Water Science & Technology

IWA PUBLISHING

Water Science & Technology Vol 87 No 7, 1686 doi: 10.2166/wst.2023.084

A health risk model for rural households based on the distribution of multi pollutants

G. V. Rathnamala [a], G. P. Shivashankara [b], R. M. Ashwini [a], H. R. Rashmi [c] and Basuraj Bhowmik [d,*]

[a] Department of Civil Engineering, GITAM School of Technology, GITAM University, Bengaluru, India
[b] Dayananda Sagar College of Engineering, Bengaluru, India
[c] Department of Civil Engineering, K.S. School of Engineering and Management, Bengaluru, India
[d] Department of Civil Engineering, Indian Institute of Technology BHU, Varanasi, India
*Corresponding author. E-mail: basuraj.civ@itbhu.ac.in

GVR, 0000-0001-5796-8897; GPS, 0000-0001-5216-0581; RMA, 0000-0002-2887-483X; BB, 0000-0001-7782-513X

ABSTRACT

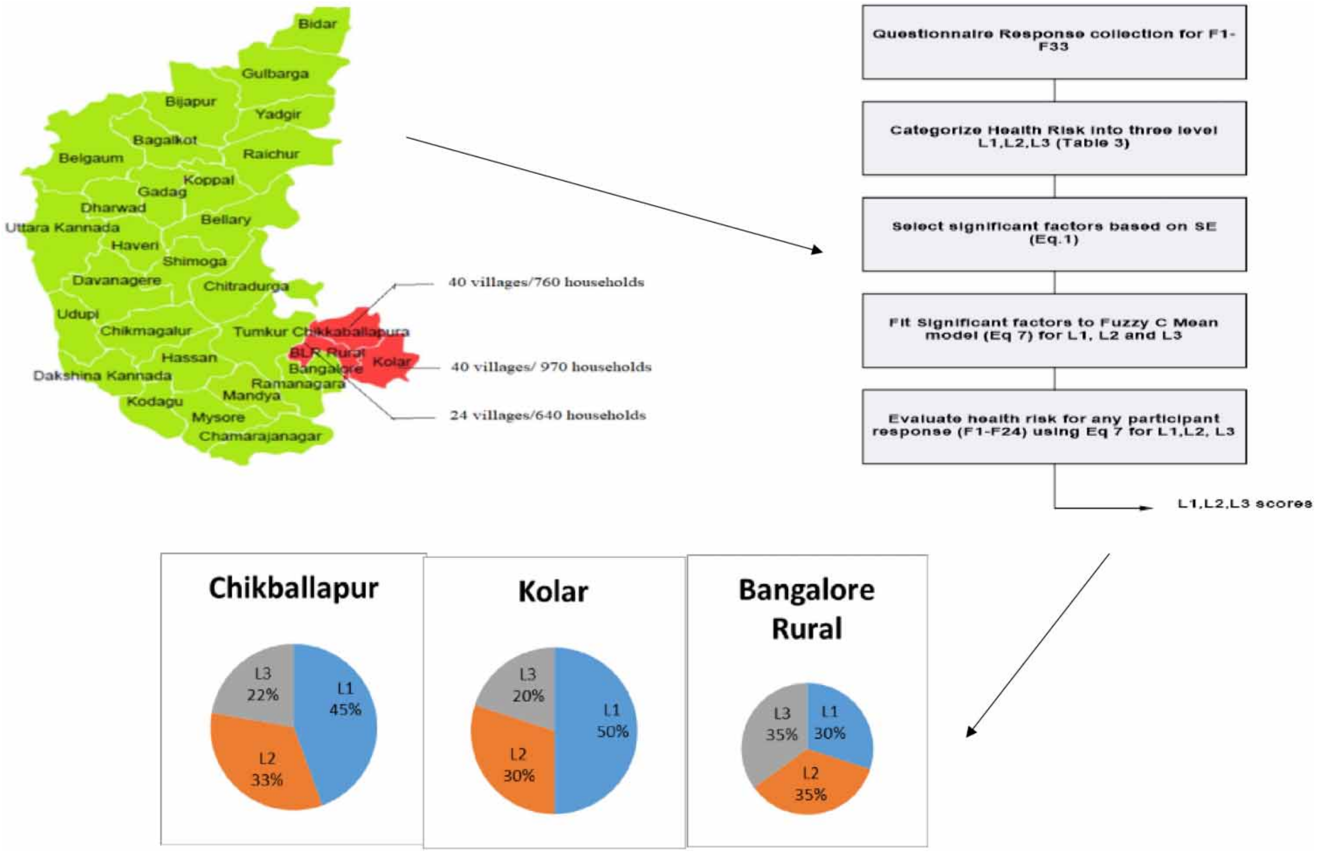
Rapid consumerism and improper waste disposal create widespread environmental degradation through the air, water sources and landfills in India's rural areas. This work develops a health risk prediction model to score villages based on quantitative and qualitative factors. Quantitative observations regarding pollutant levels and qualitative responses are collected from various households. that are risk labelled against WHO standards. The health risk model is designed to correlate the qualitative factors. A total of 2,370 rural households spread across three districts of Karnataka were selected. The study found that the health risk score predicted by the model has a higher significant correlation (0.8) to various existing pollutant factors. The study found that source of drinking water (0.87), quality of drinking water (0.81), drainage canal availability (0.72), type of drainage (0.73), stagnant water (0.71), toilet availability (0.83), maintenance frequency (0.83), cooking fuel type (0.77), cigarette use (0.71), garbage piles up (0.73) and the percentage composition of wastes (0.74) was found to have a higher positive correlation to the health of rural households. The villages with higher health risks can be identified, and suitable mitigation plans can be designed to mitigate the health risk by state authorities.

Key words: environment, health risks, pollutants, rural households

HIGHLIGHTS

- This work develops a health risk prediction model to score the villages based on quantitative and qualitative factors.
- The study found that the health risk score predicted by the model has a higher significant correlation (>0.8) to various existing pollutant factors.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Since the opening up of the economy in late 1991, India has become a global marketplace. It has increased per capita income among people. These increased income levels have accelerated consumerism. This rapid consumerism, accompanied by improper waste disposal, has resulted in environmental degradation. Though the degradation was more pronounced in cities, with rapid economic expansion, the degradation is also evident in villages. Environmental degradation through air, water and water pollution exposes people to various health risks. Environmental degradation and damage to public health are essential constraints in sustainable economic growth and social development. There are three significant pollution receipts found in villages: (i) air pollution due to dispersed particles, hydrocarbons, CO, CO_2, NO, NO_2, SO_3, etc., (ii) water pollution due to organic, inorganic and biological discharges at high levels and (iii) soil pollution through the release of chemicals, heavy metals, hydrocarbons and pesticides.

Studies on the effect of pollution on human health have become a global research interest over the last decade. They have proposed various assessment methodologies to reduce the chances of significant uncertainties (Cohen *et al.* 2004; Pope *et al.* 2009; Burnett *et al.* 2014). Several researchers have estimated health risks due to pollution in the Indian context (Madheswaran 2007; Silva 2015; Chowdhury & Dey 2016; Kumar *et al.* 2016; Maji *et al.* 2017a, 2017b, 2018; Balakrishnan *et al.* 2018; Saini & Sharma 2019; Bherwani *et al.* 2020; Manojkumar & Srimurganandam 2021). Integrated exposure-response (IER) model (Kumar *et al.* 2016) estimation of premature deaths due to $PM_{2.5}$ exposure, non-linear power model (Kumar *et al.* 2016) to estimate premature death due to air pollutants, monetary cost-based health assessment studies using methods like the cost of illness, contingent valuation, hedonic wage (Madheswaran 2007; Silva 2015; Maji *et al.* 2017a, 2017b; Bherwani *et al.* 2020) and labour output-based health risk assessment (Pandey *et al.* 2021) is some of the essential works in the Indian context. Most of the studies in the Indian context are based on mortality rate and monetary burden and focus on cities. Also, the models proposed by these researchers were tied to a single pollutant factor. The health risk from exposure to pollution occurs in the rural and urban populations (Garaga *et al.* 2018) in India. However, most of the existing works are exclusively confined to urban centres. Real-time monitoring (Bhowmik *et al.* 2022) can fill this gap and provide a more comprehensive understanding of the nature and distribution of health exposure in Indian villages. This work proposes a comprehensive health risk prediction model for Indian villages that integrates real-time monitoring (Mucchielli *et al.* 2020) of multiple pollutant factors, with a higher coherence to World Health Organization (WHO) benchmarks. Without a comprehensive study covering Indian rural households, understanding the nature and distribution of health exposure in Indian villages is very difficult. Most existing studies are based on $PM_{2.5}$; there are few works on other strong sources in the Indian context, like biomass cooking, trash burning and landfills due to agricultural pesticides and household chemicals. This works addresses this gap and proposes a comprehensive health risk prediction model for Indian villages in terms of multiple pollutant factors.

2. HEALTH RISK ASSESSMENT MODELS

In this section, an overview of different health risk assessment models is provided, including those used in both international and Indian settings.

2.1. Risk assessment arising from pollution

Pope *et al.* (2009) modelled the health risk due to pollution in terms of life expectancy. The changes in life expectancy are analysed in correlation with particulate matter in the air. The regression model is built between the air pollutant levels and life expectancy. The model is adjusted for socio-economic and demographic variables. The model is built at the macro level based on limited observation of air pollutants in selected metropolitan areas of the US.

Burnett *et al.* (2014) proposed an IER model with a relative risk (R.R.) on respiratory problems as output and ambient indoor air pollution caused by solid cooking fuels and smoking. All these pollutant factors are converted to estimate annual $PM_{2.5}$ exposure equivalents and fitted into the IER model. The model considered only air pollutant factors, and it is a macro-level indicator.

Kumar *et al.* (2016) did air quality mapping and health impact assessment for Mumbai city. From air quality observations made at a particular location, spatial variation over a large area is made using ArcGIS interpolation techniques. The health impact assessment was made at ward levels based on the air pollutant level of nitrogen dioxide (NO_2), sulphur dioxide (SO_2) and suspended particulate matter (SPM). The health cost was estimated for each ward. It is difficult to isolate the health cost due to air pollutants alone. In Mumbai, there are other significant factors like sewage, water quality, etc.

2.2. Socio-economic and cultural aspects of risk models

Silva (2015) discussed that the design and architecture must prioritize sustainable practices and take into account the needs of diverse populations in order to create healthy and functional urban environments. Along with several case studies and examples of successful sustainable design practices, including green roofs, urban agriculture and pedestrian-friendly design, the importance of interdisciplinary collaboration and community engagement in creating sustainable cities was significantly pursued.

Maji *et al.* (2017a) proposed an epidemiology-based exposure-response function. The function fitted mortality and morbidity to $PM_{2.5}$ exposure over 24-year data. The fitness function is adjusted for disability-adjusted life years (DALYs). The fitness result is transformed into economic costs. The study was conducted in Mumbai city. The same author in Maji *et al.* (2017b) extended the work for Agra city by incorporating more air pollutant factors. The model could predict health risk in terms of health cost. Nevertheless, extending this study to the village context is impossible as no dependent metrics were available for villages.

2.3. Risk models based on urban setting using particulate matter (PM)

Chowdhury & Dey (2016) developed a non-linear power law (NLP) function to estimate the relative risk in terms of mortality due to ambient $PM_{2.5}$ exposure. Satellite observations of $PM_{2.5}$ were used to predict premature death using the NLP function. Though the model was simple to apply at fine-grained district levels, it could not provide risks to other health factors like physical disabilities resulting from other pollutants.

Maji *et al.* (2018) correlated the $PM_{2.5}$ levels to health risk in terms of mortality using the data collected from 13 major cities. It is a macro-level study demonstrating a significant relationship between mortality and $PM_{2.5}$ levels.

Balakrishnan *et al.* (2018) used $PM_{2.5}$ concentration to estimate death mortality by adjusting for DALYs. The study was conducted at the macro level of states. The study can be used for budget planning but needs to be applied at the fine-grained level of villages for designing effective action plans.

Saini & Sharma (2019) predicted premature death from $PM_{2.5}$ levels using the IER model. Premature death is estimated for each specific problem of stroke, chronic obstructive pulmonary disease, lung cancer and lower respiratory infection.

Manojkumar & Srimurganandam (2021) developed a model correlating the PM concentrations to mortality and hospital admissions. The study was conducted in major Indian cities. Hospital admission count due to respiratory and cardiovascular problems is correlated using linear regression with the PM levels. With the disparity in hospitals across cities and villages, this model can only be used to assess health at the macro level.

Pandey *et al.* (2021) correlated premature deaths after adjusting for DALYs with indoor and outdoor particulate matter pollution. The study was conducted for each Indian state. The estimation was then used to fit the cost of illness method to provide the economic impact of air pollution.

Lu *et al.* (2017) used simultaneous equation modelling (SEM) to analyse the relationship between health and environmental pollution. The study was conducted across China. Air pollutants factors and wastewater emissions are collected over many years and fit the SEM model. The model was able to predict mortality in terms of pollutant factors.

2.4. Evaluation of risk models based on statistical modelling and econometric analysis

Wu *et al.* (2020) estimated healthcare expenditure with increased pollutants. The study was conducted on the pollutant data collected for about 21 years from Taiwan. The data were transformed into time series data, and wavelet analysis was conducted. The model correlated the healthcare expenditure to influencing wavelet coefficients. The model requires a large volume of data.

Hao & Gao (2019) proposed a quantitative relationship between environmental pollution and public health using the expanded Grossman health production function. Pollutant factors in sulphur dioxide and industrial smoke dust emissions are fitted to health risks in terms of mortality rates.

Karambelas *et al.* (2018) designed a correlation model for the health impact due to ambient air pollution. The model was based on an analysis of levels of $PM_{2.5}$ and O_3 and their correlation to the mortality rate over the years. All the air pollutant factors were normalized to $PM_{2.5}$ levels, and linear regression was fit between mortality and $PM_{2.5}$ levels.

Ravishankara *et al.* (2020) estimated premature death mortality in Indian states based on satellite-derived surface $PM_{2.5}$ levels. The study was fine-grained, and death mortality was estimated for six major diseases listed in Global burden of Diseases 2017.

Koul (2021) estimated death mortality after adjustment with DALYs based on three air pollutant factors: ozone, particulate matter and indoor pollution. Like Ravishankara *et al.* (2020), this study was fine-grained, with death mortality estimated for all six significant diseases listed in Global Burden of Diseases 2017.

Ranzani *et al.* (2020) analysed the health risk of indoor household pollution in terms of bone mass. The study was conducted in five semi-urban places in India. Separate linear mixed models were fitted between the $PM_{2.5}$ levels and black carbon levels to the bone mass. The lower bone mass levels are associated with higher $PM_{2.5}$ levels.

Behera *et al.* (2012) estimated the health risks due to groundwater pollutants. Well-known water quality parameters like pH, R.C., turbidity, fluoride, hardness, etc., were collected from the Jagadalpur district. The impact of water quality on perceived health was analysed through a survey study. Nevertheless, the study did not provide any model correlating groundwater pollutants to health risks.

James *et al.* (2020) analysed the impact of cooking fuels on rural women's health. A study was a community-based cross-sectional survey across four villages in Karnataka to estimate health risk in self-reported ophthalmic, cardiovascular and dermatological symptoms with exposure to various cooking fuels. The association between cooking fuels and symptoms were modelled using regression (Rathnamala *et al.* 2021).

The summary of the models is presented in Table 1.

3. METHODOLOGY

The methodology used in this context involves the development of a new model for health risk assessment that considers multiple pollutant factors with perceived health risk assessment, and is coherent with quantitative benchmark-based health risks. The study aimed to overcome the limitations of existing models that are based on limited pollutant factors

Table 1 | Survey summary

Author	Model	Pollutant variables	Health variables
Pope *et al.* (2009)	Linear regression model	$PM_{2.5}$	Life expectancy
Burnett *et al.* (2014)	Integrated exposure-response (IER) model	Indoor and outdoor air pollutants on the scale of $PM_{2.5}$	Premature death mortality
Kumar *et al.* (2016)	Interpolation techniques	SO_2, NO_2 and SPM	Health cost
Silva (2015)	Regression	Ambient air quality index	Premature death mortality
Maji *et al.* (2017a)	Epidemiology-based exposure-response function	$PM_{2.5}$	DALYs
Chowdhury & Dey (2016)	Non-linear power law function	$PM_{2.5}$	Mortality
Maji *et al.* (2018)	Regression	$PM_{2.5}$	Mortality
Balakrishnan *et al.* (2018)	Regression	$PM_{2.5}$	Premature death adjusting for DALYs
Saini & Sharma (2019)	Integrated exposure-response (IER)	$PM_{2.5}$	Stroke, chronic obstructive pulmonary disease (COPD), lower respiratory infection (LRI) and lung cancer (LNC)
Manojkumar & Srimurganandam (2021)	Linear regression	Particular matter (PM)	Hospital admission count for cardiovascular and respiratory problems
Pandey *et al.* (2021)	Cost of illness method	Particulate matter pollution, household air pollution and ozone pollution	Premature death adjusting for DALYs
Bhowmik *et al.* (2022)	Eigen perturbation	Real-time monitoring	Data analytics
Mucchielli *et al.* (2020)	Descriptive and analytical statistics	Online identification of variables	In-situ perception of streaming data
Lu *et al.* (2017)	SEM	SO_2 and wastewater emissions	Mortality
Wu *et al.* (2020)	Wavelet analysis	$PM_{2.5}$	Healthcare expenditure
Hao & Gao (2019)	Expanded Grossman health production function	Sulphur dioxide emissions, industrial smoke dust emissions	Mortality rates
Karambelas *et al.* (2018)	Linear correlation	$PM_{2.5}$, O_3	Mortality rate
Ravishankara *et al.* (2020)	Linear correlation	$PM_{2.5}$	Stoke, COPD, LRI and LNC
Koul (2021)	Linear correlation	Indoor and outdoor pollution in terms of $PM_{2.5}$	Premature death adjusting for DALYs
Ranzani *et al.* (2020)	Separate linear mixed models	The $PM_{2.5}$ levels and black carbon levels	Bone mass
Behera *et al.* (2012)	Correlation model	Groundwater pollutants	Perceived health risk
James *et al.* (2020)	Regression model	$PM_{2.5}$ due to cooking fuel	Ophthalmic, cardiovascular, dermatological symptoms

and estimate health risks in terms of mortality, which is not adequate for accounting for various health abnormalities and loss of livelihood.

To achieve this, the researchers designed a structured questionnaire with 33 questions in four dimensions: water supply, drainage, air pollutant and solid waste, explicitly tailored to the context of Indian villages. The perceived health risk factors were designed by extending the Short Form 36 Health Survey (SF-36) (Treanor & Donnelly 2015), which has been widely adopted by numerous public and private healthcare organizations across various countries (Jenkinson & Layte 1997;

Gandek *et al.* 1998; Kodraliu *et al.* 2001; Hanmer *et al.* 2006; Guerra & Shea 2007; Kontodimopoulos *et al.* 2007). However, the SF-12 was chosen for extension due to its applicability to a broad group of general and vulnerable populations (Côté *et al.* 2004; Jakobsson 2007; Pezzilli *et al.* 2007; Tang *et al.* 2008; Wee *et al.* 2008). The behavioural risk factors of SF-12 were extended with risk factors specific to pollutant contexts and used as perceived health risk factors (Rathnamala *et al.* 2020).

The methodology used in this study involved the development of a new model for health risk assessment that accounted for the unique context of Indian villages and included perceived health risk factors related to multiple pollutant factors. The study used a structured questionnaire and extended the SF-12 to create perceived health risk factors specific to the pollutant contexts in Indian villages. This methodology aimed to provide a more comprehensive and accurate health risk assessment that accounted for various health abnormalities and loss of livelihood, which was not possible with existing models that focused solely on mortality.

Table 2 presents pollutant factors that are grouped into four categories: water supply factors, drainage factors, air pollutant factors and solid waste factors. Each category includes several sub-factors that contribute to perceived health risks in the context of Indian villages. The perceived health risk factors listed in the table include hypertension, cancer, heart disease, gastrointestinal illness, asthma/COPD, psychiatric disease, frequent diarrhoea, skin problems and frequent illness. The table provides a comprehensive list of the factors that the study considered in developing a new model for health risk assessment that is coherent with quantitative benchmark-based health risks and considers multiple pollutant factors.

Table 2 | Pollutant factors

Water supply factors	Source of drinking water (F1) Storage of drinking water (F2) Replacement frequency (F3) Cleaning frequency (F4) Quality of water (F5)
Drainage factors	Canal availability (F6) Type of drainage (F7) Kind of drainage system (F8) Water stagnant (F9) Breeding of insects (F10) Toilet availability (F11) Human waste disposal (F12) Maintenance frequency (F13)
Air pollutant factors	Type of roads (F14) Place of cooking (F15) House ventilation (F16) Kitchen ventilation (F17) Type of cooking fuel (F18) Cigarette use (F19)
Solid waste factors	Garbage piled up nearby (F20) Garbage is strewn on the ground (F21) Disposal facility (F22) Pumping of livestock solid waste (F23) Percentage composition of waste (F24)
Perceived health risk factors	Hypertension (F25) Cancer (F26) Heart disease (F27) Gastrointestinal illness (F28) Asthma/COPD (F29) Psychiatric disease (F30) frequent diarrhoea (F31) Skin problems (F32) Frequent illness (F33)

Each perceived health risk factor collects responses on two scales (Yes/No). Based on the respondents' perceived health risk factors, the health risk is categorized into three levels: Level 1 (L1), Level 2 (L2) and Level 3 (L3). The mapping is given in Table 3.

Deviating from earlier works on modelling the relationship between the pollutant factors and health risk as a linear model, this work proposes a fuzzy model to estimate the health risk in terms of scores for each level (L1, L2 and L3).

Survey is conducted across 2,370 respondents from 104 villages spread across three districts of Kolar, Chikkballapura and Bengaluru Rural in Karnataka (Rathnamala & Shivashankara 2022; Figure 1).

The significant factors among the 24 factors (F1–F24) in the dimension of water supply, drainage, air pollutant and solid waste are identified by the symmetric entropy (S.E.) test between each of the factors and levels of perceived health risks. The symmetric entropy (S.E.) between the input variable a (F1–F24) and the output variable b (level of perceived health risk) is calculated as

$$\text{S.E.}(a, b) = \frac{2 \times \text{MI}(a, b)}{H(a) + H(b)} \tag{1}$$

where $\text{MI}(a, b)$ is the mutual information between the variables a and b. $H(a)$ is the entropy for the variable a.

Table 3 | Factor mapping to risks

Risks	F25	F26	F27	F28	F29	F30	F31	F32	F33
L1				√			√	√	√
L2	√					√			
L3		√	√		√				

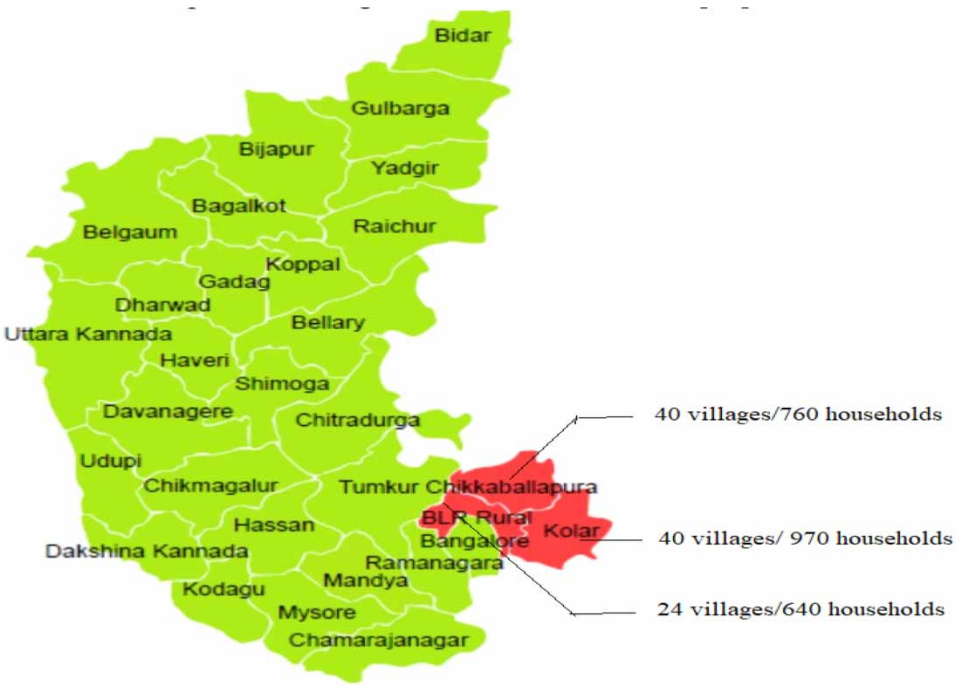


Figure 1 | Survey population in Karnataka.

Mutual information between variable a and b is calculated as

$$\mathrm{MI}(a, b) = \sum_a \sum_b \mathrm{PDF}(a, b) \log \frac{\mathrm{PDF}(a, b)}{p(a) * p(b)} \tag{2}$$

PDF(a) is the probability density function for the variable a, and PDF(a, b) is the joint probability density function.

$H(a)$ is calculated in terms of Shanon's entropy as

$$H(a) = -\int \mathrm{PDF}(a) \log (\mathrm{PDF}(a)) dx \tag{3}$$

The factors whose S.E. value is greater than 0.7 are decided as significant factors. The significant factors found from analysis of data responses of 2,370 participants are given in Figure 2.

In Figure 2, the 14 factors whose values exceeded the 0.7 threshold are highlighted in red. Using responses from 2,370 participants, a dataset was created that includes the significant factors and corresponding perceived health risk levels. The relationship between these factors and perceived health risk levels was analysed using Fuzzy C Means clustering. Specifically, the dataset (D) was subjected to Fuzzy C Means clustering with three clusters, resulting in the definition of cluster centres represented by

$$D = \{D_{e,q}, \quad e = 1, 2 \ldots k \text{ and } q = 1, 2, 3\}$$

where $D_{e,q}$ is the qth factor of the eth cluster.

The closeness of the qth factor of the rth data $f_{r,q}$, with the qth coordinate of the eth cluster is defined using the Gaussian function as

$$G(f_{r,q}, D_{e,q}, \sigma_{e,q}) = e^{\frac{(f_{r,q} - D_{e,q})^2}{\sigma_{e,p}^2}} \tag{4}$$

where

$$\sigma_{e,q} = \frac{1}{N_e} \sum_{r=1}^{N_e} (f_{r,q} - D_{e,q})^2 \tag{5}$$

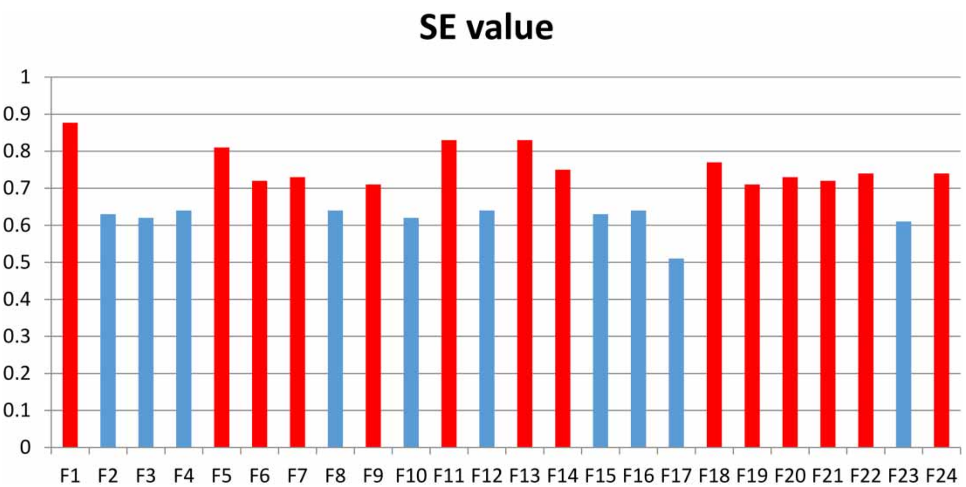


Figure 2 | Symmetric entropy value. Please refer to the online version of this paper to see this figure in colour: https://dx.doi.org/10.2166/wst.2023.084.

The closeness of the factor of rth data to the eth cluster is given as

$$\psi_{r,e} = \prod_{q=1}^{P} G(f_{r,q}, D_{e,q}, \sigma_{e,q}) \tag{6}$$

The output label for the eth cluster is found from the linear regression of input factor $f_{r,q}$ as

$$\Phi_{r,e} = W_{e,0} + \sum_{q=1}^{P} W_{e,q,f_{r,q}} \tag{7}$$

where W is the regression coefficient of the eth cluster. Since each of the rth data has membership value to all P ($P = 3$) clusters, the final label of that particular link is given by weighting the label of the link with its membership value as

$$\bar{N}(r) = \sum_{e=1}^{P} \psi_{r,e}\Phi_{r,e} \tag{8}$$

The value of $\bar{N}(r)$ calculated above may have an error with respect to $N(r)$ from training. The total error is calculated as

$$E = \sum_{r=1}^{N} ||\bar{N}(r) - N(r)||^2 \tag{9}$$

The Gaussian parameters $D_{e,q}$, $\sigma_{e,q}$ and the regression coefficients $W_{e,p}$ are tuned to reduce the error defined above using the gradient descent method.

$$D_{e,q}(t+1) = D_{e,q}(t) + \quad \eta_C \frac{\partial E}{\partial D_{e,q}} \tag{10}$$

$$\sigma_{e,q}(t+1) = \sigma_{e,q}(t) + \quad \eta_\sigma \frac{\partial E}{\partial \sigma_{e,q}} \tag{11}$$

$$W_{e,q}(t+1) = W_{e,q}(t) + \quad \eta_W \frac{\partial E}{\partial W_{e,q}} \tag{12}$$

where t is the iteration number and η_C, η_σ, η_W are the learning parameters. The iteration is stopped when the error threshold is reached. At the end of the iteration fuzzy membership function $\Phi_{r,e}$ mapping the values of significant factors to the scores for each level of health risk is obtained. The score is in the range of 0–1, with the sum of scores of all three levels as 1. The overall process of the proposed health risk estimation model is given in Figure 3.

From the data collected (F1–F33) from participants, significant factors are found using S.E. analysis (Equation (1)). The significant factors are fit for L1, L2 and L3 risk prediction into Equation (7). For any response from rural household (F1–F24), the data are fit into Equation (7) for L1, L2 and L3 to get three health risk scores as output.

4. RESULTS

The health risk level scores for the 2,370 households are distributed for Level 1, Level 2 and Level 3, as shown in Figure 4.

From the results, there are only 20% of households above 0.6 score for L1 and 40% of households below 0.4 score, 40% of households are in the score of 0.4–0.6. For L2, there are only 20% of households below 0.4 score. 50% of households are in the score of 0.4–0.6, 30% of households are above score of 0.6. For L3, there are only 30% of households below 04 score. 30% of households are in the score of 0.4–0.6. 40% of households are above 0.6. 40% of households above L3 score of 0.6 is alarming and indicates an onset of severe health risk in these households.

The distribution of the maximum of L1, L2 and L3 in each of the districts is given in Figure 5.

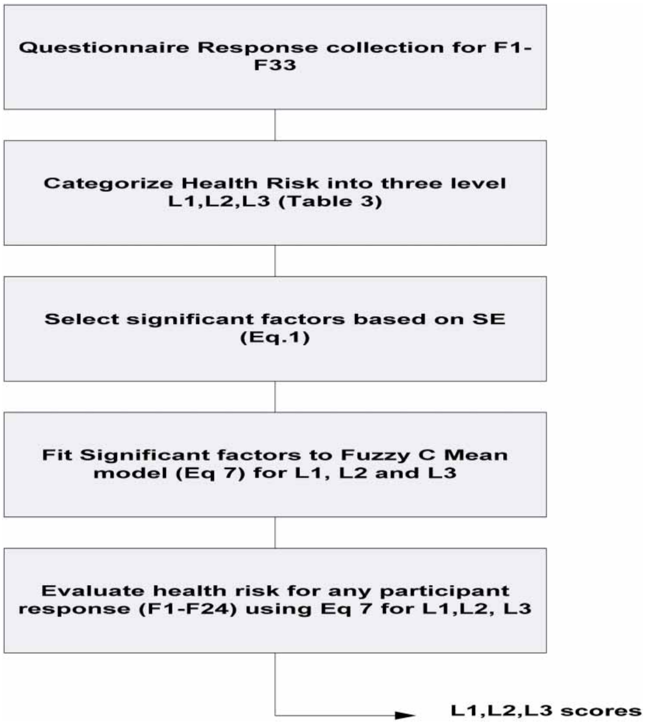


Figure 3 | Process flow.

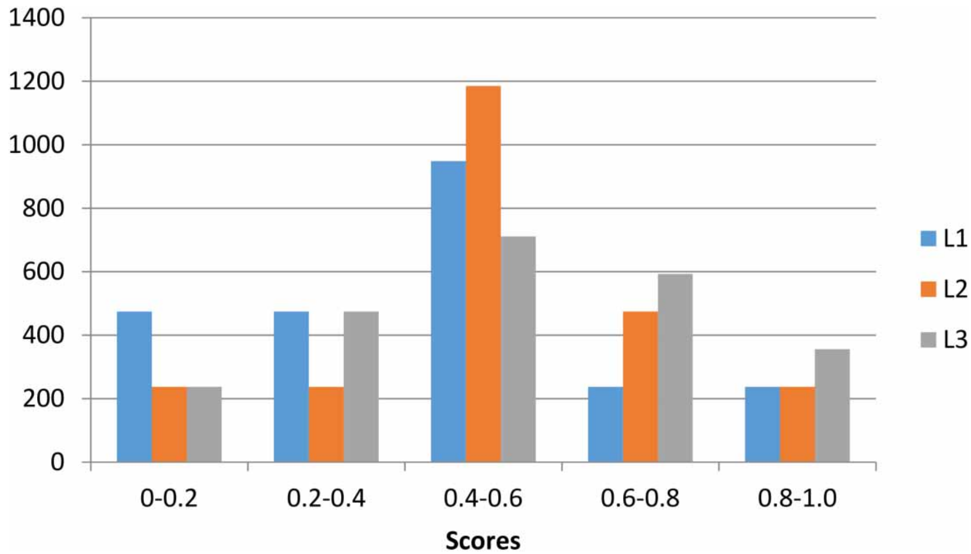


Figure 4 | Distribution of health risk scores.

The obtained results suggest that the proposed health risk prediction model is consistent with existing works that use pollutant factors to estimate health risks associated with air pollution. The study compares the perceived health score provided by the proposed model with the most used pollutant factors in existing works, such as $PM_{2.5}$.

The L3 score is highest in Bangalore rural at 35% compared with Chikballapur at 22% and Kolar at 20%. Thus, Bangalore rural is in alarming condition compared with Kolar and Chikballapur. The correlation between the most used pollutant factors in existing works and the perceived health score provided by the proposed model is compared for consistency. The results for the correlation between $PM_{2.5}$ and the proposed perceived health score are given in Figure 6. The study found that as the

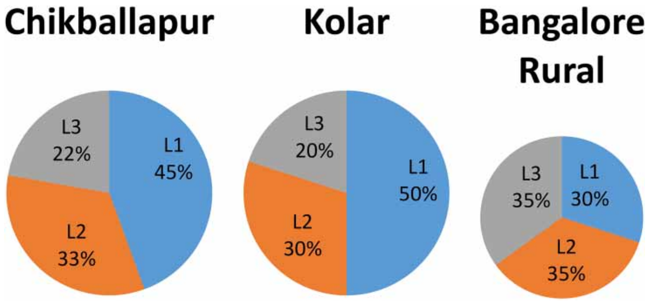


Figure 5 | Distribution of risk scores across survey districts.

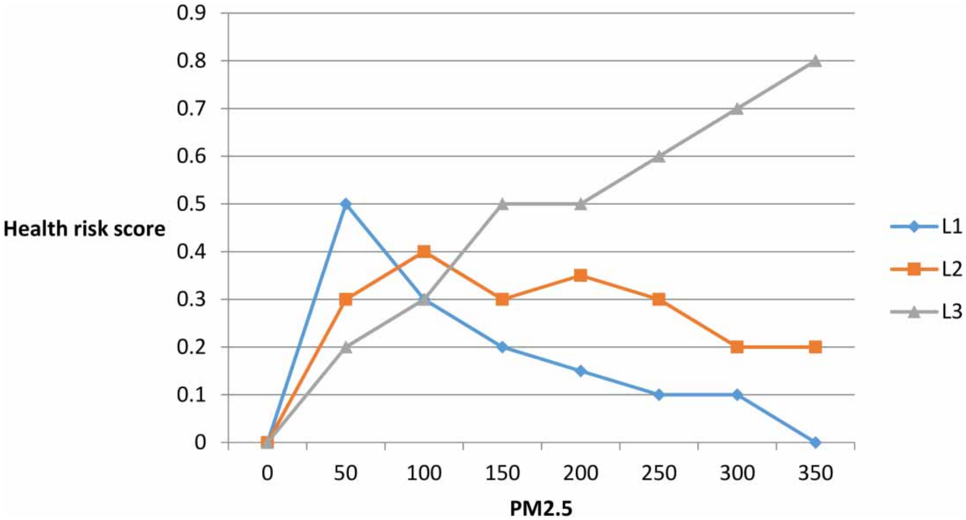


Figure 6 | Convergence to the proposed health score to $PM_{2.5}$.

concentration of $PM_{2.5}$ increases, the perceived health score provided by the proposed model (L1 and L2 score) drops, indicating a decrease in health. Conversely, the L3 score, which represents the most severe health consequences, increases with increasing $PM_{2.5}$ concentrations. This finding is consistent with previous studies that have shown that exposure to high levels of $PM_{2.5}$ can lead to severe health consequences such as respiratory and cardiovascular diseases. Furthermore, the study found a strong correlation between $PM_{2.5}$ and the L3 score, with an R^2 value greater than 0.9 (Figure 7). This indicates that the proposed model is effective in estimating the health risks associated with $PM_{2.5}$ exposure in the study area. Overall, the obtained results suggest that the proposed health risk prediction model is consistent with existing works and is effective in estimating health risks associated with air pollution in the study area. The strong correlation between $PM_{2.5}$ and the L3 score also indicates that the proposed model can be used to inform policy decisions aimed at reducing air pollution and improving public health.

The proposed model has higher R^2 for any one of scales of L1, L2 and L3 for most of the pollutant factors as seen in Table 4.

The R^2 value for most of the pollutant factors is more significant than 0.8. This signifies a higher consistency of the proposed perceived health score with most pollutant factors. The significance is achieved against one of the L1, L2 or L3 scores, justifying the reason for modelling the health risk as a fuzzy decision on pollutant factors.

Three important salient features of the proposed health risk prediction model are its simplicity, effectiveness and adaptability. Compared with the IER model (Saini & Sharma 2019), NLP function (Chowdhury & Dey 2016) and epidemiology-based exposure-response function (Maji *et al.* 2017a), the proposed health risk prediction model does not need pollutant measurements over a long period. Over time, pollutant observations are unavailable for rural Indian areas. It is costly to

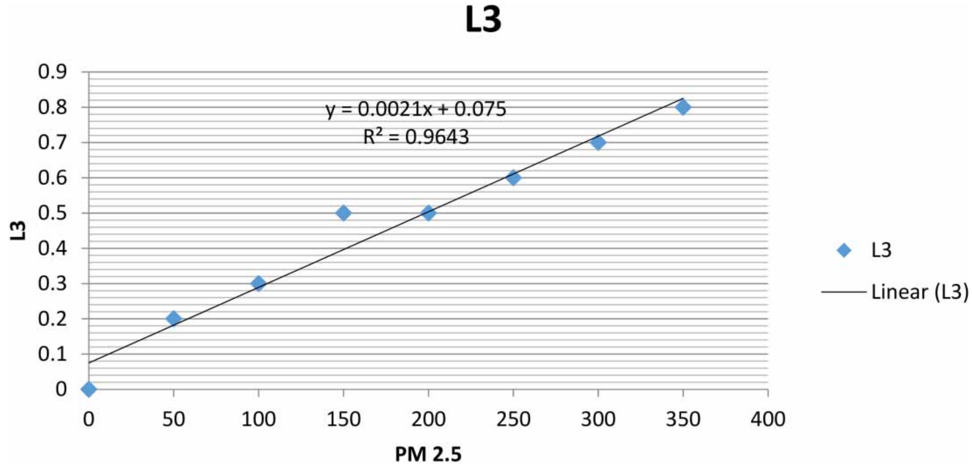


Figure 7 | Correlation between $PM_{2.5}$ and L3.

collect those parameters considering the large village distribution in India. The proposed health risk prediction model evaluates health risk at a perceived level based on a 24-item questionnaire response. The questionnaire responses are straightforward to collect. Considering the recent in-depth penetration of Smartphone revolution in India, this survey question can be easily launched as a mobile application. Feedback can be collected, and health risk scores can be provided instantly. A perceived health evaluation approach (Behera *et al.* 2012; James *et al.* 2020) lacks this effectiveness as they need pollutant measurements. Also, the approaches (Behera *et al.* 2012; Chowdhury & Dey 2016; Maji *et al.* 2017a; Saini & Sharma 2019; James *et al.* 2020) lacks adaptability. They are inflexible to adding new pollutant factors and perceived health risk factors. However, the proposed health risk estimation model scores best in adaptability as the model can be extended for new pollutants and health risk factors.

Air pollution is a significant public health concern in India, where pollutant observations are often unavailable in rural areas due to high costs and limited resources. To address this challenge, the proposed health risk prediction model offers a simple, effective and adaptable approach to evaluating health risks at a perceived level based on a 24-item questionnaire response.

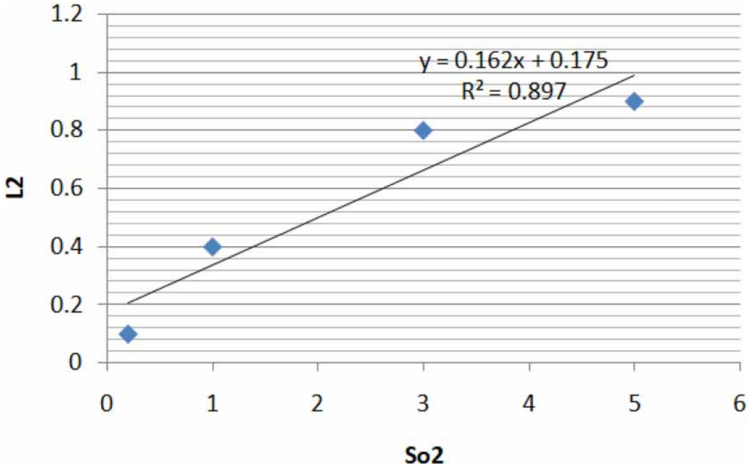
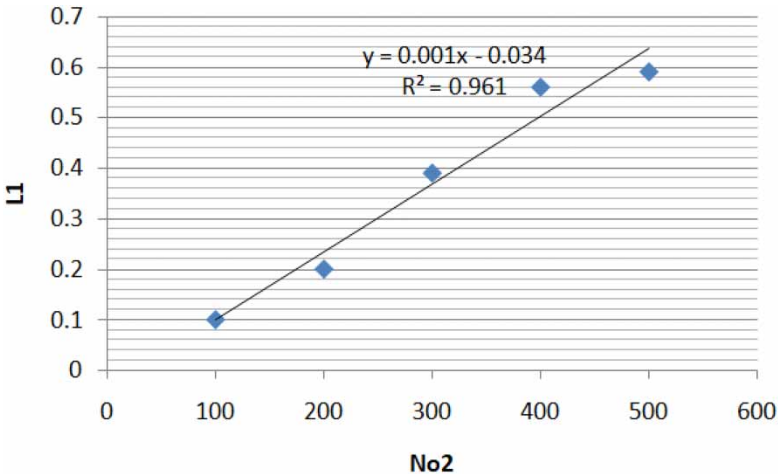
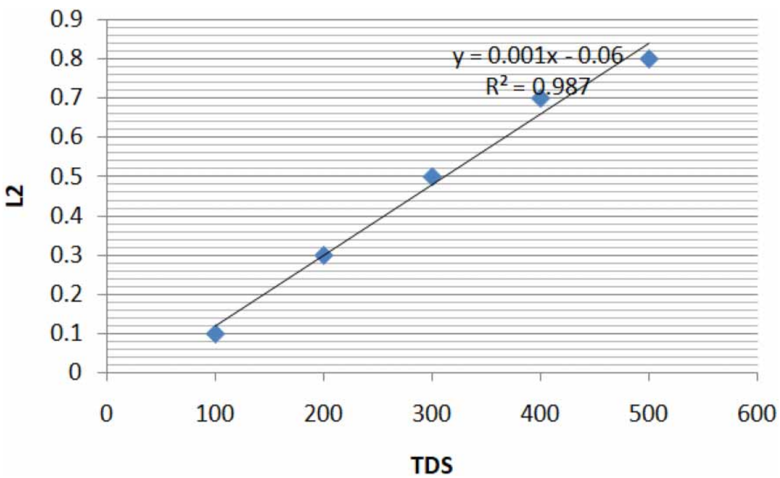
The IER model is a commonly used method to estimate the health impacts of air pollution exposure. IER uses a linear model to relate exposure to health outcomes, but it requires data on pollutant concentrations over a long period of time to estimate the exposure-response function. This approach may not be feasible in rural areas where pollutant measurements are limited, which highlights the importance of the proposed health risk prediction model's simplicity and adaptability. However, IER has been shown to be effective in estimating health risks in urban areas where pollutant measurements are available (Fann *et al.* 2012).

Another method that has been used to evaluate health risks associated with air pollution is the NLP function. This method assumes a non-linear relationship between exposure and health outcomes and is particularly useful for short-term exposure assessments. However, it also requires pollutant concentration data and is therefore limited in its application in areas where such data is unavailable. A study conducted in China showed that the NLP function had a better performance in predicting daily hospital admissions for respiratory diseases than other models (Liu *et al.* 2019).

Epidemiology-based exposure-response functions have also been used to estimate health risks associated with air pollution. These functions are based on observed associations between air pollution and health outcomes in epidemiological studies. However, they also require pollutant concentration data and may not be feasible in areas where such data is limited. A study conducted in Canada used an epidemiological approach to estimate the burden of air pollution on premature mortality (Brook *et al.* 2010).

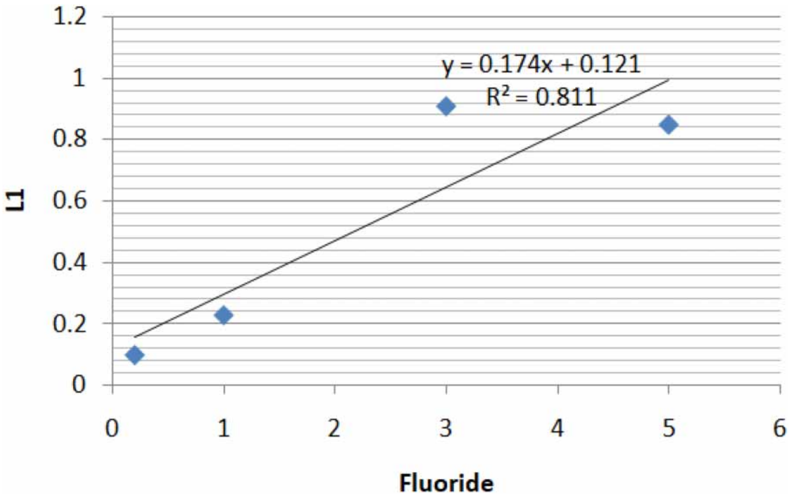
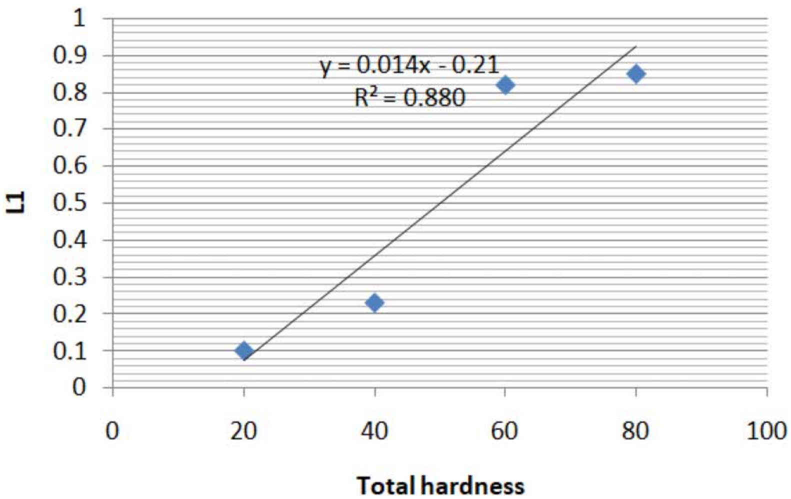
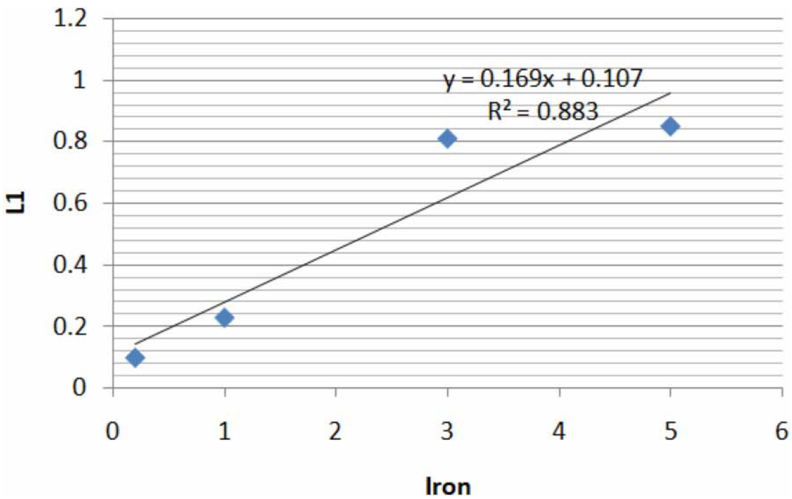
Perceived health evaluation approaches have been used to assess health risks associated with air pollution in areas where pollutant measurements are limited. These approaches use self-reported health outcomes to estimate the health impacts of air pollution exposure. However, perceived health evaluation approaches lack effectiveness as they rely on subjective measures of health outcomes and may not accurately reflect actual health impacts. A study conducted in China compared perceived health status with actual health outcomes and found that the two were not always consistent (Ye *et al.* 2013).

Table 4 | Correlation of THE proposed health risk score to pollutant factors

Pollutant factor	R^2	Fitness
SO_2	L2 = 0.89	
NO_2	L1 = 0.84	
Total dissolved salt (TDS)	L2 = 0.98	

(*Continued.*)

Table 4 | Continued

Pollutant factor	R^2	Fitness
Fluoride (F)	L2 = 0.81	
Total hardness (T.H.)	L1 = 0.86	
Iron	L1 = 0.88	

The proposed health risk prediction model offers an adaptable approach to evaluating health risks associated with air pollution. The model can be extended to include new pollutant factors and perceived health risk factors, which is a significant advantage over the other methods discussed. Additionally, the use of a questionnaire-based approach to collect data on perceived health outcomes makes the proposed model easily deployable as a mobile application. This feature can significantly reduce the cost and time associated with data collection, especially in areas with limited resources.

5. CONCLUSION

The proposed health risk prediction model is a novel approach to estimate health risks associated with air pollution, water pollution and landfill factors in rural households in India. The model uses a 24-item questionnaire to provide three-scale qualitative scores for rural households. Compared with expensive chemical tests based on inferences, the proposed model is simple and cost-effective, making it suitable for rural Indian villages where pollutant measurements may not be readily available. Additionally, the model can be realized using semi-skilled staff, further reducing the cost and technical expertise required.

One of the main benefits of the proposed model is its simplicity and cost-effectiveness. As mentioned, existing approaches based on chemical tests and pollutant measurements can be expensive and may not be feasible in rural Indian villages due to logistical and financial constraints. The proposed model overcomes these limitations by using a questionnaire-based approach that is easy to administer and cost-effective.

Furthermore, the proposed model was found to have higher consistency compared to benchmark air pollutant, water pollutant and landfill factor methods. This indicates that the proposed model can provide accurate and reliable estimates of health risks associated with pollution in rural households in India.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Balakrishnan, K., Dey, S., Gupta, T., Dhaliwal, R. S., Brauer, M., Cohen, A. J., Stanaway, J. D., Beig, G., Joshi, T. K., Aggarwal, A. N. & Sabde, Y. 2019 The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of Disease Study 2017. *Lancet* **3** (1), 26–39.

Behera, B., Das, M. & Rana, G. S. 2012 Studies on groundwater pollution due to iron content and water quality in and around Jagdalpur, Bastar district, Chattisgarh, India. *Journal of Chemical and Pharmaceutical Research* **4** (9), 3803–3807.

Bherwani, H., Nair, M., Kapley, A. & Kumar, R. 2020 Valuation of ecosystem services and environmental damages: an imperative tool for decision making and sustainability. *European Journal of Sustainable Development Research* **4** (4), 33.

Bhowmik, B., Hazra, B. & Pakrashi, V. 2022 *Real-time Structural Health Monitoring of Vibrating Systems*. CRC Press, Boca Raton, FL, USA.

Brook, R. D., Rajagopalan, S., Pope III, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y., Luepker, R. V., Mittleman, M. A. & Peters, A. 2010 Particulate matter air pollution and cardiovascular disease. *Circulation* **121** (21), 2331–2378.

Burnett, R. T., Pope, C. A., Ezzati, M., Olives, C., Lim, S. S., Mehta, S., Shin, H. H., Singh, G., Hubbell, B., Brauer, M. & Anderson, H. R. 2014 An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environmental Health Perspectives* **122** (4), 397–403.

Chowdhury, S. & Dey, S. 2016 Cause-specific premature death from ambient $PM_{2.5}$ exposure in India: estimate adjusted for baseline mortality. *Environment International* **91**, 283–290.

Cohen, A. J. 2004 *Urban air pollution, in comparative quantification of health risks: Global and regional burden of disease attributable to selected major risk factors*. M. Ezzati, A. D. Lopez & C. J. L. Murray (eds). World Health Organization, Geneva, Switzerland, pp. 1353, 1433.

Côté, I., Grégoire, J. P., Moisan, J. & Chabot, I. 2004 Quality of life in hypertension: the SF-12 compared to the SF-36. *Journal of Population Therapeutics and Clinical Pharmacology* **11** (2), 232–238.

Fann, N., Nolte, C. G., Dolwick, P., Spero, T. L., Brown, A. C., Phillips, S. & Anenberg, S. 2012 The geographic distribution and economic value of climate change-related ozone health impacts in the United States in 2030. *Journal of the Air & Waste Management Association* **62** (6), 644–655.

Feng, S., Gao, D., Liao, F., Zhou, F. & Wang, X. 2013 The health effects of ambient PM2. 5 and potential mechanisms. *Ecotoxicology and Environmental Safety* **128**, 67–74.

Gandek, B., Ware, J. E., Aaronson, N. K., Apolone, G., Bjorner, J. B., Brazier, J. E., Bullinger, M., Kaasa, S., Leplege, A., Prieto, L. & Sullivan, M. 1998 Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA project. *Journal of Clinical Epidemiology* **51** (11), 1171–1178.

Gao, N., Li, C., Ji, J., Yang, Y., Wang, S., Tian, X. & Xu, K. F. 2019 Short-term effects of ambient air pollution on chronic obstructive pulmonary disease admissions in Beijing, China (2013–2017). *International Journal of Chronic Obstructive Pulmonary Disease*, 297–309.

Garaga, R., Sahu, S. K. & Kota, S. H. 2018 A review of air quality modeling studies in India: local and regional scale. *Current Pollution Reports* **4** (2), 59–73.

Guerra, C. E. & Shea, J. A. 2007 Health literacy and perceived health status in Latinos and African Americans. *Ethnicity & Disease* **17** (2), 305–312.

Hanmer, J., Lawrence, W. F., Anderson, J. P., Kaplan, R. M. & Fryback, D. G. 2006 Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores. *Medical Decision Making* **26** (4), 391–400.

Hao, L. & Gao, J. 2019 Influence of environmental pollution on residents' health: evidence from Hubei, China. *Natural Environment and Pollution Technology* **18** (1), 211–216.

Jakobsson, U. 2007 Using the 12-item Short Form health survey (SF-12) to measure quality of life among older people. *Aging Clinical and Experimental Research* **19**, 457–464.

James, B. S., Shetty, R. S., Kamath, A. & Shetty, A. 2020 Household cooking fuel use and its health effects among rural women in southern India – a cross-sectional study. *PLoS ONE* **15** (4), e0231757.

Jenkinson, C. & Layte, R. 1997 Development and testing of the UK SF-12. *Journal of Health Services Research & Policy* **2** (1), 14–18.

Karambelas, A., Holloway, T., Kinney, P., Fiore, A., Defries, R., Kiesewetter, G. & Heyes, C. 2018 Urban versus rural health impacts attributable to $PM_{2.5}$ and O_3 in northern India. *Environmental Research Letters* **13**, 064012.

Kodraliu, G., Mosconi, P., Groth, N., Carmosino, G., Perilli, A., Gianicolo, E. A., Rossi, C. & Apolone, G. 2001 Subjective health status assessment: evaluation of the Italian version of the SF-12 Health Survey. Results from the MiOS project. *Journal of Epidemiology and Biostatistics* **6** (3), 305–316.

Kontodimopoulos, N., Pappa, E., Niakas, D. & Tountas, Y. 2007 Validity of SF-12 summary scores in a Greek general population. *Health and Quality of Life Outcomes* **5** (1), 1–9.

Koul, P. & Collaborators, India 2021 Health and economic impact of air pollution in the states of India: the Global Burden of Disease Study 2019. *The Lancet Planetary Health* **5**, e25–e38.

Kumar, A., Gupta, I., Brandt, J., Kumar, R., Dikshit, A. K. & Patil, R. S. 2016 Air quality mapping using GIS and economic evaluation of health impact for Mumbai city, India. *Journal of the Air & Waste Management Association* **66** (5), 470–481.

Lu, Z.-N., Bai, Y., Wang, J.-Z., Li, Y. & Liu, Q.-Q. 2017 The dynamic relationship between environmental pollution, economic development and public health: evidence from China. *Journal of Cleaner Production* **166**, 134–147.

Madheswaran, S. 2007 Measuring the value of statistical life: estimating compensating wage differentials among workers in India. *Social Indicators Research* **84**, 83–96.

Maji, K. J., Dikshit, A. K. & Deshpande, A. 2017a Disability-adjusted life years and economic cost assessment of the health effects related to $PM_{2.5}$ and PM_{10} pollution in Mumbai and Delhi, in India from 1991 to 2015. *Environmental Science and Pollution Research* **24**, 4709–4730.

Maji, K. J., Dikshit, A. K. & Deshpande, A. 2017b Assessment of city level human health impact and corresponding monetary cost burden due to air pollution in India taking Agra as a model city. *Aerosol and Air Quality Research* **17**, 831–842.

Maji, K. J., Arora, M. & Dikshit, A. K. 2018 Premature mortality attributable to $PM_{2.5}$ exposure and future policy roadmap for 'airpocalypse' affected Asian megacities. *Process Safety and Environment Protection* **118**, 371–383.

Manojkumar, N. & Srimurganandam, B. 2021 Health effects of particulate matter in major Indian cities. *International Journal of Environmental Health Research* **31** (3), 258–270.

Mucchielli, P., Bhowmik, B., Hazra, B. & Pakrashi, V. 2020 Higher-order stabilized perturbation for recursive eigen-decomposition estimation. *Journal of Vibration and Acoustics* **142**, 6.

Pandey, A., Brauer, M., Cropper, M. L., Balakrishnan, K., Mathur, P., Dey, S., Turkgulu, B., Kumar, G.A., Khare, M., Beig, G. & Gupta, T. 2021 Health and economic impact of air pollution in the states of India: the Global Burden of Disease Study 2019. *The Lancet Planetary Health* **5** (1), e25–e38.

Pezzilli, R., Morselli-Labate, A. M., Fantini, L., Campana, D. & Corinaldesi, R. 2007 Assessment of the quality of life in chronic pancreatitis using Sf-12 and EORTC Qlq-C30 questionnaires. *Digestive and Liver Disease* **39** (12), 1077–1086.

Pope, C. A., Ezzati, M. & Dockery, D. W. 2009 Fine-particulate air pollution and life expectancy in the United States. *New England Journal of Medicine* **360** (4), 376–386.

Ranzani, O., Milà, C., Kulkarni, B., Kinra, S. & Tonne, C. 2020 Association of ambient and household air pollution with bone mineral content among adults in peri-urban South India. *JAMA Network Open* **3** (6), e1918504.

Rathnamala, G. V., Ashwini, R. M., Sai, P., Chowdary, K. Y. & Harshitha, C. 2020 Analysis of drinking water quality and mode of storage in Dodballapur Taluk. *Solid State Technology* **63** (5), 945–957.

Rathnamala, G., Ashwini, R. & Babitha, N. 2021 Domestic environmental destructions due to lack of solid waste management in rural areas. *Advances in Mathematics Scientific Journal* **10** (3), 1807–1819.

Rathnamala, G. V. & Shivashankara, G. P. 2022 An evidence-based approach towards identifying household emerging pollutants in rural areas in Southern Karnataka. *Journal of Water, Sanitation and Hygiene for Development* **12** (6), 498–516.

Ravishankara, A. R., David, L., Pierce, J. & Venkataraman, C. 2020 Outdoor air pollution in India is not only an urban problem. *Proceedings of the National Academy of Sciences* **117** (6), 2751–2752.

Saini, P. & Sharma, M. 2019 Cause and age-specific premature mortality attributable to $PM_{2.5}$ exposure: an analysis for million-plus Indian cities. *Science of the Total Environment* **7**, 10.

Silva, L. T. 2015 Environmental quality health index for cities. *Habitat International* **45**, 29–35.

Tang, L. Y., Nabalamba, A., Graff, L. A. & Bernstein, C. N. 2008 A comparison of self-perceived health status in inflammatory bowel disease and irritable bowel syndrome patients from a Canadian national population survey. *Canadian Journal of Gastroenterology and Hepatology* **22**, 475–483.

Treanor, C. & Donnelly, M. 2015 A methodological review of the Short Form Health Survey 36 (SF-36) and its derivatives among breast cancer survivors. *Quality of Life Research* **24**, 339–362.

Wee, C. C., Davis, R. B. & Hamel, M. B. 2008 Comparing the SF-12 and SF-36 health status questionnaires in patients with and without obesity. *Health and Quality of Life Outcomes* **6**, 1–7.

Wu, C. F., Li, F., Hsueh, H. P., Wang, C. M., Lin, M. C. & Chang, T. 2020 A dynamic relationship between environmental degradation, healthcare expenditure and economic growth in wavelet analysis: empirical evidence from Taiwan. *International Journal of Environmental Research and Public Health* **17** (4), 1386.

First received 11 November 2022; accepted in revised form 10 March 2023. Available online 21 March 2023