MJFAS
Malaysian Journal of Fundamental and Applied Sciences

Amin *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 21 (2025) 2120-2135

RESEARCH ARTICLE

Prediction Model of Health Index for Petrochemical Industrial Area, Paka, Terengganu and Gebeng, Pahang, Malaysia

Siti Noor Syuhada Muhammad Amin[a]*, Azman Azid[b], Muhammad Yusran Abdul Aziz[a], Sharifah Wajihah Wafa Syed Saadun Tarek Wafa[c], Nurul Alia Azizan[c]

[a]UniSZA Science and Medicine Foundation Centre, Gong Badak Campus, Universiti Sultan Zainal Abidin, 21300 Kuala Nerus, Terengganu, Malaysia; [b]Faculty of Bioresources and Food Industry, Besut Campus, Universiti Sultan Zainal Abidin, 22200 Besut, Terengganu, Malaysia; [c]Faculty of Health Sciences, Gong Badak Campus, Universiti Sultan Zainal Abidin, 21300 Kuala Nerus, Terengganu, Malaysia

Abstract Atmospheric dust can contain various heavy metals which are known to have detrimental effects on human health. Understanding the spatial distribution of these heavy metals is essential for evaluating the potential risks connected with exposure to atmospheric dust. The purpose of this study is to develop a prediction model that uses the concentrations of heavy metals in the study area to evaluate health hazards in both adults and children. High volume air samplers were loaded with filter paper at a constant flow rate for dust sampling in Paka and Gebeng. The aqua regia method was used to treat the dust samples. The data were analysed by using chemometrics and Artificial Neural Network (ANN) to show the spatial and prediction model of health index from heavy metals concentrations in the study area. The results showed the total heavy metal concentrations were found to be in the following decreasing order of iron (Fe) (0.218 mg/L ± 0.192), zinc (Zn) (0.083 mg/L ± 0.059), lead (Pb) (0.079 mg/L ± 0.119), cadmium (Cd) (0.004 mg/L ± 0.003), copper (Cu) (0.004 mg/L ± 0.004) and arsenic (As) (0.001 mg/L ± 0.001) for northeast monsoon whereas Fe (0.407 mg/L ± 0.270), Zn (0.110 mg/L ± 0.092), Pb (0.088 mg/L ± 0.118), Cd (0.008 mg/L ± 0.008), Cu (0.004 mg/L ± 0.005) and As (0.004 mg/L ± 0.001) for southwest monsoon. Three principal components factor for northeast and southwest monsoon respectively were extracted from Principal Components Analysis (PCA), which are based on eigenvalue (>1.0). During northeast monsoon, Factor 1 revealed Fe and As. Factor 2 revealed Cu, Cd and Zn and Factor 3 revealed Pb. During southwest monsoon, Factor 1 revealed Pb, Cd and Zn. Factor 2 revealed Fe and Cu and Factor 3 revealed As. The estimations of HQ for pathways in this study decreased in the order of ingestion>dermal contact>inhalation. Health index (HI) prediction model value for adults decreased in the order of Fe>Pb>Cd>As>Zn>Cu whereas Fe>Pb>Cd>As>Cu>Zn HI prediction model value for children. HI values of these metals for children were higher than adults. However, the values of health risk obtained in this study are in the receivable range (HI<1.0). Fe concentrations were recorded highest in both areas and seasons, while PCA revealed three factor analysis of heavy metals for both areas and seasons with As being the most identifying sources of heavy metals through sensitivity analysis. ANN was an efficient technique to compute prediction models of health index in the study area. This study suggested that the prediction models approached, will provide a better insight into air quality information to understand potential environmental health hazard towards people and mitigation strategies plan in the future.

Keywords: Health index, health quotient, atmospheric dust, petrochemical industrial area, heavy metals.

***For correspondence:**
syuhadaamin@unisza.edu.my

Received: 03 Feb. 2025
Accepted: 13 May 2025

Introduction

In the industrial sector, atmospheric particles are one of the main issues. Different types of chemicals, including metals, organic compounds and inorganic compounds can trigger production of the atmospheric particles [1]. Atmospheric particles in the ambient affect many different actions in the surroundings, such as reducing how far we can see, deposition of acid rain and influencing the amount of radiation we receive. In addition to directly generating these effects, these particles may also indirectly contribute by creating clouds [2]. The earth's crust is made up of trace metals, which are naturally occurring substances that are stable and unbreakable. In addition to dust carried by the wind, human activities like metal mining, using fossil fuels, conducting metallurgical and combustion processes, or working in industrial plants can all be sources of heavy metals in the atmosphere [3]. In this study, the filter was used to extract precise information regarding the particulate matter with a size greater than 2.5 micrometers (> 2.5 µm).

If found in the water, soil or air, some heavy metals can be detrimental to both the ambient and human health because of their high density. This has resulted in serious health issues for people in the world. In Directive 2008/50/EC, the European Commission established a boundary of 0.5 x10-6 mg/L for lead (Pb), whereas the target values for cadmium (Cd), nickel (Ni) and arsenic (As), are 0.005 x10-6 mg/L, 0.2 x10-6 mg/L and 0.006 x10-6 mg/L respectively [3].

Moreover, high exposure to trace metals can result in the accumulation of metals in the body, which can lead to various health problems, such as respiratory ailments, cardiovascular incidents and even early death [4]. Humankind can be exposed to these elements in the environment through inhalation, ingestion and dermal contact. All these considerations make it clear why heavy metals, even in trace amounts can affect human health [5].

When developing a plan to manage air quality and reduce pollution in a sustainable manner, it is crucial to identify the quantity of sources of heavy metals in samples of atmospheric dust. By determining the levels of zinc (Zn), copper (Cu), lead (Pb), iron (Fe), arsenic (As) and cadmium (Cd) in the air dust in Paka and Gebeng, two industrial areas in Malaysia, this study attempts to address the lack of data on air quality in these areas. This selection of heavy metals was made because they have been linked to emissions from transport and industry, which can potentially lead in effect towards human health [6].

Hotspots of heavy metals contamination area are areas where the concentration of heavy metals in the surrounding area is significantly higher than the surrounding areas. This can be due to a variety of factors, such as industrial emissions, mining activities, and the use of pesticides. People who live or work in these areas may be exposed to heavy metals through the air, water, food, and soil. In this study, both Paka and Gebeng were petrochemical industrial area and were assumed to have an impact towards human health from the contamination of heavy metals.

Materials and Methods

Study Area and Sampling Location

16 points of location were selected in order to sample for both the southwest monsoon (August and September 2017) as well as the northeast monsoon (December 2017 and February 2018) (Table 1). The sampling site's location (Figure 1) was chosen with accessibility, safety, contamination prevention, and environmental feature representation in mind. The majority of the energy sector, oil and gas operations, and other factors contributed to the selection of these places by raising air pollution and health risks. The consistent temperature, high humidity and copious amounts of rainfall are characteristics of Malaysia's climate. Winds are usually not very strong. There are two monsoon regimes that define the weather in the research area. The northeast monsoon (November to March) and the southwest monsoon (May until September). In the southwest monsoon season, the wind speed is light, typically less than 15 knots. The predominant winds during the northeast monsoon season come from the east or northeast and range in 10 to 20 knots of speed. While the weather was comparatively drier during the southwest monsoon, the northeast monsoon brings torrential rain. Transition times are when the two monsoons change from one another. The study included background sites with little industrial activity for comparison. It also considered factors like weather and construction to minimize external biases. The purpose of choosing the listing industry is due to the need for transparency in identifying major pollution sources and their environmental impact. Specifying industries allows for a precise analysis of emission patterns, strengthening the study's credibility in assessing air quality and associated health risks. This ensures a more accurate evaluation of industrial contributions to pollution.

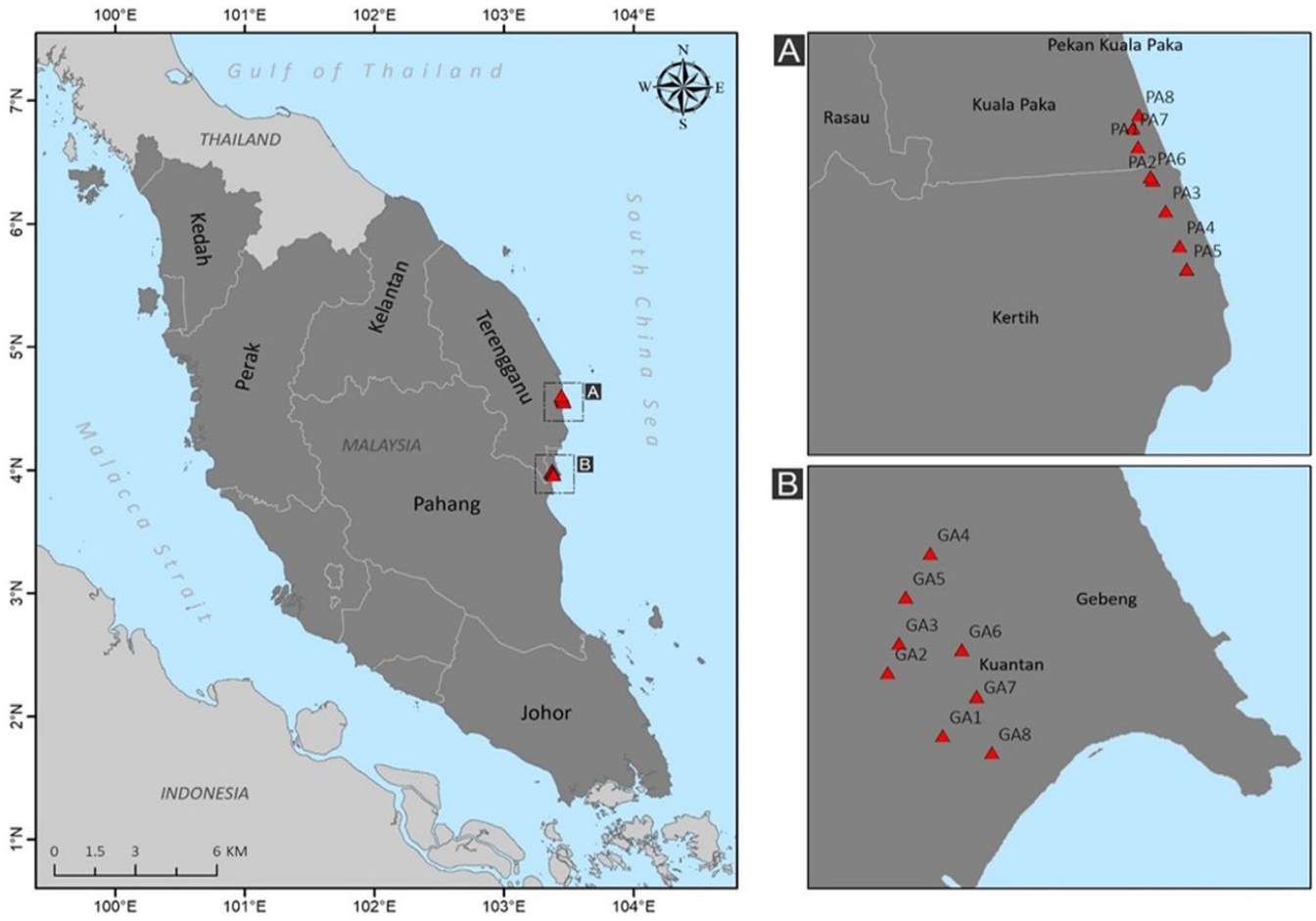

Figure 1. The sampling site in Paka/Kerteh and Gebeng area

Table 1. The site of sampling in Gebeng and Paka/Kerteh area

Site	Industrial Area Name	Longitude	Latitude
Paka 1 (P1)	Kemaman Technology and Industrial Park S/B	N04˚35'58.0"	E103˚26'41.1"
Paka 2 (P2)	Petronas Filtering (Terengganu) S/B	N04˚35'16.7"	E103˚26'59.0"
Paka 3 (P3)	Kertih Terminal S/B	N04˚34'37.6"	E103˚27'14.2"
Paka 4 (P4)	Petronas Filtering (Terengganu) S/B	N04˚33'54.0"	E103˚27'30.9"
Paka 5 (P5)	Terengganu Crude Oil Terminal	N04˚33'24.7"	E103˚27'39.5"
Paka 6 (P6)	Petronas Acetyls (BP) S/B	N04˚35'20.7"	E103˚26'56.3"
Paka 7 (P7)	Taman Hadhari Paka Mosque	N04˚36'22.6"	E103˚26'35.1"
Paka 8 (P8)	Taman Paka Indah 3	N04˚36'38.6"	E103˚26'41.9"
Gebeng 1 (G1)	Kuantan Training Industrial Institute	N03˚58'03.9"	E103˚22'25.9"
Gebeng 2 (G2)	Petronas Chemical (BASF) Sdn. Bhd.	N03˚58'48.1"	E103˚21'49.5"
Gebeng 3 (G3)	Petronas Chemical (BASF) Sdn. Bhd.	N03˚59'08.5"	E103˚21'56.8"
Gebeng 4 (G4)	Lynas Advanced Material Plant	N04˚00'11.7"	E103˚22'17.8"
Gebeng 5 (G5)	Petronas Chemicals MTBE/PDH Sdn. Bhd.	N03˚59'41.0"	E103˚22'01.4"
Gebeng 6 (G6)	Eastman Chemical (M) Sdn. Bhd.	N03˚59'04.0"	E103˚22'38.8"
Gebeng 7 (G7)	Petronas Chemicals MTBE Sdn. Bhd.	N03˚58'31.4"	E103˚22'48.6"
Gebeng 8 (G8)	Taman Balok Makmur	N03˚57'51.9"	E103˚22'58.6"

The regulatory and compliance factors from Department of Environment Malaysia (DOE) might have designated Paka and Gebeng as priority areas for monitoring due to their industrial significance. The Compliance with Ambient Air Quality Guidelines (Malaysian Ambient Air Quality Standard) (MAAQS)) also used in dictate monitoring site selection.

Sample Collection

The sampling was carried out between August 2017 and February 2018 across a seven-month period. The sampling was done three times (August, September and October for southwest monsoon) and (November, December and January for northeast monsoon) [6] in two days (Day 1: P1 – P4, Day 2: P5 – P8) for Paka and (Day 1: G1 – G4, Day 2: G5 – G8) for Gebeng starting in the morning. The sampling was done in several times in order to get several readings from the filters (trapped heavy metals) after the digestion and analyzing process. A team of researchers (Team A) took the samples in Paka and another team (Team B) took samples in Gebeng. The sample collection planning was detailed out in the Table 2 with the total samples 576. Monsoon variations affect air pollution by changing wind patterns, rainfall, and temperature, which can either disperse or trap pollutants. Heavy rains help clean the air, while dry and stagnant conditions can lead to higher pollution levels.

The study focuses on the northeast and southwest monsoons, as these seasons exhibit distinct and dominant wind patterns that significantly influence pollutant dispersion. In contrast, the inter-monsoon periods are characterized by variable and unpredictable wind directions, leading to inconsistent pollution transport dynamics that are challenging to model with reliability. By excluding the inter-monsoon seasons, the study ensures a clearer analysis of long-term pollutant behavior, capturing the most impactful meteorological conditions affecting air quality.

Each area contained eight desired points and the sampling was conducted for eight hours for each point via filter paper (WhatmanTM No. 42, ≤0.007% ash, pore size 2.5 µm, 150 mm diameter). The filter paper was put inside an air sampler (Portable Air Sampler, L10B MUNIRO, Aeroqual Series 500) that was made to gather particulate matter by drawing in 100 L of air per minute with an efficiency of 100.0%. Filter papers were carefully soaked in a desiccator with fresh silica gel both before and after sampling, for two days, to reduce the effect of humidity and extract precise information regarding the particulate matter with a size greater than 2.5 micrometers (> 2.5 µm). This was done because the filter paper is hygroscopic, meaning that its weight changes as a function of humidity.

The air sampler was elevated to 1.5 m [7], as this height is expected to be similar to a human's breath level. This is accomplished by passing air though a filter for a fixed time and at a fixed speed to collect the sample of particulate matter retained on the filter within the air sampler. After sampling, the filters were placed in a clean plastic bag to protect them from light and then left to dry in a desiccator for 48 hours. The particles collected on the filters were weighed using an electronic micro-balance (SartoriusTM SECURA225D1SPR) that could measure with 0.01 mg accuracy. The air sampler operating temperature are -5.0 oC to +30.0 oC, run time from full charge can be up to 30 hours, its nominal flow rate is 5 to 10 L/min.

The portable air sampler was calibrated prior to each sampling session using a standardized flow calibration device to ensure accurate and consistent airflow rates. Calibration procedures followed manufacturer guidelines and were verified against a certified reference standard to minimize measurement errors. Regular calibration checks were conducted throughout the study to maintain data reliability and compliance with quality assurance protocols.

Sample Digestion

High-purity analytical-grade reagents were used in this study to ensure an accurate analysis. All lab apparatus and glassware were pre-cleaned by using a 5.0% nitric acid solution (HNO_3, ACS Reagent, 70.0%, Sigma-Aldrich) and then thoroughly rinsed with deionized water. A recognized standard reference material (NIST® SRM® 1648a) was used for digestion in order to perform elemental analysis on the samples and determine the concentration of heavy metals in airborne particulate matter.

The quality and accuracy of quantitative analysis, specifically with regard to urban particulate matter, were assessed using the SRM®1648a standard. Nitric acid and hydrochloric acid were combined in a 1:3 ratio (aqua regia method) as part of a digestive process in completing the task. This approach is often used to determine the presence of heavy metal elements, thus proving to be a common choice for solving environmental issues.

50 mL of aqua regia was added to separate 100 mL beakers containing SRM®1648a and WhatmanTM filter papers. After placing watch glasses over the beakers, they were heated to 140.0ºC until they were

nearly dry. Subsequently, the solutions were filtered off and the beakers were rinsed with 10.0% HNO_3 three times. When the process had been completed, the solutions were cooled to room temperature, transferred to a 100 mL volumetric flask, and diluted to volume with 10.0% HNO_3.

The accuracy, reliability, and precision of the results were tested by preparing three replicate samples and analysing them with the ICP-MS using a series of diluted standard solutions from the Merck ICP Multi-element standard solution IV. The recoveries of the elements and ions were in the range of 89.0% - 101.0%, as compared to the SRM®1648a (Table 3). The recoveries percentage of the elements in this study were acceptable as it was aligned with the study done by Nurhaini *et al*., [8] that mentioned the recoveries elements in the range of 82.0% - 101.0%.

Table 2. The sample collection planning

Term	Monsoon	Month	Day	Station
A (Paka)	Southwest	August	1	P1 – P4
			2	P5 – P8
		September	1	P1 – P4
			2	P5 – P8
		October	1	P1 – P4
			2	P5 – P 8
	Northeast	November	1	P1 – P4
			2	P5 – P8
		December	1	P1 – P4
			2	P5 – P8
		January	1	P1 – P4
			2	P5 – P8
B (Gebeng)	Southwest	August	1	G1 – G4
			2	G5 – G8
		September	1	G1 – G4
			2	G5 – G8
		October	1	G1 – G4
			2	G5 – G8
	Northeast	November	1	G1 – G4
			2	G5 – G8
		December	1	G1 – G4
			2	G5 – G8
		January	1	G1 – G4
			2	G5 – G8

In ensuring the reliability and consistency of sample collection, comprehensive quality assurance (QA) and quality control (QC) measures were implemented for both Team A and Team B. All team members underwent standardized training on sampling techniques, equipment handling, and contamination prevention to minimize variability in data collection. Standard Operating Procedures (SOPs) were strictly followed, covering key aspects such as calibration of instruments, sample preservation, and field documentation. Regular cross-validation between teams was conducted to identify any discrepancies and ensure procedural adherence. Additionally, periodic audits and duplicate sampling were performed to verify data accuracy and maintain methodological integrity throughout the study. The air sampling procedure began with the positioning of the portable air sampler at the designated site, ensuring proper placement according to standardized guidelines to avoid obstructions and external interferences. Before each sampling session, the sampler was calibrated using a flowmeter to maintain accurate and consistent airflow rates, following manufacturer specifications and quality assurance protocols. During the active sampling phase, ambient air was drawn through the inlet, with the airflow direction carefully monitored to ensure efficient pollutant capture. The collected air passed through the filter media, where particulate matter and gaseous pollutants were retained for further analysis. Upon completion, the samples were carefully sealed, labeled, and transported under controlled conditions to the laboratory to prevent contamination and ensure the integrity of the collected data.

Table 3. The standard recovery of heavy metal using certified reference materials (NIST®SRM® 1648a) of urban particulate matter

Metals	Certified Value (ppm)	Measured Value (ppm)	Recovery (%)
As	115.5 ± 3.9	114.3 ± 3.7	99.0
Cd	73.7 ± 2.3	72.8 ± 1.6	98.8
Cu	610.0 ± 70	601.8 ± 67	98.7
Fe	0.0392 ± 0.0021	0.0396 ± 0.0008	101.0
Ni	81.1 ± 6.8	79.8 ± 6.1	98.4
Pb	0.0066 ± 0.0003	0.0059 ± 0.0001	89.4
Zn	4800 ± 270	4789 ± 272	99.8

Analytical Instrument

The elemental analysis was performed by using ICPMS (Model ELAN DRC-e, Perkin Elmer, Shelton, CT) due to its high specificity, detectability, reliability and wide working concentration range. ICPMS was chosen for the quantitative measurement of heavy metals [9]. ICPMS provides the capability for highly precise and sensitive multi-element analysis of an extensive range of samples [10].

Inductively Coupled Plasma (ICP) source and mass spectrometer are combined in an ICP-MS primarily for high temperature applications. The atoms of the elements in the sample are converted to ions by the ICP source. These ions are then identified and isolated using the mass spectrometer. Then, argon gas fills the concentrically formed tubes of the ICP torch. The radio frequency (RF) load coil is connected to a radio frequency (RF) generator. While the load coil is powered by oscillating electric current, the generator and magnetic fields are positioned at the end of the torch. The argon flowing through the ICP torch is struck by a spark, which causes the argon atoms to lose their electrons and create argon ions.

These ions are drawn into the oscillating fields where they strike other argon atoms and discharge or create plasma made of argon. The sample is typically delivered as an aerosol into the ICP plasma. The sample aerosol is completely desolated once it is introduced to the ICP torch and the elements are subsequently transformed into gaseous atoms and finally ionised near the end of the plasma. With a temperature range of 6,000-10,000 K, the argon ICP plasma release is a great ion source. A novel quantitative analysis technique was developed prior to analysis and a daily performance check of ICPMS was conducted.

Metals Determination

In order to attain standard calibration factors, 10µg/L of the multi-element calibration standard was employed and diluted. Every one of the 25 metal elements were tested. Only six metals, meanwhile, were recognized and provided readings. Other relevant fields were automatically filled in after the elements (As, Fe, Cu, Pb, Cd and Zn) that will be measured were chosen. The filter that was not exposed to the air sampling procedure served as the control filter in this study. The device has a record of this technique. The blank solution (ultra-pure water) and stock solutions have been used to calibrate and record each element in the device. All samples were analysed and data was recorded following calibration.

The samples were diluted ten times using ultra-pure nitric acid 1.0% (v/v) (Merck, Germany) and deionized water 18 micro-ohm (µΩ) in order to preserve the ICPMS sensitivity. To filter the sample, a 0.45 µm nylon syringe filter was employed. Using a pipette, 1 mL of the undiluted sample was taken and put into a test tube along with 9 mL of dilution liquid (Milli-Q® water). As a result, the fluid was diluted ten times. Prior to the analysis, seven standard solutions (0, 10, 20, 50, 100, 200 and 300 µg/L) in final volumes of 100 mL and one blank (Milli-Q® water) were created in order to calibrate the instrument. The blank solution and several metal stock solutions were used to finish the instrument's calibration.

The calculated volumes of each metal were added to 100 ml flasks and the flasks were filled with deionized water. These volumes were then used to compute the metal concentration in a linear graph. With a correlation value (r) of 0.999, this graph demonstrated that the outcome was satisfactory.

Data Analysis

The main characteristics of heavy metals in the atmospheric environment of Paka and Gebeng Industrial Areas for the southwest and northeast monsoons in this region were conducted using descriptive analysis. The analysis including mean, minimum, maximum and standard deviation are included in the descriptive statistic. All heavy metals were tested for significant differences between sampling sites using the analysis of variance (ANOVA) technique.

Principal Component Analysis (PCA) was utilized to identify the primary causes of air pollution associated with heavy metals in the research area by removing unnecessary parameters with little impact on the original data [11–12]. PCA was selected as a statistical tool due to its capacity to display the most important factors that may have contributed to the research area and to pinpoint the pollutant source. Principal Component Analysis (PCA) was applied to identify key patterns in the dataset by reducing dimensionality while preserving important variations. The method involved computing eigenvalues and eigenvectors from the covariance matrix to extract principal components. The relevant equations, including the variance-covariance matrix and transformation formula, provide a mathematical foundation for this process.

Theoretically, PCA generates latent variables, or scores, which are linear combinations of the heavy metal concentrations' original data components. Latent variables with an eigenvalue greater than 1.0 are deemed significant in PC analysis for heavy metals. XLSTAT version 2019 was used to statistically compute and evaluate the data. PCA can be expressed based on the following equation:

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + a_{im}x_{mj}$$

where, z is the component score, a is a component loading, x is the measured heavy metal variables, i is the component number and m is the total of heavy metal variables. PCA will produce the list of variables that are dominant from less significant until the most significant variables that influence positive maximum and negative minimum peak of the spatial variation pattern.

Sensitivity Analysis (SA) was used to validate the test, with the main criterion utilized being the R_2 value determined from ANN predictions. After PCA was performed to identify the sources of pollution which had an effect on the atmosphere of the study area. This was done by utilizing the 'leave-one-out' method [13-14]. The sensitivity analysis was conducted to assess how variations in input data influence the model's output. By systematically adjusting key parameters, we evaluated their impact on pollutant concentrations and risk assessments. This approach ensures that the model remains robust and reliable under different environmental conditions.

A prediction model was created using an Artificial Neural Network (ANN). In order to make predictions and identify the most important factors affecting heavy metals, RH, and API, this study uses a back-propagation ANN. The three layers that comprise this model are the input layer, hidden layer and output layer. The output layer (independent test set) includes a single node, but the hidden layer has nodes and calculates the total number of inputs (8 parameters). The Heath Index values are the output layers.

JMP Pro v. 16.0 software was used to generate an ANN model utilizing the data from air sampling for heavy metals. As the model technique, holdback with TanH sigmoid activation was used [15-16]. A combination of linear and independent variables, each with its own intercept term, make up the neural network equation. The formula was developed from the data processing within ANN.

Model Performance Determination

In order to measure the accuracy of the predictions made with the model when compared to the experimental values, three criteria were used: the coefficient of determination, root mean square error and the Nash-Sutcliffe efficiency coefficient [17-19]. The Nash-Sutcliffe Efficiency Coefficient (NSC) can be utilized to measure the accuracy of hydrological models when forecasting. This metric ranges from negative infinity to 1, with 1 indicating a perfect match between predicted data and observed data. A score of 0 means that the model's prediction is as reliable as the observed data's mean and a score below 0 is registered when the observed mean is more effective at forecasting than the model.

Health Risk Assessment

The model utilized to quantify the exposure of humans to heavy metals is based on the one established by the Environmental Protection Agency of the United States. According to the Exposure Factors Handbook, the daily dose (D) of pollutants, measured in milligrams per kilogram per day (mg/kg/day), that can be obtained through ingestion, dermal exposure and inhalation.

Validation of Health Index Prediction Model

In order to validate the Health Index Prediction Model (HIPM) developed, the formula used in health risk assessment was employed to determine the actual Health Index. Afterwards, the actual HIPM was compared with the predicted HIPM to evaluate the validity of the model. The comparison between the actual and predicted HIPM results was then obtained.

The Artificial Neural Network (ANN) methodology employed in this study is designed to ensure transparency, reproducibility, and model reliability. The network architecture comprises an input layer with neurons representing key environmental variables such as temperature, humidity, wind speed, and pollutant concentrations, followed by one or more hidden layers with an optimized number of neurons determined through empirical tuning or established heuristics. Bias handling is explicitly addressed by specifying whether biases are included and their initialization strategy. To prevent overfitting, regularization techniques such as dropout, early stopping, and L1/L2 regularization are applied to enhance model generalization. The learning process is governed by an optimization algorithm. By systematically outlining these methodological components, the ANN framework ensures robust pollutant dispersion modeling and enhances the accuracy of air quality predictions.

Results and Discussion

Normality Test

The normality of the data was assessed using the Jarque-Bera test [20]. This statistical test determines whether the series follows a normal distribution. The analysis revealed that the data did not exhibit normal distribution, as evidenced by a p-value (<0.0001) lower than the significance level alpha (0.05). As a result, the alternative hypothesis (H_a) was accepted rather than the null hypothesis (H_0), showing that the extracted variable did not follow a normal distribution. Data transformation was used to solve this problem and stop any variable from controlling performance [21].

Kaiser-Meyer-Olkin (KMO) and Bartlett's Tests

KMO and Bartlett's tests were run before the analysis to determine whether the data were appropriate for PCA and to assess its sufficiency [18]. All metrics in this study were used for PCA, with the exception of NO and NOx, which had a significant amount of missing data. Although a small percentage of missing data can be handled using PCA [22], a high number of missing data can cause the analysis to produce incorrect conclusions. Furthermore, PCA is susceptible to missing data brought on by a lack of data for particular parameters [23]. The KMO test was used to evaluate the samples' suitability, which may have been affected by underlying causes [24]. The KMO sample adequacy measure is shown in Table 4. For the KMO test, a value larger than 0.5 was chosen as the reference point [25]. As a result of the KMO values exceeding 0.5, which indicate adequate data for PCA extraction, the findings obtained demonstrated that the measure of sampling adequacy (MSA) was acceptable.

Table 4. Kaiser-Meyer-Olkin measure of sampling adequacy

PM_{10}	0.537
$PM_{2.5}$	0.541
SO_2	0.829
NO_2	0.625
O_3	0.720
CO	0.679
KMO	**0.612**

The Bartlett's sphericity test was performed in addition to KMO to assess the parameters' association with the suitability of the data for PCA construction. With a significance level of 0.05, the test result revealed a p-value less than 0.0001. As a result, the data were eligible for multivariate analysis because there was enough information to perform PCA and a significant correlation between the parameters.

As the p-value was lower than the significance level of 0.05, the null hypothesis (H_0) was rejected in favour of the alternative hypothesis (H_a). This means that at least one of the correlations between parameters was significantly different from zero. Consequently, the air quality parameters were found to be correlated and not orthogonal, allowing for a diverse interpretation of the data's variability.

Sub-index of Five Pollutants

The sub-index of five air pollutants (criteria pollutants) is calculated. The API reading can be ascertained straight from the PM10 number if it is less than 50. To find the API reading, however, a calculation is necessary if the PM10 number is higher than 50 (Table 5). The sub-index calculation translates pollutant levels into an Air Pollution Index (API) to assess air quality. Each pollutant has set concentration ranges, and a formula is applied based on where the value falls. This method ensures a clear and consistent way to measure and communicate pollution risks.

Table 5. The sub-index computation of five air contaminants (criteria contaminants)

Contaminants	Time (Average)	Concentration (Average)	Formula
CO	8-hr	c. < 9ppm	API = c. x 11.11111
		9 < c. < 15	API = 100 + {[c. - 9] x 16.66667}
		15 < c. < 30	API = 200 + {[c. - 15] x 6.66667}
		c. > 30ppm	API = 300 + {[c. - 30] x 10}
O_3	1-hr	c. < 0.2ppm	API = c. x 1000
		0.2 < c.< 0.4	API = 200 + {[c. - 0.2] x 500}
		c. > 0.4ppm	API = 300 + {[c. - 0.4] x 1000}
NO_2	1-hr	c. < 0.17ppm	API = c. x 588.23529
		0.17 < c. < 0.6	API = 100 + {[c. - 0.17] x 232.56}
		0.6 < c. < 1.2	API = 200 + {[c. - 0.6] x 166.667}
		c. > 1.2ppm	API = 300 + {[c. - 1.2] x 250}
PM10	24-hr	c. < 50µg.m-3	API = c.
		50 < c. < 350	API = 50 + {[c. - 50] x 0.5}
		350 < c. < 420	API = 200 + {[c. - 350] x 1.4286}
		420 < c. < 500	API = 300 + {[c. - 420] x 1.25}
		c. > 500µg.m-3	API = 400 + {[c. - 500]}
SO_2	24-hr	c. < 0.04ppm	API = c. x 2500
		0.04 < c. < 0.3	API = 100 + {[c. - 0.04] x 384.61}
		0.3 < c. < 0.6	API = 200 + {[c. - 0.3] x 333.333}
		c. > 0.6ppm	API = 300 + {[c. - 0.6] x 500}

Note: c. = concentration. Source: Department of Environment, DOE [19]

Principal Component Analysis

PCA was used to identify the environmental factors that are the main driven of the development of separate groups and to minimize the dimensionality of the data set. Furthermore, PCA was used to distinguish the key parameters which could be implemented as inputs for the prediction model concerning six elements namely Pb, Fe, As, Zn, Cd and Cu. PCA identifies the main patterns in the dataset by finding linear combinations of the original variables, called principal components (PCs). These PCs explain the maximum amount of variance in the data. Factor scores, or PC scores, show how each variable contributes to the PCs, while factor loadings represent the correlation between the variables and PCs. The highest factor scores or loadings for the PCs that explain most of the variance.

The selection of principal components was based on eigenvalues, with those capturing the most variance retained for analysis. Each component's significance was assessed through factor loadings, ensuring meaningful representation of the data. This approach highlights key patterns while reducing dimensionality for clearer interpretation. The three main components (PCs) were extracted from the PCA of the petrochemical industrial region using eigenvalues greater than 1 (>1.0). In terms of the overall variation, the chosen PCs for the southwest and northeast contributed 86.939% and 76.444%, respectively. Heavy metals including As, Fe, Pb, Cu, Cd and Zn had the highest communality values for the chosen PCs in the southwest monsoon, with respective values of 95.4%, 88.6%, 86.9%, 84.6%, 83.4% and 82.7% (Table 6). There are two categories of highest and moderate for the communality values of the PCs selected in the northeast. According to Table 7, the highest communality values were found for As (91.2%), Fe (86.8%) and Pb (86.2%), whereas the intermediate values were found for Cu (73.9%), Cd (69.6%) and Zn (50.9%). These separate communalities served to provide information on how well the PCA model performed for the different variables, which further supports the number of PCs that was chosen. Each principal component was linked to pollution sources or health risks. Factor loadings helped identify whether they came from traffic, industry, or other sources. This makes it easier to understand how pollutants affect air quality and health.

Table 6. Varifactors after varimax rotation and identifying of heavy metals in southwest monsoon

Parameter	Component			Communality
	RPC1	RPC 2	RPC 3	
Fe	-0.055	**0.940**	-0.003	0.886
Cu	0.446	**0.803**	0.055	0.846
Pb	**0.857**	-0.100	0.353	0.869
Cd	**0.872**	0.253	0.097	0.834
As	0.052	0.045	**0.975**	0.954
Zn	**0.846**	0.165	-0.290	0.827
Value for Eigan	2.752	1.350	1.115	
Variance (%0	45.867	22.496	18.577	
Cumulative variance (%)	45.867	68.363	86.939	

Table 7. Varifactors after varimax rotation and identifying of heavy metals in northeast monsoon

Parameter	Component			Communality
	RPC1	RPC 2	RPC 3	
Fe	0.916	-0.096	0.139	0.868
Cu	-0.278	0.812	0.044	0.739
Pb	0.060	0.105	0.921	0.862
Cd	-0.056	0.560	-0.616	0.696
As	0.955	0.010	-0.012	0.912
Zn	0.400	0.591	-0.014	0.509
Value for Eigan	2.0714	1.4679	1.0474	
Variance (%0	34.524	24.465	17.456	
Cumulative variance (%)	34.524	58.988	76.444	

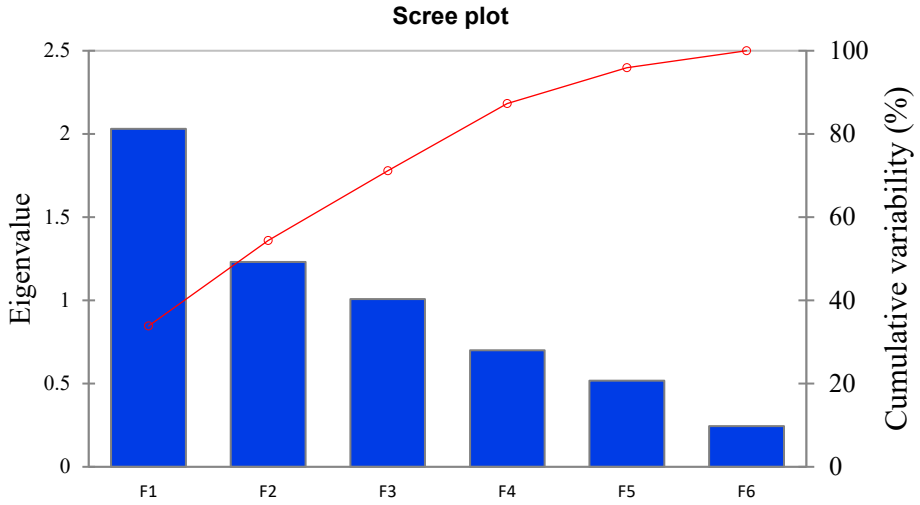

Figure 2. Scree Plots for PCA

Sensitivity Analysis

The test done by PCA towards the identifying of the sources of pollution which contributed to the atmosphere in the study area are confirmed by using SA [26-27]. Table 8 had detailed out the SA calculation. In order to obtain these results, the leave-one-out technique was initiated. One of the variables used as input for the ANN is taken out from the list of independent variables, while the other variables are left to be included as input neurons.

Overall, the R^2 result for API (all parameter) for Malaysia was 0.293. In order to estimate the contribution of each pollutant, the R^2 result was subtracted from API for each pollutant. The resulting differences in R^2 values ranged from 0.195 to 0.293. The highest result belonged to As, which contributed 32.03 % to the pollution level in the petrochemical area during both monsoons. On the other hand, Cd and Zn had the lowest percentage contribution of 0.0% each, even though they still had some effect on API and health. Even though Cd and Zn had a low percentage contribution, both of the metals can still impact the Air Pollution Index (API) and health through their toxicity and long-term exposure risks. As is released in large amounts from industries, coal burning, and petrochemical activities. It easily spreads in the environment and builds up in the human body, making it a major concern for air quality and health. As a result, it has a stronger impact on pollution levels and associated risks rather than other metals.

Table 8. Sensitivity analysis in identifying the sources of heavy metals in the study area

Parameters	R^2	Difference R^2	% Contribution
API	0.293		
Fe	0.201	0.092	30.07
Cu	0.263	0.030	9.80
Pb	0.207	0.086	28.16
Cd	0.293	0.0	0.0
As	0.195	0.098	32.03
Zn	0.293	0.0	0.0
		0.306	100

Health Risk Assessment in Paka and Gebeng

The values of HQ and HI parameters for Paka (Southwest Monsoon) are listed in Table 9, Paka (Northeast Monsoon) (Table 10), Gebeng (Southwest Monsoon) (Table 11) and Gebeng (Northeast Monsoon) (Table 12). The similar trend of the results of HQ and HI are showed for both children and adults. The values of HQ for all three pathways are decreased in the order of ingestion>dermal contact>inhalation. The contribution of HQIng to HI (the total risk) is highest compared to another health quotient. HQ values change over time, either rising or falling depending on factors like the environment, policies, or economic shifts.

Overall, the HI value decreased in the following order of Fe>Pb>Cd>As>Cu>Zn. Although some of the assumptions used for modelling were rather straightforward and basic, the health risks of the six metals that were studied fall within an acceptable range. Nearly all of the metals under study had HQ and HI values below the safe threshold of 1 for both adults and children. Zn and Cu scores were the lowest, but Fe and Cd HI values were near the permissible threshold. It should be mentioned that children's HI values for the metals were actually higher than adults. The higher HI values in children indicate greater susceptibility to pollution due to their developing physiology and higher intake relative to body weight.

The reference dose (RfD) is an estimate of the maximum tolerable risk to human health from daily exposure over the long term, taking into accounts exposures to more susceptible individuals. This threshold can be used to determine whether an exposure is likely to cause adverse health effects. Lower than average daily doses (D) are generally considered safe while higher than average daily doses may increase the risk of adverse health effects. Hazard Quotients (HQ) can be used to measure this risk, and multiple estimates can be combined to calculate the Hazard Index (HI), which can be used to measure the risk of mixed metal contaminants and calculate the Health Index (HI).

$$HI = \sum_{i=1}^{3} HQi$$

According to the proportion of contribution, the respective contributions of the two parameters (Fe and Pb) following As are 30.07% and 28.16%. Conversely, Cu made a 9.8% contribution. According to the study, the identifying sources of heavy metals were As>Fe>Pb>Cu>Cd>Zn, in descending order. Fe and Pb had high levels suggest ongoing emissions that can harm air quality and health.

Heavy metals in the atmosphere pose significant risks to both environmental and human health, with Fe, Pb, and As contributing the most to pollution. Their presence is often linked to industrial activities,

vehicle emissions, and combustion processes, leading to long-term air quality deterioration. Exposure to these metals, especially through ingestion and inhalation, can cause serious health effects, including respiratory issues, neurological disorders, and developmental problems in children. While some metals like Zn and Cu have lower risk levels, the persistence of Fe and Cd near the safety threshold highlights the need for stricter pollution controls. Addressing these emissions is crucial to safeguarding public health and reducing long-term ecological damage.

Although Fe and Zn are essential trace elements, their presence in the atmosphere at elevated levels can still pose environmental and health concerns. Industrial emissions, vehicle exhaust, and metal processing can release these elements in forms that contribute to air pollution. Inhalation or ingestion of excess Fe and Zn particles may lead to respiratory irritation and other health effects, particularly in sensitive individuals. Their inclusion in the study is important to assess their role in overall pollution levels and potential interactions with more toxic heavy metals. Understanding their atmospheric behavior helps provide a more comprehensive picture of air quality and associated risks.

Table 9. Health risk from heavy metals for Paka (Southwest season)

Concentration (mg/(kg.d))		HQ_{Ing}		HQ_{Inh}		HQ_{Derm}		HI	
		Children	Adult	Children	Adult	Children	Adult	Children	Adult
As	0.0060	0.55×10^{-4}	0.74×10^{-5}	0.85×10^{-9}	0.92×10^{-9}	0.17×10^{-5}	0.13×10^{-5}	0.57×10^{-4}	0.86×10^{-5}
Cd	0.0146	0.19×10^{-3}	0.25×10^{-4}	0.35×10^{-8}	0.38×10^{-8}	0.34×10^{-4}	0.25×10^{-4}	0.22×10^{-3}	0.50×10^{-4}
Cu	0.0225	0.78×10^{-5}	0.10×10^{-5}	0.14×10^{-9}	0.15×10^{-9}	0.14×10^{-5}	0.10×10^{-5}	0.92×10^{-5}	0.21×10^{-5}
Fe	1.0900	0.52×10^{-1}	0.69×10^{-2}	0.13×10^{-7}	0.14×10^{-7}	0.13×10^{-3}	0.94×10^{-4}	0.52×10^{-1}	0.70×10^{-2}
Pb	0.1950	0.71×10^{-3}	0.96×10^{-4}	0.13×10^{-7}	0.14×10^{-7}	0.43×10^{-4}	0.32×10^{-4}	0.75×10^{-3}	0.13×10^{-3}
Zn	0.2767	0.12×10^{-4}	0.16×10^{-5}	0.22×10^{-9}	0.24×10^{-9}	0.53×10^{-6}	0.40×10^{-6}	0.12×10^{-4}	0.20×10^{-5}

Table 10. Health risk from heavy metals for Paka (Northeast season)

Concentration (mg/(kg.d))		HQ_{Ing}		HQ_{Inh}		HQ_{Derm}		HI	
		Children	Adult	Children	Adult	Children	Adult	Children	Adult
As	0.0020	1.83×10^{-5}	2.45×10^{-6}	2.85×10^{-10}	3.06×10^{-10}	5.61×10^{-7}	4.20×10^{-7}	1.88×10^{-5}	2.87×10^{-6}
Cd	0.0120	1.53×10^{-4}	2.06×10^{-5}	2.91×10^{-9}	3.12×10^{-9}	2.76×10^{-5}	2.06×10^{-5}	1.81×10^{-4}	4.12×10^{-5}
Cu	0.0040	1.38×10^{-6}	1.85×10^{-7}	2.41×10^{-11}	2.59×10^{-11}	2.42×10^{-7}	1.81×10^{-7}	1.62×10^{-6}	3.66×10^{-7}
Fe	0.2000	9.48×10^{-3}	1.27×10^{-3}	2.38×10^{-9}	2.56×10^{-9}	2.30×10^{-5}	1.72×10^{-5}	9.50×10^{-3}	1.29×10^{-3}
Pb	0.4440	1.62×10^{-3}	2.18×10^{-4}	3.07×10^{-8}	3.27×10^{-8}	9.73×10^{-5}	7.26×10^{-5}	1.72×10^{-3}	2.90×10^{-4}
Zn	0.1833	7.80×10^{-6}	1.05×10^{-6}	1.48×10^{-10}	1.59×10^{-10}	3.52×10^{-7}	2.62×10^{-7}	8.15×10^{-6}	1.31×10^{-6}

Table 11. Health risk from heavy metals for Gebeng (Southwest season)

Concentration (mg/(kg.d))		HQ_{Ing}		HQ_{Inh}		HQ_{Derm}		HI	
		Children	Adult	Children	Adult	Children	Adult	Children	Adult
As	0.0060	5.48×10^{-5}	7.36×10^{-6}	8.53×10^{-10}	9.18×10^{-10}	1.68×10^{-6}	1.26×10^{-6}	5.65×10^{-5}	8.61×10^{-6}
Cd	0.0146	1.87×10^{-4}	2.50×10^{-5}	3.54×10^{-9}	3.79×10^{-9}	3.36×10^{-5}	2.50×10^{-5}	2.21×10^{-4}	5.00×10^{-5}
Cu	0.0225	7.78×10^{-6}	1.04×10^{-6}	1.36×10^{-10}	1.46×10^{-10}	1.36×10^{-6}	1.02×10^{-6}	9.15×10^{-6}	2.06×10^{-6}
Fe	1.0900	5.15×10^{-2}	6.93×10^{-3}	1.30×10^{-8}	1.39×10^{-8}	1.25×10^{-4}	9.35×10^{-5}	5.16×10^{-2}	7.02×10^{-3}
Pb	0.1950	7.11×10^{-4}	9.57×10^{-5}	1.34×10^{-8}	1.44×10^{-8}	4.27×10^{-5}	3.18×10^{-5}	7.54×10^{-4}	1.28×10^{-4}
Zn	0.2767	1.18×10^{-5}	1.58×10^{-6}	2.23×10^{-10}	2.40×10^{-10}	5.30×10^{-7}	3.95×10^{-7}	1.23×10^{-5}	1.98×10^{-6}

Table 12. Health risk from heavy metals for Gebeng (Northeast season)

	Concentration (mg/(kg.d))	HQ_{Ing}		HQ_{Inh}		HQ_{Derm}		HI	
		Children	Adult	Children	Adult	Children	Adult	Children	Adult
As	0.0020	1.83×10^{-5}	2.45×10^{-6}	2.85×10^{-10}	3.06×10^{-10}	5.61×10^{-7}	4.20×10^{-7}	1.88×10^{-5}	2.87×10^{-6}
Cd	0.0070	8.95×10^{-5}	1.20×10^{-5}	1.70×10^{-9}	1.82×10^{-9}	1.61×10^{-5}	1.20×10^{-5}	1.06×10^{-4}	2.40×10^{-5}
Cu	0.0190	6.57×10^{-6}	8.81×10^{-7}	1.14×10^{-10}	1.23×10^{-10}	1.15×10^{-6}	8.58×10^{-7}	7.72×10^{-6}	1.74×10^{-6}
Fe	0.7567	3.58×10^{-2}	4.81×10^{-3}	9.01×10^{-9}	9.70×10^{-9}	8.71×10^{-5}	6.49×10^{-5}	3.59×10^{-2}	4.88×10^{-3}
Pb	0.1123	4.11×10^{-4}	5.51×10^{-5}	7.73×10^{-9}	8.30×10^{-9}	2.46×10^{-5}	1.83×10^{-5}	4.36×10^{-4}	7.35×10^{-5}
Zn	0.2100	8.93×10^{-6}	1.20×10^{-6}	1.70×10^{-10}	1.82×10^{-10}	4.03×10^{-7}	3.00×10^{-7}	9.34×10^{-6}	1.50×10^{-6}

Prediction Model

The relative importance of the parameters in an artificial neural network (ANN) was determined by the magnitude and direction of their effect on the output. In this study, eight variables had been used as input variables and one layer of hidden node had applied to compute two prediction model in HIc (Health Index prediction for children) and HIa (Health Index prediction model for adults).

The results of an ANN model with two different models (HIc and HIa) are summarised in Table 13. Both models consist of overall, northeast and southwest training sets. R^2 and -Loglikelihood values are listed for three different sets of each model, including Training, Validation and Testing. The greater value of R^2, means the better and the model's fit to the data. In this study, HIa model shows the higher R^2 in training, validation and testing set as compared to HIc models. HIa performed better than HIc in validation and testing because it more accurately captured exposure risks, likely due to better parameter selection or dataset alignment. However, both models revealed fit and accurate data as R^2 in the range of 0.70-1.00.

This prediction model is essential for assessing long-term health risks associated with air pollution, providing a systematic approach to quantify exposure impacts. It helps policymakers and researchers understand pollution trends, enabling proactive mitigation strategies. Without such a model, health risks may be underestimated, delaying necessary interventions to protect public health. The model should be applied in areas experiencing industrial emissions, urban pollution, or changing environmental conditions to evaluate potential health hazards. By incorporating real-time data, it can offer timely assessments for both short-term and chronic exposure risks. Its flexibility allows adaptation to different regions with varying pollution sources and population demographics. NSC is used to compare the variance of a model's residuals to the measured data. If the NSC value is close to 1, it indicates that the model predictions perfectly match the observed data. By contrast, a lower value suggests that the observed mean is a more reliable predictor than the model. Values less than 0 indicate that the model does not perform better than the mean of the observed data. In other words, an efficiency of less than zero (NSC < 0) occurs when the observed mean is a better forecaster than the model. Principally, the closer the model efficiency is to 1, the more accurate the model.

Table 13. R^2 for HIc and HIa prediction model

	HIc		HIa	
	R2	-Loglikelihood	R2	-Loglikelihood
Overall	1.0000	-2749.362	0.9999	-2968.98
Training	0.9998	-1355.204	0.9999	-1561.012
Validation	0.9406	-555.150	0.7821	-287.075
Testing	1.0000	-2749.362	0.9999	-2968.98
Northeast				
Training	1.0000	-1335.047	1.0000	-1721.135
Validation	0.8821	-652.8114	1.0000	-909.1749
Testing	0.8439	-228.994	0.9929	-197.4187
Southwest				
Training	1.0000	-1335.047	1.0000	-1524.145
Validation	0.8821	-653.8114	1.0000	-811.5004
Testing	0.8365	-284.8571	0.8732	-181.0782

The results show that both HIc and HIa models perform well in predicting health risks, but HIc is generally more accurate. The R^2 values indicate how well the models fit the data, with HIc showing higher accuracy, especially in validation and testing phases. This suggests HIc provides more reliable predictions across different conditions. For the northeast monsoon, HIa performed exceptionally well in validation, achieving a perfect R^2 (1.0000), but this could indicate overfitting, meaning it may not work as well in other situations. On the other hand, HIc remained stable across different data sets, showing consistent performance. For the Southwest monsoon, HIc again showed better performance in testing and validation. The -Loglikelihood values, which measure model fit, were lower for HIa in some cases, indicating a less optimal fit. Overall, these results suggest that HIc is the more reliable model for predicting health risks.

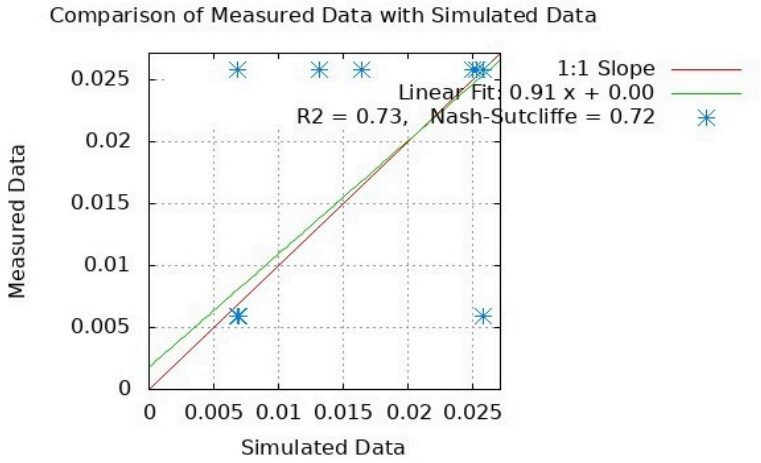

Figure 3. HIPM in children utilizing Linear Fit and Nash-Sutcliffe analysis comparing actual and simulated data

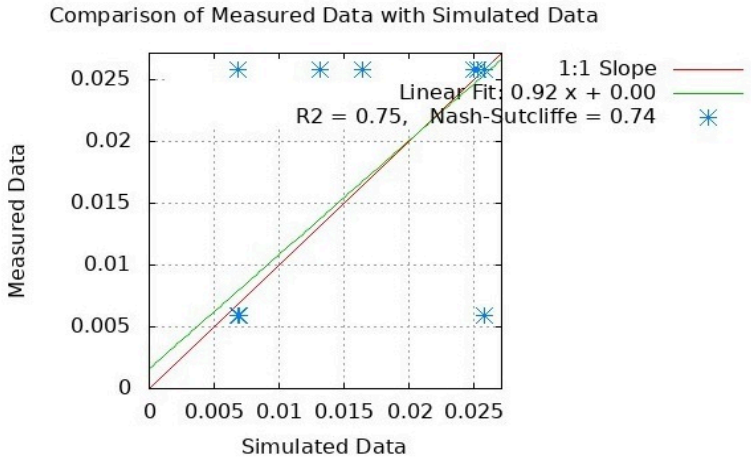

Figure 4. HIPM in adults utilizing Linear Fit and Nash-Sutcliffe analysis comparing actual and simulated data

The predicted HIPM equation model was developed as shown in Table 14. It displays an overview of the ANN-derived equations for the HI prediction model for HIc and HIa. The input parameters acquired using PCA were used to develop the equations. The health index pertaining to the concentration of heavy metals in the petrochemical industrial region can be predicted using these formula.

Table 14. Equation derived from ANN for the prediction model of health index

Health Index	Models (ANN)
Children	HIc Predicted = 0.0095*(TANH(0.5*(28.4389*Fe)+(4795.1207*Cu)+(45.246*Pb)-(672.6587*Cd)-(7947.7098*As)+(18.9376*Zn)-(0.1168*RH)+(0.8102*API)-30.8318))+0.0163 ± SD
Adults	HIa Predicted = 0.0010*(TANH(0.5*(10.3135*Fe)+(29.4515*Cu)+(13.8422*Pb)-(116.9227*Cd)-(534.9393*As)+(24.6396*Zn)+(0.0230*RH)+(0.0524*API)-7.8073))+0.0027 ± SD

Conclusions

In conclusion, the regional variability of the concentration of specific heavy metal dust in the atmosphere was ascertained using descriptive analysis. The origins of heavy metals in the air dust surrounding the research area were determined by using PCA. Sensitivity analysis was used to confirm the sources of heavy metals found in PCA. The sensitivity analysis with leave-one-out technique was carried out to identify the significant input-output relationship and to identify the most significant heavy metals parameters that contribute to the health risk in petrochemical industrial area, Malaysia.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgement

This project was supported by Universiti Sultan Zainal Abidin (UniSZA) under Dana Penyelidikan UniSZA (UniSZA/2016/DPU/01) and Geran Dalaman Universiti (UniSZA/2024/PSU-TDP/13).

References

[1] Salthammer, T., Schieweck, A., Gu, J., Ameri, S., & Uhde, E. (2018). Future trends in ambient air pollution and climate in Germany – Implications for the indoor environment. *Building and Environment, 143*, 661–670. https://doi.org/10.1016/j.buildenv.2018.07.038

[2] Shao, L., Liu, P., Jones, T., Yang, S., Wang, W., Zhang, D., Yang, C.-X., Xing, J., Hou, C., Zhang, M., Feng, X., Li, W., & BeruBe, K. (2022). A review of atmospheric individual particle analyses: Methodologies and applications in environmental research. *Gondwana Research, 110*, 347–369. https://doi.org/10.1016/j.gr.2021.06.013

[3] Adani, M., Mircea, M., D'Isidoro, M., Costa, M. P., & Silibello, C. (2015). Heavy metal modelling study over Italy: Effects of grid resolution, lateral boundary conditions and foreign emissions on air concentrations. *Water, Air, & Soil Pollution, 226*(1), 46. https://doi.org/10.1007/s11270-015-2324-7

[4] Ripin, S. N. M., Hasan, S., Kamal, M. L., & Hashim, N. M. (2014). Analysis and pollution assessment of heavy metal in soil, Perlis. *Malaysian Journal of Analytical Sciences, 18*(1), 155–161.

[5] Mitra, S., Chakraborty, A. J., Tareq, A. M., Emran, T. B., Nainu, F., Khusro, A., Idris, A. M., Khandaker, M. U., Osman, H., Alhumaydhi, F. A., & Simal-Gandara, J. (2022). Impact of heavy metals on the environment and human health: Novel therapeutic insights to counter the toxicity. *Journal of King Saud University - Science, 34*(3), 101865. https://doi.org/10.1016/j.jksus.2022.101865

[6] Aksu, A. (2015). Source of metal pollution in the urban atmosphere (A case study: Tuzla, Istanbul). *Journal of Environmental Health Science and Engineering, 13*(1), 79. https://doi.org/10.1186/s40201-015-0220-9

[7] U.S. Environmental Protection Agency. (2024). *National Ambient Air Quality Standards (NAAQS) for PM*. https://www.epa.gov/pm-pollution/national-ambient-air-quality-standards-naaqs-pm

[8] Nurhaini, F. F., Lestiani, D. D., Syahfitri, W. Y. N., Kusmartini, I., Sari, D. K., Kurniawati, S., & Santoso, M. (2023). Determination of nutrient and toxic elements in food reference materials by suspension preparation and TXRF analysis. *International Food Research Journal, 30*(2), 463–471.

[9] Xing, G., Sardar, M. R., Lin, B., & Lin, J. M. (2019). Analysis of trace metals in water samples using NOBIAS cheate resins by HPLC and ICP-MS. *Talanta, 204*, 50–56. https://doi.org/10.1016/j.talanta.2019.05.049

[10] Canizo, B. V., Escudero, L. B., Pérez, M. B., Pellerano, R. G., & Wuilloud, R. G. (2018). Intra-regional classification of grape seeds produced in Mendoza province (Argentina) by multi-elemental analysis and chemometrics tools. *Food Chemistry, 242*, 272–278. https://doi.org/10.1016/j.foodchem.2017.09.047

[11] Ismail, A. S., Abdullah, A. M., & Samah, M. A. A. (2017). Environmetric study on air quality pattern for assessment in northern region of Peninsular Malaysia. *Journal of Environmental Science and Technology,*

10(3), 186–196.
[12] Azid, A., Rani, N. A. A., Samsudin, M. S., Khalit, S. I., Gasim, M. B., Kamarudin, M. K. A., Yunus, K., Saudi, A. S. M., & Yusof, K. M. K. K. (2017). Air quality modelling using chemometric techniques. *Journal of Fundamental and Applied Sciences, 9*(2S), 443–466. https://doi.org/10.4314/jfas.v9i2s.32
[13] Azid, A., Juahir, H., Toriman, M. E., Endut, A., Rahman, M. N. A., Kamarudin, M. K. A., Latif, M. T., Saudi, A. S. M., Hasnam, C. N. C., & Yunus, K. (2016). Testing and evaluation selection of the most significant variables of air pollutants using sensitivity analysis. *Journal of Testing and Evaluation, 44*(1), 376–384. https://doi.org/10.1520/JTE20140294
[14] Xiao, S., Oladyshkin, S., & Nowak, W. (2020). Forward-reverse switch between density-based and regional sensitivity analysis. *Applied Mathematical Modelling, 84*, 377–392. https://doi.org/10.1016/j.apm.2020.02.028
[15] Feenstra, D., Molotnikov, A., & Birbilis, N. (2021). Utilisation of artificial neural networks to rationalise processing windows in directed energy deposition applications. *Materials & Design, 198*, 109342. https://doi.org/10.1016/j.matdes.2020.109342
[16] Gotwalt, C. M., Jones, B. A., & Steinberg, D. M. (2009). Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics, 51*(1), 88–95. https://doi.org/10.1198/tech.2009.0011
[17] Ma, L., Hurtado, A., Eguilior, S., & Borrajo, J. F. L. (2016). A model for predicting organic compounds concentration change in water associated with horizontal hydraulic fracturing. *Science of the Total Environment, 625*, 1164–1174. https://doi.org/10.1016/j.scitotenv.2017.12.319
[18] Hutchins, M. G., Abesser, C., Prudhomme, C., Elliott, J. A., Bloomfield, J. P., Mansour, M. M., & Hitt, O. E. (2018). Combined impacts of future land-use and climate stressors on water resources and quality in groundwater and surface water bodies of the upper Thames river basin, UK. *Science of the Total Environment, 631*, 962–986. https://doi.org/10.1016/j.scitotenv.2018.03.028
[19] Barzegar, R., Moghaddam, A. A., Adamowski, J., & Ozga-Zielinski, B. (2018). Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. *Stochastic Environmental Research and Risk Assessment, 32*(3), 799–813. https://doi.org/10.1007/s00477-017-1404-y
[20] Kim, N. (2016). A robustified Jarque–Bera test for multivariate normality. *Economics Letters, 140*, 48–52.
[21] Lee, D. (2020). Data transformation: A focus on the interpretation. *Korean Journal of Anesthesiology, 73*(6), 503–508.
[22] van Ginkel, J. R., van der Ark, L. A., Emons, W. H. M., & Meijer, R. R. (2023). Handling missing data in principal component analysis using multiple imputation. In *Essays on contemporary psychometrics* (pp. 141–161). Springer. https://doi.org/10.1007/978-3-031-10370-4_8
[23] Franceschi, F., Cobo, M., & Figueredo, M. (2018). Discovering relationships and forecasting PM10 and PM2.5 concentrations in Bogotá, Colombia, using artificial neural networks, principal component analysis, and k-means clustering. *Atmospheric Pollution Research, 9*(5), 912–922.
[24] Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software, 22*(4), 464–475.
[25] Ul-Saufie, A. Z., Yahaya, A. S., Ramli, N. A., Rosaida, N., & Hamid, H. A. (2013). Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principal component analysis (PCA). *Atmospheric Environment, 77*, 621–630.
[26] Department of Environment Malaysia. (2020). *New Malaysian ambient air quality standard*.
[27] Cabaneros, S. M. S., Calautit, J. K. S., & Hughes, B. R. (2017). Hybrid artificial neural network models for effective prediction and mitigation of urban roadside NO2 pollution. *Energy Procedia, 142*, 3524–3530. https://doi.org/10.1016/j.egypro.2017.12.242