	0.00093*RH+0.00036*SP+0.00241*CW-0.00012*MSTR-0.0005*BC-0.252
LM10	0.00057*RH+0.00024*SP-0.0189*CW-0.0038*MSTR+0.038*BC-0.011	LM21	0.0006*RH+0.00036*SP+0.00241*CW-0.00427*MSTR-0.0005*BC+0.204
LM11	0.0009*RH+0.00024*SP-0.0111*CW-0.0024*MSTR+0.038*BC+0.560	LM22	0.0006*RH+0.00036*SP+0.00241*CW-0.0067*MSTR-0.0005*BC-0.01
		LM23	**-0.00604*RH-0.01817*SP-0.0430*CW-0.00004*MSTR-0.0005*BC+16.129**

Figure 14. Decision tree model and predictive equations for log(PM_1) concentrations

moisture content of the loaded material, and loader bucket capacity. Previous emission models have traditionally ignored meteorological conditions. However, this study demonstrates that these conditions also play an important role in modelling particulate matter concentrations.

Previous studies have limited PM_{10} emission modelling of loading activities to coal and aggregate mines. Certain parameters were used to predict PM_{10} levels during coal mining activities, including $0.75 \times PM_{15}$ (USEPA, 1991; USEPA, 1995), moisture content of loaded material (NPI, 2012), and both moisture content of loaded material and wind speed during overburden removal in coal mines (NPI, 2012). PM_{10} was estimated as a fraction of TSP (35%) in aggregate quarry loading operations (USEPA, 2006). In this study, PM_{10} concentrations were modelled using several parameters, including air temperature, dew point temperature, station pressure, relative humidity, wind speed (including headwind and crosswind), moisture content of loaded material, and loader bucket capacity.

To the best of the authors' knowledge, no academic study has been published on modelling PM_1 concentration in open pit mining. However, there are two different formulas for calculating $PM_{2.5}$ emissions generated during loading in open pit coal mining and aggregate mining. The formula for $PM_{2.5}$ emissions from coal mining is $0.019 \times$ TSP (Axetell, Cowherd, 1984; USEPA, 1991; USEPA, 1995; USEPA, 1998a; NPI, 2012), while the formula for aggregate loading is $0.053 \times$ TSP (USEPA, 2006). In this study, independent formulas were developed to predict the concentrations of $PM_{2.5}$ and PM_1.

The PM concentration models include key parameters such as air temperature, dew point temperature, station pressure, relative humidity, wind speed (including headwind and crosswind components), moisture content of the loaded material, and loader bucket capacity.

Some prediction models in this study use logarithmic values of dependent variables, assuming the regressors are non-random and the error terms are independent and identically distributed. When the log-based fitted values of the dependent variables are directly back-transformed, it is assumed that the errors in the log model are normally distributed. However, if this assumption does not hold, this approach may produce biased predictions. To address this, Duan (1983) proposed a nonparametric method called the smearing estimate, which does not require any specific assumption about the distribution of the regression errors. In this method, a smearing estimate is applied for the reverse transformation to obtain a prediction of the original variable after a transformation has been used for estimation. It is recommended to follow this methodology when using the logarithm-based estimation formulas generated in this study.

Previous research has shown that open-cast coal mining releases significant amounts of dust into the workplace. Emission estimating equations are site-specific, and factors developed for one site may not yield accurate results at another. Therefore, it is essential to adjust emission factors for each location (Ghose, 2004; Huertas et al., 2012; Papagiannis et al., 2014; Richardson et al., 2019). In this study, the emission estimation equations differed, even though the parameters influencing them were similar in the quarries examined. These findings are consistent with the existing literature.

Drawing the conclusion that a single independent variable affects the PM concentrations model is very challenging, primarily due to significant fluctuations in the dependent variable. The decision tree models developed using the training set methodology show a higher correlation coefficient than the other two methods, with similar outcomes when the PM concentrations data are widely distributed. In contrast, all three test methods yield similar correlation coefficients for narrower data ranges. The final PM concentration models produced by the three testing methods include the same number of nodes, use the same methods for inserting variables into regression analysis, and have the same minimum number of samples for each node. Because the mean absolute error (MAE) and root mean square error (RMSE) in PM concentration models varied minimally, the correlation coefficient and number of nodes were the most relevant predictors of model success. The meta-learning strategy, the search method, and the minimum number of samples permitted in a node all affected the performance of the PM concentration models.

Limitations

Samples representing material properties were collected at regular intervals of 5 minutes per kilogram. The moisture content test was performed on the same day because the characteristics of the samples could change by the next day. To achieve more consistent results, laboratory staff and resources should allow for the simultaneous examination of multiple moisture samples.

This study used a PM meter and a weather meter to measure PM concentrations during loading. Simultaneous use of these instruments from various directions is likely to yield more precise results in assessing PM concentrations.

Conclusions and recommendations

This study examined particulate matter levels and meteorological parameters during loading operations at gypsum and limestone quarries near Sivas, Türkiye. The researchers recorded the characteristics of the mining equipment and periodically collected samples to analyse the properties and moisture content of the loaded materials in the laboratory. They developed separate predictive models for TSP, PM_{10}, PM2.5, and PM1 concentrations using two different software packages (SPSS and WEKA): one for statistical analysis and one for data mining.

In the statistical analyses, the original data sets were generally not suitable for a normal distribution. The researchers applied various mathematical transformations, including the natural logarithm, logarithm, reciprocal, and thousandths, to the data. The data sets transformed using the natural logarithm and logarithm conformed to a normal distribution. Stepwise regression analysis was used to create representative models with normally distributed data sets, which demonstrated a modest ability (adj. $R^2 < 15\%$) to predict particulate matter concentrations. The researchers then used the decision tree approach with both raw and transformed data sets to develop PM prediction equations. The final models included equations with the highest correlation coefficients and the fewest nodes for all PM size categories. In this respect, the M5P algorithm stands out as an effective and powerful tool, particularly for developing decision support systems based on field data.

The release of particulate matter is a complex phenomenon influenced by many factors. No single variable can fully explain the significant variations observed in PM levels. Meteorological factors such as air temperature, relative humidity, dew point temperature, wind speed (including headwind and crosswind), and station pressure all contribute to the amount of PM emitted. Equipment characteristics, loader bucket capacity, and the moisture content of the loaded material also affect PM concentrations.

Existing research on dust generation in mining has primarily focused on coal mines and personal dust exposure. However, there is limited research on particulate matter released from machinery and equipment in non-coal open pit mines. More comprehensive studies are needed to quantify particulate matter concentrations from all mining operations. This will provide a clearer understanding of the impact of these concentrations on workers and the surrounding environment.

Table 4

PM release prediction equations for loading activity in open-pit mining operations (Duran et al., 2021)

Author	Mine	PM release prediction equation				Unit
		TSP	PM_{15}	PM_{10}	$PM_{2,5}$	
Chakraborty et al., 2002; Lal, Tripathy, 2012	Coal (O/B)	$E = [0.018*\{(100 - M)/M\}^{1.4} * [S/(100 - S^{0.4}] * (U * H * X * L)^{0.1}]$				g/s
NPI, 2012	Coal (O/B)	$E = 0.74 * 0.0016 * (U/2.2)^{1.3} * (M/2.0)$-1.4		$E = 0.35 * 0.0016 * (U/2.2)^{1.3} * (M/2.0)^{-1.4}$		kg/ton
Chaulya, 2006	Iron (Waste)	$E = [0.0081 * [(100 - M)/M\}^{1.4} *\{S/(100 - S)^{0.3}] * U^{1.1}* (H * X * L)^{0.1}$				g/s
USEPA, 2006	Gravel	$E = 0.74 * 0.0016 * (U/2.2)^{1.3} * (M/2.0)^{-1.4}$	$E = (0.48/0.74) * TSP$	$E = (0.35/0.74) * TSP$	$E=0.07 * TSP$	kg/ton
Axetell, Cowherd 1984; USEPA, 1998a	Coal (Production)	$E = 1.16 * M^{-1.2}$	$E = 0.119 * M^{-0.9}$		$TSP * 0.019$	lb/ton
USEPA, 1991; USEPA, 1995	Coal (Production)	$E = 0.580 * M^{-1.2}$	$E=0.0596 * M^{-0.9}$	$PM\ 15 * 0.75$	$TSP * 0.019$	kg/ton
Chakraborty et al., 2002; Lal, Tripathy, 2012	Coal (Production)	$E =[\{(100 - M))/M\}^{0.1} * \{S/(100 - S)^{0.3} * H^{0.2} * \{U/(0.2 + 1.05 * U)\} * (X * L)/(15.4 + 0.87 * X * L))]$				g/s
NPI, 2012	Coal (Production)	$E = 0.580 * M^{-1.2}$		$E = 0.0447 * M^{-0.9}$		kg/ton
Chaulya, 2006	Iron (Ore)	$E = [H * \{(100 - M/M\}^{0.4} * \{0.555 * S/(100 - S)\} * (U^2 * X * L)^{0.1}$				g/s

E: Emission factor
M: Moisture content (mass %)
S: Silt content (%)
U : Wind speed (m/h)
X : Loading frequency (#/h)
H: Dumping height (m)
L: Bucket capacity (m^3)

Models developed for similar mining operations can be used to forecast particulate concentrations from loading activities during the initial planning stage of a new mining project. These models offer valuable insights that support the development and implementation of sustainable mining practices, ensuring that environmental impacts are thoroughly considered from the outset. The prediction equations developed in this study can also be used to estimate the dust concentrations to which personnel working in South African quarries are exposed.

Declarations

Funding

This work was supported by the Scientific Research Project Fund of Sivas Cumhuriyet University under Grant [M-780].

Competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Data and model availability

All data processing scripts, WEKA configuration files, trained models, and a representative subset of the field dataset are archived on Zenodo (DOI: 10.xxxx/zenodo.xxxxxx) and mirrored on GitHub (release v1.0). Due to operational confidentiality, raw continuous data are not public but may be inspected upon reasonable request.

Author contributions

Zekeriya Duran and Bülent Erdem envisioned and planned the research project. Zekeriya Duran and Mehmet Genç carried out the material preparation and data gathering tasks. Zekeriya Duran and Tuğba Doğan conducted the traditional statistical analyses. Zekeriya Duran developed the decision tree models. Zekeriya Duran and Bülent Erdem wrote the initial draft of the manuscript. All contributors provided feedback on all versions of the written work.

References

Ahmad, A.U., Balakrishnan, U.V., Jha, P.S. 2021. A study of multicollinearity detection and rectification under missing values. *Turkish Journal of Computer and Mathematics Education*, 12(1S), pp. 399–418. https://doi.org/10.17762/turcomat.v12i1S.1880

Akçetin, E., Çelik, U. 2014. The performance benchmark of decision tree algorithms for spam e-mail detection. *Journal of internet applications and management*, vol. 5, no. 2), pp. 43–56. https://doi.org/10.5505/iuyd.2014.43531

Akın, E., Şahin, M.E. 2024. A review on deep learning and artificial neural network models. EMO Scientific Journal: *Scientific Refereed Journal of Electrical, Electronics, Computer, Biomedical, Control Engineering*, vol. 14, no. 1, pp. 27–38.

Aksu, G. 2018. Investigation of data mining methods used for estimating pisa success (In Turkish). Hacettepe University *Graduate School of Educational Sciences* (Doctoral Thesis), Ankara.

Albayrak, A.S. 2005. Alternative to the least squares technique in multiple linear connection, biased estimation techniques and an application (in Turkish), *International Journal of Management Economics and Business*, vol. 1, no. 1, pp. 106–126.

Alsultanny, Y.A. 2020. Machine learning by data mining REPTree and M5P for predicating novel information for PM10. Cloud Computing and Data Science, pp. 40–48.

Altair. 2024. The regression trees returned by the operators W-M5P and W-REPTree. https://community.rapidminer.com/discussion/440/the-regression-trees-returned-by-the-operators-w-m5p-and-w-reptree [accessed 27 May 2024]

Araújo, P.S.I., Costa, B.D., J.B de Moraes, R. 2014. Identification and characterization of particulate matter concentrations at construction jobsites. Sustainability, 6, pp. 7666–7688.

Arregocés, H.A., Rojano, R., Restrepo, G. 2021. Effects of lockdown due to the COVID-19 pandemic on air quality at Latin America's largest open-pit coal mine. *Aerosol and Air Quality Research*, vol. 21, no. 8, p. 200664.

Avcı, T. 2020. The effects of macroeconomic factors and bank loans on inflation: the case of Turkey. *Journal of Corporate Governance, Insurance, and Risk Management*, vol. 7, no. 2, pp. 70–86.

Aydemir, E. 2018. *Artificial intelligence with Weka* (in Turkish), Ankara: Seçkin Publishing House.

Aydemir, E., Kaysi, F., Yavuz, M. 2020. Estimation of income obtained by tree algorithms using drug sales data. *Anatolian Journal of Computer Sciences*, vol. 5, no. 1, pp. 14–21.

Axetell, K., Cowherd, C. 1984. *Improved emission factors for fugitive dust from western surface coal mining sources*. Industrial Environmental Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, USA.

Barros, R.C., De Carvalho, CPLFA., Freitas, A.A. 2015. Automatic design of decision-tree induction algorithms. *SpringerBriefs in Computer Science*, 176p, London.

Bükey, A.M., Çetin, B.I. 2017. The analysis of the factors influencing income distribution by the least squares method in Turkey (in Turkish), *Research Journal of Public Finance*, vol. 3, no. 1 pp. 103–117.

Büyükuysal, M.Ç., Öz, İ.İ. 2016. An alternative method to least squares in presence of multicollinearity: ridge regression (in Turkish), *Journal of Düzce University Health Sciences Institute*, vol. 6, no. 2, pp. 110–114.

Cao, J., Chow, C.J., Lee, S.C.F., Watson, G.J. 2013. Evolution of PM2.5 measurements and standards in the U.S. and future perspectives for China. *Aerosol and Air Quality Research*, vol. 13, no. 4, pp. 1197–1211.

Chai, T., Draxler, R.R. 2014. Root mean square error (RMSE) or mean absolute error (MAE)? - *Arguments against avoiding RMSE in the literature*. Geoscientific Model Development, vol. 7, pp. 1247–1250.

Chakraborty, M.K., Ahmad, M., Singh, R.S., Pal, D., Bandopadhyay, C., Chaulya, S.K. 2002. Determination of the emission rate from various opencast mines operations. *Environmental Modelling & Software*, vol. 17, no. 5, pp. 467–480.

Chaulya, S.K. 2006. Emission rate formulae for surface iron ore mining activities. *Environmental Modeling & Assessment*, vol. 11, no. 4, pp. 361–370.

Chien, C.F., Chen, L.F. 2008. Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry, Expert Systems with Applications, vol. 34, no. 1, pp. 280-290.

Chow, C.J., Watson, J.G. 1998. *Guideline on speciated particulate monitoring*. U.S. Environmental Protection Agency, Research Triangle Park.

Copeland, M. 2016. What's the difference between artificial intelligence, machine learning and deep learning, https://blogs.nvidia.com/blog/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/ Accessed 2 Sep 2025.

Czajkowski, M., Kretowski, M. 2010. Globally induced model trees: an evolutionary approach. 11th International Conference on Parallel Problem Solving from Nature, September vol. 11–15, pp. 324–333, Krakow.

Çınaroğlu, S. 2017. Comparison of Machine Learning Regression Methods to Predict Health Expenditures, Uludağ University Journal of The Faculty of Engineering, vol. 22, no. 2, pp. 179–200. DOI: 10.17482/uumfd.338805.

Del Campo-Avila, J., Moreno-Vergara, N., Trella-Lopez, M. 2011. Analyzing factors to increase the influence of a Twitter user. Advances in Intelligent and Soft Computing, vol. 89, pp. 69–76.

Deng, L., Dong, Y. 2014. Deep learning: methods and applications, Foundations and trends® in signal processing 7, 3–4, pp. 197–387.

Dreshaj, A., Millaku, B., Shala, S., Selimaj, A., Shabani, H. 2017. Sources of air pollution, environmental impacts and exploitation of natural resources in Kosovo. *Proceedings of CBU International Conference on Innovations in Science and Education*, March 22-24, pp. 1275–1280, Prague.

Duan, N. 1983. Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association*, vol. 78, pp. 605–610.

Duncan, J., Stolarczyk, L.G. 2015. Mining complex geology, mitigation of float dust, and developing autonomous machine capability using horizon sensing technology for coal seam boundary detection. *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 115, no. 11, pp. 1013–1020. http://dx.doi.org/10.17159/2411-9717/2015/v115n11a5

Duran, Z., Erdem, B., Doğan, T. 2021. Particulate matter emissions in surface mining operations: literature research (in Turkish), Journal of ESOGU Engineering and Architecture Faculty, vol. 29, no. 3, pp. 450–465, https://doi.org/10.31796/ogummf.945742

Duran, Z. 2022. Particulate matter emission measurement and modeling of variation in some surface mines with meteorological conditions, material and heavy mining equipment properties (In Turkish). Doctoral dissertation, *Sivas Cumhuriyet University*.

Econorm. 2024. Turnkey ambient air portable dust analyzers (in Turkish). http://www.econorm.com.tr/urunler/ortamda-ve-online-toz-olcum-cihazlari/turnkey-dustmate-anlik-toz-olcum-cihazi/ [accessed 27 May 2024]

EPA. 2020. *What is PM?* U.S. Environmental Protection Agency, Washington, DC. [(https://www3.epa.gov/region1/airquality/pm-what-is.html)](https://www3.epa.gov/region1/airquality/pm-what-is.html)

Erbay, Ş., Beydoğan, H.Ö. 2017. Educators' attitudes towards educational research (in Turkish), *Ahi Evran University Kırşehir Journal of Education Faculty*, vol. 18, no. 3, pp. 246–260.

Evyapan, F., Mungan, D., Akgün, M., Arbak, P. 2012. Air *quality and health* (in Turkish). T.R. *Ministry of Health and Turkish Thoracic Association*, ISBN: 978-1-84984-008-8, Ankara.

Gargano, M.L., Raggad, B.G. 1999. Data mining-a powerful information creating tool. OCLC Systems & Services, vol. 15, no. 2, pp. 81–90.

Gautam, S., Prusty, B.K., Patra, A.K. 2012. Pollution due to particulate matter from mining activities, Reciklaža i održivi razvoj 5, pp. 53–58.

George, D., Mallery, M. 2010. SPSS for Windows step by step: a simple guide and reference, 17.0 update (10th ed.). Allyn & Bacon, 386p, Boston.

Ghose, M.K. 2004. Emission factors for the quantification of dust in India coal mines. *Journal of Scientific and Industrial Research*, vol. 63, pp. 763–768.

Ghose, M.K., Majee, S.R. 2007. Characteristics of hazardous airborne dust around an Indian surface coal mining area. *Environmental Modeling & Assessment*, vol. 130, pp. 17–25.

Gültepe, Y. 2019. A Comparative Assessment on Air Pollution Estimation by Machine Learning Algorithms. *European Journal of Science and Technology*, vol. 16, pp. 8–15.

Henedy, S.N., Naser, A.H., Imran, H., Bernardo, L.F., Teixeira, M.M., Al-Khafaji, Z. 2022. Development of Prediction Models for the Torsion Capacity of Reinforced Concrete Beams Using M5P and Nonlinear Regression Models. J. Compos. Sci., vol. 6, pp. 366.

Hime, N., Cowie, C., Marks, G. 2015. Review of the health impacts of emission sources, types and levels of particulate matter air pollution in ambient air in NSW. *NSW Environment Protection Authority and NSW Ministry of Health, Environmental Health Branch*.

Huertas, J.I., Camacho, D.A., Huertas, M.E. 2012. Standardized emissions inventory methodology for open-pit mining areas. Environmental Science and Pollution Research, vol. 19, pp. 2784–2794.

Hwang, S.Y., Yang, D.N., Kwon, H.J., Lee, J.Y. 2014. Measurement of PM10 concentration in one dental clinic. International Journal of Clinical Preventive Dentistry, vol. 10, no. 4, pp. 273-280. http://dx.doi.org/10.15236/ijcpd.2014.10.4.273

Karabulut, C. 2019. Development of E-Commerce in Turkey After 2000. *Journal of Management, Economic and Marketing Research,* vol. 3, no. 5, pp. 192–207. doi: 10.29226/TR1001.2019.152

Kara, Ş.E., Şamlı, R. 2021. Comparative analysis of machine learning algorithms in WEKA for cost estimation of software projects. European Journal of Science and Technology, vol. 23, pp. 415–426. doi: 10.31590/ejosat.877296

Kayaalp, K., Süzen, A.A. 2018. Deep Learning and Applications in Türkiye (In Turkish), Institution of Economic Development and Social Researches Publications, pp. 93.

Keskinkılıç, B., Karakaya, K., Kutlu, Ş., Zal, H. 2010. A health approach to Turkey's air pollution and climate change problems (In Turkish). T.R. *Ministry of Health General Directorate of Health Services*, Ankara.

Kim, K-H., Kabir, E., Kabir, S. 2015. A review on the human health impact of airborne particulate matter. *Environment International*, vol. 74, pp. 136–143.

Kumar, E., Kumar, L.A. 2017. A methodology for assessing the cumulative impact of a mineral transportation system on air quality. *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 117, no. 4, pp. 401–408. http://dx.doi.org/10.17159/2411-9717/2017/v117n4a13

Lal, B., Tripathy, S.S. 2012. Prediction of dust concentration in open cast coal mine using artificial neural network. Atmospheric Pollution Research,vol. 3, pp. 211–218. https://doi.org/10.5094/APR.2012.023

Lashgari, A., Kecojevic, V. 2016. Comparative analysis of dust emission of digging and loading equipment in surface coal mining. *International Journal of Mining, Reclamation and Environment*, vol. 30, no. 3, pp. 181–196. https://doi.org/10.1080/17480930.2015.1028516

Löndahl, J., Pagels, J., Swietlicki, E., Zhou, J.C., Ketzel, M., Massling, A., Bohgard, M. 2006. A set-up for field studies of respiratory tract deposition of fine and ultrafine particles in humans. *Journal of Aerosol Science*, vol. 37, no. 9, pp. 1152–1163.

Montgomery, D.C., Peck, E.A., Vining, G.G. 2012. *Introduction to linear regression analysis*. Wiley, 679p, Hoboken.

Muma, D., Besa, B., Manchisi, J., Banda, W. 2020. Effects of mining operations on air and water quality in Mufulira district of Zambia: A case study of Kankoyo Township. *Journal of the Southern African Institute of Mining and Metallurgy, vol.* 120, no. 4, pp. 287–298. http://dx.doi.org/10.17159/2411-9717/952/2020

Murzin, M.A., Gorlenko, N.V. 2020. Analyzing effects of dust from mining enterprises on the environment, Agritech-IV-2020. Proceedings of IOP Conf. Series: Earth and Environmental Science 677 (2021) 042091. https://doi.org/10.1088/1755-1315/677/4/042091

Njeguš, A., Nikolić, V., Jovanović, V. 2015. The selection of optimal data mining method for small-sized hotels. *International Scientific Conference of IT and Business-Related Research*, April 16, pp. 519–524, Belgrade.

NPI. 2012. *Emission estimation technique manual for mining-version 3.1.* Australian Government, Canberra.

NSW, EPA. 2015. *NSW coal mining benchmarking study best practice measures for reducing non-road diesel exhaust emissions*. NSW Environment Protection Authority (EPA), Sydney.

Onan, A. 2015. Comparative performance analysis of decision tree algorithms in the corporate bankruptcy prediction. Journal of information technologies, vol. 8, no. 1, pp. 9–19. https//doi.org/10.17671/btd.36087

Önder, M., Yiğit, E. 2009. Assessment of respirable dust exposures in an opencast coal mine. *Environmental Monitoring & Assessment*, vol. 152, no. 1-4, pp. 393–401. https://doi.org/10.1007/s10661-008-0324-4

Öztürk, E. 2012. Transformation and segmentation operations on images for increasing the effectiveness of compression methods (In Turkish). *Trakya University Graduate School of Natural and Applied Sciences* (Master's Thesis), Edirne.

Öztürk, N. 2016. Assessment of noise, vibration and dust exposure of the heavy equipment operators in surface mines (In Turkish). Expertise thesis, T.R. Ministry of Labour and Social Security, Directorate General of Occupational Health and Safety, Ankara.

Papagiannis, A., Roussos, D., Menegaki, M., Damigos, D. 2014. Externalities from lignite mining-related dust emissions. *Energy Policy*, vol. 74, pp. 414–424. https://doi.org/10.1016/j.enpol.2014.08.026

Patra, A.K., Gautam, S., Kumar, P. 2016. Emissions and human health impact of particulate matter from surface mining operation - a review. *Environmental Technology & Innovation*, vol. 5, pp. 233–249. https://doi.org/10.1016/j.eti.2016.04.002

Petavratzi, E., Kingman, S., Lowndes, I. 2005. Particulates from mining operations: A review of sources, effects and regulations. *Minerals Engineering*, vol. 18, no. 12, pp. 1183–1199. https://doi.org/10.1016/j.mineng.2005.06.017

Poisson, A., Guezou, J.C., Öztürk, A., İnan, S., Haluk, T., Gürsoy, H., Kavak, K.S., Özden, S. 1996. Tectonic setting and evolution of the sivas basin, central anatolia, turkey. *International Geology Review*, pp. 838–853. https://doi.org/10.1080/00206819709465366

Quinlan, J.R. 1992. Learning with continuous classes. 5th Australian Joint Conference on Artificial Intelligence, pp. 343–348, Singapore.

Richardson, C., Rutherford, S., Agranovski, E.I. 2019. Open cut black coal mining: Empirical verification of PM2.5 air emission estimation techniques. *Atmospheric Research*, vol. 216, pp. 151–159. https://doi.org/10.1016/j.atmosres.2018.10.008

Rojano, R.R., Arregocés, H.A., Restrepo, G. 2020. PM10 Concentrations measured in open pit coal mines in northern colombia: seasonal variations, trends and source. *WIT Transactions on Ecology and the Environment*, vol. 244. https://doi.org/10.2495/AIR200081

Sabti, A.A., Rashid, S.M., Hummadi, A.S. 2019. Interrelationships between writing anxiety dimensions and writing goal orientation among Iraqi EFL undergraduates, *International Journal of Instruction*, vol. 12, no. 4, pp. 529-544, https://doi.org/10.29333/iji.2019.12434a

Schober, P., Boer, C., Schwarte, L.A. 2018. Correlation coefficients: appropriate use and interpretation, Anesthesia & Analgesia, vol. 126, no. 5, pp. 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

SÇR. 2023. Sivas Province 2023 State of Environment Report, T.C. Sivas Governorship Provincial Directorate of Environment, Urbanization and Climate Change (in Turkish), 217s, Sivas.

Sihag, P., Singh, B., Said, A., Azamathulla, H.M. 2021. Prediction of Manning's coefficient of roughness for high-gradient streams using M5P. *Water Supply*, vol. 22, no. 3, pp. 2707–2720. https://doi.org/10.2166/ws.2021.440

Singh, P.K., Roy, M.P., Paswan, R.K., Sarim, M., Kumar, S., Ranjan, Jha.R. 2016. Rock fragmentation control in opencast blasting. Journal of Rock Mechanics and Geotechnical Engineering, vol. 8, no. 2, pp. 225–237.

Shrestha, N. 2020. Detecting multicollinearity in regression analysis. American Journal of Applied Mathematics and Statistics, vol. 8, no. 2, pp. 39–42. https://doi.org/ 10.12691/ajams-8-2-1

Stackexchange. 2024. M5P Interpretations and Questions. https://stats.stackexchange.com/questions/228724/m5p-interpretations-and-questions [accessed 27 May 2024]

Tabachnick, B.G., Fidell, L.S. 2013. Using multivariate statistics (6th ed.), Pearson, Boston.

Tanni, S.E., Patino, C.M., Ferreira, J.C. 2020. Correlation vs. regression in association studies. Jornal Brasileiro de Pneumologia, 46(1): e20200030. https://doi.org/10.1590/1806-3713/e20200030

Trechera, P., Moreno, T., C´ordoba, P., Moreno, N., Zhuang, X., Li, B., Li, J., Shangguan, Y., Dominguez, A.O., Kelly, F., Querol, X. 2021. Comprehensive evaluation of potential coal mine dust emissions in an open-pit coal mine in Northwest China. *International Journal of Coal Geology*, vol. 235, pp. 1–17. https://doi.org/10.1016/j.coal.2021.103677

Tripathy, D.P., Dash, T.R., Badu, A., Kanungo, R. 2015. Assessment and modelling of dust concentration in an opencast coal mine in India. *Global Nest Journal*, vol. 17, no. 4, pp. 825–834.

Trivedi, R., Chakraborty, M.K., Sangode, A.G., Tewary, B.K. 2010. A study of air pollution load assessment around opencast coal project in India. *Indian Journal of Environmental Protection*, vol. 30, no. 3, pp. 198–206.

TS 1900-1. 2006. *Methods of testing soils for civil engineering purposes in the laboratory - Part 1: Determination of physical properties* (in Turkish), Turkish Standards Institute, Ankara.

TS 4447. 1985. *Gypsum- Ground (for Agricultural Use)* (in Turkish), Turkish Standards Institute, Ankara.

Turnkey. 2024. DustMate. https://turnkey-instruments.com/product/dustmate/ [accessed 27 May 2024]

USEPA. 1991. *Review of surface coal mining emissions factors*. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

USEPA. 1995. *AP-42 compilation of air pollutant emission factors*, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

USEPA. 1998a. *Revision of emission factors for AP-42 Section 11.9 Western surface coal mining*. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

USEPA. 1998b. *Particulate matter ($PM_{2.5}$) speciation guidance document*. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

USEPA. 2006. Revision of emission factors for AP-42. Chapter 13: miscellaneous source. section 13.2.4: aggregate handling and storage piles. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Usha, T.M., Balamurugan, S.A. 2016. Seasonal based electricity demand forecasting using time series analysis. Circuits and Systems, vol. 7, no. 10, http://dx.doi.org/10.4236/cs.2016.710283

Vidinli, N., Özkan, E.K., Topçu, D., Yamurluklu, Y., Gedikli, F.G., Kürkçü, E.A. 2016. Guidance on dust control in the cement sector (In Turkish). T.R. Ministry of Labour and Social Security, Ankara.

Wang, W., Xu, Z. 2004. A heuristic training for support vector regression. Neurocomputing, vol. 61, pp. 259–275. https://doi.org/10.1016/j.neucom.2003.11.012

WHO. 1999. *Hazard prevention and control in the work environment: airborne dust*. World Health Organization, Geneva.

WHO. 2023. Ambient (outdoor) air pollution. World Health Organization, Geneva. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health [accessed 23 September 2023].

WEKA. 2024. Weka machine learning software. Waikato University, Hamilton. https://sourceforge.net/p/weka/mailman/search/?q=The%20second%20number%20in%20the%20model%20on%20the%20Weka%20m5P%20decision%20tree%20leaves%20is%20obtained%20by%20dividing%20the%20mean%20squared%20error%20of%20the%20linear%20model%20on%20each%20leaf%20by%20the%20global%20absolute%20deviation.&limit=50 [accessed 27 May 2024]

WHO. 2006. *Health risks of particulate matter from long-range transboundary air pollution*. World Health Organization, Copenhagen.

Witten, I.H., Frank, E., Hall, M.A. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington: *Morgan Kaufmann Publishers*. https://doi.org/10.1016/C2009-0-19715-5

Yazıcı, A.C., Öğüş, E., Ankaralı, S., Canan, S., Ankaralı, H., Akkuş, Z. 2007. Artificial neural networks: review, Journal of Surgical Medical Sciences, vol. 27, pp. 65–71. ◆

MINESAFE 2026
MINE HEALTH AND SAFETY CONFERENCE

ENGINEERING HEALTH & SAFETY FOR SUSTAINABLE FUTURE-READY MINES

HYBRID CONFERENCE

11-12 NOVEMBER 2026 — CONFERENCE
13 NOVEMBER 2026 — INDUSTRY AWARDS DAY
GALLAGHER CONVENTION CENTRE, JOHANNESBURG, SOUTH AFRICA

ZERO HARM

AMSPSA | MINERALS COUNCIL SOUTH AFRICA | MMMA | SAIMM THE SOUTHERN AFRICAN INSTITUTE OF MINING AND METALLURGY | SANIRE SOUTH AFRICAN NATIONAL INSTITUTE OF ROCK ENGINEERING

Abstract submission – 31 July 2026 I Extended Abstract/Papers submission- 31 August 2026

BACKGROUND

We are excited to announce our upcoming industry Conference and Industry Awards Day, dedicated to advancing mining safety and health, innovation, and sustainability and recognising the excellent work in creating ZERO HARM future ready mines.
This year's theme, "Engineering Health, Safety for Sustainable Future-Ready Mines", underscores our collective commitment to achieving zero harm while embracing modernisation, technology, and collaboration as the foundation for safer, future-ready mining operations.

This conference will serve as a vital platform for knowledge-sharing and idea exchange among key stakeholders, including mining companies, the Department of Mineral Resources and Energy (DMRE), the Minerals Council South Africa, labour unions, and health and safety practitioners at all levels in the minerals industry.

CONFERENCE THEMES

1. Driving Zero Harm Forward
- Reinforcing the vision of eliminating fatalities.
- Sharing progress, lessons learned, and renewed commitments.

2. Technology in Action
- Case studies of mines implementing breakthrough technologies.
- Practical examples of problems solved through innovation.

3. Future-Ready Mine Design
- Strategies for long-term sustainability and resilience.
- Integrating safety and health into planning and operational frameworks.

4. Engineering Safety Excellence
- Advances in engineering solutions that enhance safety.
- Best practices in design, maintenance, and risk management.

5. Engineering in the Digital Age
- Modernisation and digital transformation in mining operations.
- Leveraging data, automation, and AI for safer outcomes.

6. Technology Advancements and Innovation
- Exploring robotics, wearables, sensors, and predictive analytics.
- Emerging technologies shaping the future of mining safety.

7. Health and Wellbeing in Mining
- Protecting workers' physical and mental health.
- Occupational health strategies, wellness programs, and medical innovations.

8. Sustainability and Responsible Mining
- Environmental stewardship and climate-conscious practices.
- Linking safety with sustainability for long-term industry resilience.

Why Attend?
- Gain insights from global case studies and pioneering projects.
- Network with industry leaders, engineers, and technology innovators.
- Explore practical solutions for building safer, sustainable mines.
- Be part of shaping the future of mining through collaboration and innovation.

Join us as we work together towards Safe Mines, Healthy Lives, and Sustainable Futures!

PARTNERSHIP OPPORTUNITIES

Sponsorship opportunities are available. Companies wishing To partner on this event should contact the Conference Coordinator.

WHO SHOULD ATTEND

The conference should be of value to:
- Safety practitioners
- Mine management
- Mine health and safety officials
- Engineering managers
- Underground production supervisors
- Surface production supervisors
- Environmental scientists
- Minimizing of waste
- Operations manager
- Processing manager
- Contractors (mining)
- Including mining consultants, suppliers and manufacturers
- Education and training • Energy solving projects
- Water solving projects
- Unions
- Academics and students
- DMRE
- Occupational health practitioners
- Leadership and ManagementTechnical and Engineers

CALL FOR PAPERS

SUBMIT AN ABSTRACT

Call for papers on the topics of safety, health and environment
Prospective authors are invited to submit titles and abstracts of their presentations in English and not longer than 500 words. Abstracts should be submitted to:
Include Abstract Submission Deadline: **31 July 2026**
Extended Abstract/Paper Submission: **31 August 2026**

For further information contact:
Gugu Charlie, Conference Co-ordinators
E-mail:gugu@saimm.co.za | Tel: +27 11 538-0238
| Web: www.saimm.co.za