Synthesis of water, sanitation, and hygiene (WaSH) spatial pattern in rural India: an integrated interpretation of WaSH practices

Ankita Pran Dadhich · Pran N. Dadhich · Rohit Goyal

Received: 5 April 2022 / Accepted: 4 July 2022 / Published online: 8 July 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Rural areas largely lack access to improved drinking water, sanitation, and hygiene (WaSH) facilities in India. This requires documentation of WaSH practices at the local level for better understanding and sustainable development. In this paper, a global positioning system (GPS)-based household survey was carried out in 67 villages of Phagi tehsil using individual questionnaires to evaluate the existing WaSH conditions spatially at the panchayat level. Three sub-indices were used for WaSH risk areas mapping and prediction with the integration of machine learning algorithms. Survey results indicate the improvement in the availability of toilet facilities; however, a gap was found between toilet ownership and its usage by villagers. Data show that only six panchayats have almost zero open defecation practices among the 32 panchayats of Phagi tehsil. The findings highlight that presence of toilets in house, water supply in toilets, and high literacy rate lead to an increase in toilet usage by the population. WaSH index scores indicate that panchayats like Mandawari, Mendwas, Chandma Kalan, and Rotwara have worst conditions and fall in the high-risk category. Moreover, support vector machine regression (SVMR) results reveal that WaSH scores are mainly affected by open defecation ($r = 0.94$), water supply in toilets ($r = 0.92$), and female members’ participation in sanitation facilities decision-making ($r = 0.53$), followed by literacy rate ($r = 0.33$). Findings demonstrate the association between gender inequalities and WaSH conditions, and the potential of the WaSH index as a monitoring tool by local policymakers to shrink the WaSH gaps.

Keywords Water-sanitation-hygiene · Spatial assessment · Defecation practices · WaSH index · SVMR models · GIS

Introduction
Accessibility to safe drinking water, adequate sanitation, and appropriate hygiene services are the necessities of people and essential to protecting them from infectious disease outbreaks. Latest estimates reveal that about 1.7 million deaths in the world are due to diarrhea which reflects the lack of safely managed drinking water, sanitation, and hygiene (WaSH) facilities (Gupta and Obani 2016). In India, 0.8 million deaths were reported due to diarrheal and intestinal infections mainly due to unsafe WaSH practices (GBD 2017). According to UNICEF and WHO (2019), about 2.2 billion people around the world do not have access to safe drinking water, 4.2 billion people lack proper sanitation facilities, 673 million people practice open defecation, and 3 billion do not have basic hand-washing facilities in the socio-economically challenged countries. Improved WaSH facilities are vital to prevent the transmission of diseases such as diarrhea, cholera, dysentery, hepatitis A, typhoid, and COVID-19, and this will help to create resilient communities. Hence, the international community designed sustainable development goals (SDG 6.1 and SDG 6.2) to provide safe and affordable WaSH facilities for the entire population and to end open defecation.
The practice of open defecation results in environmental (water) contamination that leads to waterborne diseases which are also responsible for high child mortality (Chan et al. 2021). India has achieved significant success in this area because of the Government of India’s flagship programs such as the Swachh Bharat Mission (SBM), Total Sanitation Campaign, and the National Rural Drinking Water Programme (NRDWP) at district levels in each state. SBM has led to wide-scale construction of toilets to end open defecation across the country by providing financial support to below-poverty-line households, landless laborers, small and marginal farmers, women-headed households, and differently able people. However, 34 percent of Indian states face high water contamination levels, and about 718 districts have no access to safe water (WHO 2017). The situation is worse in rural areas of Rajasthan state, which is characterized by limited surface water bodies, scanty rainfall, and frequent droughts. Here, women/girls have to travel a minimum of 3–5 km daily to fetch water for their household needs. Furthermore, women face more health risks than men from unsanitary conditions and sometimes become the victims of violence while defecating in the open (World Bank 2011; Lee 2017). Gender inequality is also another critical issue in rural households of Rajasthan, as women are not decision-makers regarding sanitation and water facilities (Doron and Jeffrey 2014). These attitudes and practices in rural Rajasthan require a detailed household survey to monitor and evaluate the issues related to WaSH such as open defecation, toilet usage, drinking water quality, gender inequality, and the progress toward the Swachh Bharat Mission.

In areas with scarcity of surface water resources, inadequate sanitation, and hygiene services, geographical information system (GIS) can be an efficient and powerful tool to visualize the existing situation of WaSH and identify the risk areas (Maina 2015; De Moura and Procopiuck 2020). The detailed spatial patterns of drinking water, sanitation, and hygiene conditions at the local level could help to provide sustainable measures at the least cost. Several indexes were defined to evaluate the WaSH conditions (Cronk et al. 2015; Hashemi 2020; Dickin et al. 2021); however, local policymakers require a simple and easy-to-use index for quick assessment. Hence, in this research, a simple index was developed to monitor and understand the WaSH practices. Nowadays, with the development of computational intelligence techniques, several linear and nonlinear methods such as discriminant analysis, partial least squares, artificial neural networks, and support vector machines are available for categorization and regression (Noori et al. 2013; Leong et al. 2019). Kernel-based techniques like support vector machine regression (SVMR) offer more advantages over their ANN equivalents as they can model nonlinear systems well to produce more accurate results, can resolve small samples, and allow interpretation of the calibration models (Yoon et al. 2017; Haghiabi et al. 2018). Machine learning algorithms have the potential to detect patterns in the collected data and predict unknown variables (Froemelt et al. 2019; Shah et al. 2020). Hence, the issues impacting safe water availability, usage of toilets, defecation practices, and hygiene behavior prevailing in rural areas could be better evaluated using machine learning techniques. The objectives of the present study are as follows.

- Documentation and evaluation of the behavior of existing water, sanitation, and defecation practices in rural areas of Rajasthan using a global positioning system (GPS)-based survey.
- Classification and development of the WaSH index using water, sanitation, and hygiene indicators for the assessment of the current WaSH practices of the rural population.
- Spatial risk mapping to ascertain the WaSH practices of rural people at the panchayat level.
- WaSH index prediction by integrating machine learning algorithms to understand the attitude of the villagers.

**Material and methods**

The study area for this work is the Phagi tehsil (Fig. 1) of Jaipur district in the state of Rajasthan in western India. It is located about 51 km away from Jaipur city. Geographically, Phagi tehsil covers a 1114.3 km² area with an average elevation of 383 m. Phagi tehsil is characterized by a semi-arid climate and two ephemeral rivers, namely, Bandi and Masi, flow through the study area. The average annual rainfall is 564 mm. The total population of Phagi tehsil is 191,126 (Census 2011) with 99,226 males and 91,900 females. The average sex ratio (female to male) is 0.926, and the average total literacy rate is 61.7% with 76.22% male and 46.17% female literacy rate. A global positioning system (GPS) based household survey was carried out in different villages of Phagi tehsil to gain insight into the water, sanitation, and hygiene behavior prevailing in rural areas at the panchayat level.

The high population density was observed in Chandma Kalan (383.7 per sq km), Renwal (351.9 per sq km), and Pakhadi (316.9 per sq km) panchayats; however, the highest household size was recorded in Mandawari (7.7) and Har-sulia (7.2) panchayats. The total literacy rate is higher in the

Springer
Phagi panchayat (69.02%) and lowest in Chandma Kalan (54.53%). The male literacy rate is high in Phagi (82.73%), followed by Gohindi (82.23%), Renwal (81.58%), and Mohabbatpura (81%), and lowest in Kishorepura (69.4%). However, the female literacy rate is highest in Renwal (55%) and lowest in Mandor panchayat (36.9%).

Data collection and analysis

The household survey data collection took place from December 2019 to March 2020. Since the study area is vast, so random sampling technique was used for the selection of the representative number of respondents from 67 villages in the Phagi tehsil. A total of 319 respondents were taken for the proposed study from 32 panchayats of Phagi tehsil. Basic information collected about the respondents for this study was age, gender, number of family members, education level, economic status, and female participation in sanitation facilities decision-making. The information related to WaSH indicators such as the source of water, drinking water quality, toilet facility availability at home, toilet facility usage, water supply in the toilet, toilet cleanliness, ventilation, and soap hand-wash practices were also collected to assess the existing WaSH conditions and defecation practices in different villages of Phagi tehsil. Actual toilet functionality, ventilation, and cleanliness were physically observed during the survey. Groundwater level and groundwater quality data like total dissolved solids (TDS), fluoride, total hardness (TH), and chloride were collected from State Ground Water Department (SGWD), Jaipur, for the year 2019. The spatial analyst tool of ArcGIS software (ESRI 2011) was used to interpolate the groundwater level and groundwater quality data using the ordinary kriging technique, followed by zonal statistics calculation for 32 panchayats of the study area. Thematic layers such as the tehsil boundary and panchayat boundary were also generated in the ArcGIS environment. The detailed methodological design of the research is described in Fig. 2.

WaSH index estimation and risk mapping

The performance of water, sanitation, and hygiene indices are sensitive to the set of indicators used for their calculation. Hence, the WaSH indicators were classified into three sub-indices as water sub-index, sanitation sub-index, and hygiene sub-index for WaSH index estimation and evaluation of risk areas. Different categories of qualitative and quantitative data (Table 1) were standardized using the grade-weighted method (Yu et al. 2019) based on expert opinion and the actual characteristics. For this purpose, every individual indicator is categorized into different classes and assigned weights ranging from 0 (worst) to 1 (best), with 0.2 intervals, for scaling the data and assessment of sub-indices (Tsesmelis et al. 2020). This is followed by
the calculation of the sub-indices scores based on the aggregation of the indicators. The aggregation of indicators was performed using the weighted arithmetic mean, as shown in Eq. (1). This process was also adopted for the estimation of the WaSH Index, as it incorporated all three sub-indices mentioned above.

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

(1)

The WaSH index scores were further categorized into four risk categories ranging from ‘no risk’ to ‘high risk’ and mapped at the panchayat level.

**WaSH index analysis using support vector machine regression**

In this research, the WaSH index was analyzed using SVMR. The basic principle of the SVMR is the mapping of the inputs either linearly or non-linearly into a probable higher aspect of feature space. It constructs a classifier using available samples and avoids misclassification in future predictions (Kurniawan et al. 2021). The SVMRs effectiveness for classification and regression completely depends on the function type of kernel (linear, polynomial, and radial), \(\varepsilon\)-insensitive loss function, and capacity parameter \(C\) (Singh et al. 2011). The SVMR comprises the structural risk minimization (SRM) principle, which is better than the empirical risk minimization (ERM) principle used in neural networks conventionally (Talesh et al. 2019). For SVMR, the dataset is distributed into three sets for training, validation, and testing. The training data was further used to prepare the model automatically. The model network was designed with eight parameters (Table 2) as input parameters (age, literacy, level of education, economic status, participation of females in sanitation facilities decision-making, open defecation practice, water supply in the toilet, and water level) and the WaSH index as the output parameter.

Three different models of SVMR such as partial least squares (PLS), standard support vector regression (S-SVR), and least squares support vector regression (LS-SVR) were used in this research for the prediction of the WaSH index and to understand the variable relation with WaSH index. PLS algorithm develops the original \(X\) space onto a new one and determines the linear correlation among the new variables and the \(Y\) values. The latent variable plays a significant role in the parameters of the algorithm, which bears the covariance structures between the new \(X\) space and the \(Y\) values. After the observation of the data samples from each block variable, PLS changes the matrix of zero-mean variables \(X\) \((n \times N)\) and the \((n \times M)\) matrix of zero-mean variables \(Y\) into the form
| Sub-index | Indicator description                                      | Class                        | Weights | Mean of weighted values | SD* |
|-----------|-----------------------------------------------------------|------------------------------|---------|-------------------------|-----|
| **Water** | Source of water (source of water supply in the house)     | Government supply           | 1.0     | 0.69                    | 0.19 |
|           |                                                            | Government supply + groundwater | 0.6     |                         |     |
|           |                                                            | Tanker                      | 0.4     |                         |     |
|           |                                                            | Borewell                    | 0.2     |                         |     |
|           |                                                            | Handpump                    |         |                         |     |
|           | Drinkable quality (drinking water quality according to respondent) | Yes                          | 1.0     | 0.58                    | 0.2  |
|           |                                                            | No                           | 0.0     |                         |     |
|           | Total dissolved solids – mg/L (TDS in groundwater used for drinking purpose) | <1500                        | 1.0     | 0.7                     | 0.2  |
|           |                                                            | 1500–2000                   | 0.8     |                         |     |
|           |                                                            | 2000–2500                   | 0.6     |                         |     |
|           |                                                            | 2500–3000                   | 0.4     |                         |     |
|           |                                                            | >3000                       | 0.2     |                         |     |
|           | Chloride – mg/L (chloride in groundwater used for drinking purpose) | <400                         | 1.0     | 0.56                    | 0.21 |
|           |                                                            | 400–600                     | 0.8     |                         |     |
|           |                                                            | 600–800                     | 0.6     |                         |     |
|           |                                                            | 800–1000                    | 0.4     |                         |     |
|           |                                                            | >1000                       | 0.2     |                         |     |
|           | Fluoride – mg/L (fluoride in groundwater used for drinking purpose) | <2.0                         | 1.0     | 0.6                     | 0.23 |
|           |                                                            | 2.0–2.5                     | 0.8     |                         |     |
|           |                                                            | 2.5–3.0                     | 0.6     |                         |     |
|           |                                                            | 3.0–3.5                     | 0.4     |                         |     |
|           |                                                            | >3.5                        | 0.2     |                         |     |
|           | Total hardness – mg/L as CaCO₃ (hardness in groundwater used for drinking purpose) | <200                         | 1.0     | 0.7                     | 0.18 |
|           |                                                            | 200–300                     | 0.8     |                         |     |
|           |                                                            | 300–400                     | 0.6     |                         |     |
|           |                                                            | 400–500                     | 0.4     |                         |     |
|           |                                                            | >500                        | 0.2     |                         |     |
| **Sanitation** | Toilet facility available at home (toilet ownership) | Yes                          | 1.0     | 0.78                    | 0.22 |
|           |                                                            | No                           | 0.0     |                         |     |
|           | Toilet facility usage (whether toilet facility is used in the house or not) | Yes                          | 1.0     | 0.65                    | 0.24 |
|           |                                                            | No                           | 0.0     |                         |     |
|           | Toilet facility used by all family members | Yes                          | 1.0     | 0.59                    | 0.25 |
|           |                                                            | No                           | 0.0     |                         |     |
|           | Toilet facility used by children | Yes                          | 1.0     | 0.54                    | 0.26 |
|           |                                                            | No                           | 0.0     |                         |     |
|           | Water supply in the toilet (running tap facility in toilet) | Yes                          | 1.0     | 0.59                    | 0.26 |
|           |                                                            | No                           | 0.0     |                         |     |
|           | Toilet connected with septic tank | Yes                          | 1.0     | 0.62                    | 0.25 |
|           |                                                            | No                           | 0.0     |                         |     |
| **Hygiene** | Toilet cleanliness | Yes                          | 1.0     | 0.63                    | 0.19 |
|           |                                                            | No                           | 0.0     |                         |     |
|           | Ventilation in toilet | Yes                          | 1.0     | 0.41                    | 0.12 |
|           |                                                            | No                           | 0.0     |                         |     |
|           | Soap hand-wash after defecation | Yes                          | 1.0     | 0.58                    | 0.16 |
|           |                                                            | No                           | 0.0     |                         |     |
|           | Soap hand-wash before meal | Yes                          | 1.0     | 0.25                    | 0.11 |
|           |                                                            | No                           | 0.0     |                         |     |

*SD refers to standard deviation*
where extracted score vectors (components, latent vectors) were defined as \((n \times p)\) matrices of the \(T\), \(U\), whereas the \((M \times p)\) matrix \(Q\) and the \((N \times p)\) matrix \(P\) define matrices of loadings. The \((n \times N)\) matrix \(E\) and the \((n \times M)\) matrix \(F\) describe the matrices of residuals. S-SVR acknowledges the existence of non-linearity in the dataset and is used as an efficient prediction model. The function of linear regression is shown below (Eq. 4).

\[
f(x) = \omega \cdot \phi(x) + b
\]

where nonlinear function such as radial is defined as \(f(x)\) with the weight and bias vectors \(\omega\) and \(b\), respectively. LS-SVR incorporates the benefits of the S-SVR and does not use the epsilon-insensitive approach to penalize the squared error of each point. Thus, LS-SVR explains several linear equations of the solutions instead of the quadratic programming problem of S-SVR. The LS-SVR model can be generated by implementing a nonlinear mapping function \(f(x)\) (Eq. 5).

\[
f(x) = w^T \cdot \phi(x) + b
\]

In this equation (Eq. (5)), the weight of the vector and the bias term has been represented by \(w\) and \(b\), respectively. SVMR models were applied with RBF kernel function to predict the target variable, WaSH index using a set of independent variables. The performance of the SVMR models was measured by computing three validation statistical indexes, i.e., coefficient of determination \((R^2)\), coefficient of correlation (CC), and root mean square error (RMSE).

### Results and discussion

#### Evaluation of water, sanitation, and defecation practices

From 67 villages of Phagi tehsil, a total of 319 questionnaire-based surveys were conducted with one individual from each household. Taking into account that in rural areas women are suffering more from health-related issues due to unsafe sanitation and hygiene practices, women were the preferred interviewees in the study. The number of female respondents was 195 (61.1%), and the study had 124 (38.9%) male respondents (Table 3). Age distribution indicates more participation of the 21–40 age group individual in the survey. Literacy data reveals that male respondents have higher literacy (68.6%) than female respondents (47.7%). Education level data shows a high number of male respondents with

### Table 2  Categorization and weighting of input variables used in the SVMR model

| Variables description                                                                 | Class       | Weights | Mean of weighted values | SD* |
|---------------------------------------------------------------------------------------|-------------|---------|-------------------------|-----|
| Age (age of respondents)                                                              | <30         | 1.0     | 0.69                    | 0.09|
|                                                                                        | 30–40       | 0.8     |                         |     |
|                                                                                        | 40–50       | 0.6     |                         |     |
|                                                                                        | 50–60       | 0.4     |                         |     |
|                                                                                        | >60         | 0.2     |                         |     |
| Literacy                                                                              | Yes         | 1.0     | 0.56                    | 0.14|
|                                                                                        | No          | 0.0     |                         |     |
| Education level (respondents with secondary education, i.e., above 9th standard)    | Yes         | 1.0     | 0.26                    | 0.17|
|                                                                                        | No          | 0.0     |                         |     |
| Economic status (above poverty line respondents)                                     | Yes         | 1.0     | 0.59                    | 0.19|
|                                                                                        | No          | 0.0     |                         |     |
| Open defecation practice (respondents open defecation practices)                    | No          | 1.0     | 0.65                    | 0.24|
|                                                                                        | Yes         | 0.0     |                         |     |
| Females’ role in decision-making (participation of females in sanitation facility    | Yes         | 1.0     | 0.25                    | 0.13|
| decision-making)                                                                      | No          | 0.0     |                         |     |
| Water level – meter below ground level (groundwater level in villages)              | <8.0        | 1.0     | 0.68                    | 0.22|
|                                                                                        | 8.0–12.0    | 0.8     |                         |     |
|                                                                                        | 12.0–16.0   | 0.6     |                         |     |
|                                                                                        | 16.0–20.0   | 0.4     |                         |     |
|                                                                                        | >20.0       | 0.2     |                         |     |
| Water supply in the toilet (running tap facility in toilet)                         | Yes         | 1.0     | 0.59                    | 0.26|
|                                                                                        | No          | 0.0     |                         |     |

*SD refers to standard deviation*
primary 31% and secondary 38% education in comparison to female respondents, i.e., 29% primary and 19% secondary education, respectively.

The survey results (Table 4) show that the main source of water in the Phagi tehsil is piped and groundwater (60%). The government water supply in most of the villages is through public taps located at varying intervals on the street. However, the villagers have to depend on groundwater also because of irregular (supplied every 2–3 days, not daily) government water supply. In the absence of borewell and handpump water sources, 6% of respondents are using tanker water (water supply through private water tankers). Overall, groundwater is the main source of water in the study area. About 57.7% of respondents are satisfied with water quality; however, 42.3% of respondents stated that water is hard and not suitable for drinking. The groundwater quality data analysis reveals that the TDS level is more than 1000 mg/L in all the panchayats of the study area, which reflects the unsuitability of water for drinking purposes (Adimalla and Wu 2019). Total hardness (TH) as calcium carbonate (CaCO₃) ranged from 178.3 to 567.3 mg/L, with a mean of 301.8 mg/L. As per Sawyer and McCarty’s (1967) classification, groundwater is considered as soft with TH < 75 mg/L as CaCO₃, moderate hard with 75–150 mg/L, 150–300 mg/L as hard, and > 300 mg/L as very hard.

Results indicate that 19 panchayats have hard groundwater, while 13 panchayats fall in the category of very hard groundwater with 567.31 mg/L as CaCO₃ in Lasadia, 554.42 mg/L as CaCO₃ in Nimeda, 404.2 mg/L as CaCO₃ in Mohabbatpura, and 385.57 mg/L as CaCO₃ in Menda was. High hardness in groundwater may be due to carbonate sources (Koffi et al. 2017). The concentration of chloride ranged from 371 to 1409 mg/L, while the maximum allowable limit for chloride is 600 mg/L (WHO 2017). Very high concentration of chloride was observed in all the panchayats except Kishorepura (445.65 mg/L), Chittora (409.8 mg/L), Mohanpura (371.03 mg/L), Renwai (539.05 mg/L), Phagi (572.77 mg/L), Chakwara (525 mg/L), and Choru (594.77 mg/L). The excess chloride content in groundwater is considered an index of pollution and is known to have adverse impacts on human health (Li et al. 2018). In the study area, a very high concentration of fluoride (1.5 to 4.4 mg/L) was also observed in all the panchayats. Exposure to high fluoride content in drinking water usually results in dental and skeletal fluorosis (Sharma et al. 2015).

The present study results highlight the gap between toilet ownership and its usage by household members. About 75.5% of households interviewed during the survey had toilet facilities in the house due to the government flagship

Table 3  Descriptive characteristics of respondents

| S.N | Characteristics       | Male respondents (n = 124), n (%) | Female respondents (n = 195), n (%) |
|-----|----------------------|----------------------------------|-----------------------------------|
| 1   | Age                  |                                  |                                   |
|     | 15–20                | 16 (39)                          | 25 (61)                           |
|     | 21–40                | 52 (36.6)                        | 90 (63.4)                         |
|     | 41–60                | 35 (37.6)                        | 58 (62.4)                         |
|     | > 60                 | 21 (48.8)                        | 22 (51.2)                         |
| 2   | Literacy             |                                  |                                   |
|     | Yes                  | 85 (47.8)                        | 93 (52.2)                         |
|     | No                   | 39 (27.7)                        | 102 (72.3)                        |
| 3   | Level of education   |                                  |                                   |
|     | Primary (1st to 8th standard) | 38 (40.4) | 56 (59.6) |
|     | Secondary (above 9th standard) | 47 (56)   | 37 (44)  |

Table 4  Water sources and quality

| S.N | Main water sources and drinking quality | Respondents (n = 319), n (%) |
|-----|----------------------------------------|-----------------------------|
| 1   | Source of water                        |                             |
|     | Piped water                            | 91 (28.5)                   |
|     | Piped water + ground water             | 99 (31)                     |
|     | Tanker                                 | 19 (6.0)                    |
|     | Borewell                               | 74 (23.2)                   |
|     | Handpump                               | 36 (11.3)                   |
| 2   | Drinkable quality                      |                             |
|     | Yes                                    | 184 (57.7)                  |
|     | No                                     | 135 (42.3)                  |
| 3   | Groundwater quality parameters         |                             |
|     | TDS (mg/L)                             | Minimum 1248.7, Maximum 3037.1, Mean 2010.1 |
|     | TH (mg/L as CaCO₃)                     | Minimum 178.3, Maximum 567.3, Mean 301.8 |
|     | Chloride (mg/L)                        | Minimum 371, Maximum 1409, Mean 737.3 |
|     | Fluoride (mg/L)                        | Minimum 1.5, Maximum 4.4, Mean 2.9 |
programs; however, only 62.4% of those households actually use them, and the remaining converted these toilets into the storage area. Other studies (Coffey et al. 2014; Barnard et al. 2013; Lee 2017) have also shown that latrine ownership has increased, but more than a third of those were not being used by the households. It is found that out of 32 panchayats, only three panchayats, namely Mandawari (19%), Mendwas (42%), and Rotwara (33%), have less than 50% toilet ownership (Fig. 3a), which indicates the improvement in sanitation practices due to government efforts. Toilet usage is very less in Chanda Kalan (36%), Mendwas (17%), Mandawari (19%), Rotwara (33%), and Gohindi (38%) panchayats (Fig. 3b). However, 56.7% of respondents stated that the toilet facility is used by all family members.

The reasons for not using toilets were water scarcity (27.5%), tradition (47.5%) especially among the elders of the village, bad odor (15%), and costly maintenance (10%). The results reveal that only six panchayats out of 32 panchayats in Phagi tehsil have almost zero open defecation practices and 12 panchayats fall in more than 40% open defecation category. Predominantly open areas such as barren land and agricultural fields outside the village were used for defecation (Geetha and Srinivasan 2014). Respondents considered open defecation as a social outing, and it eliminates the need to maintain the toilet. During the survey, it was observed that open defecation practice is least in the areas with less open areas and fenced agricultural fields. Since a large number of the rural populations in India are still defecating in open areas; hence, proper awareness is required to alleviate the problems associated with open defecation (Anuradha et al. 2017).

It is clear from Fig. 3b that water supply in toilets is affecting the toilet usage in households as both are following the almost same trend. The survey results demonstrate that nine panchayats have less than 50% literacy and the literacy rate is high in males compared to females. The high literacy level of women helps to improve children’s health issues and gain access to safe WaSH facilities (Bisung and

---

**Fig. 3** Sanitation and defecation practices at panchayat level for Phagi tehsil: (a) toilet ownership and open defecation practice, (b) toilet usage and running tap facility in toilet, and (c) total literacy rate and female participation in sanitation facilities decision-making.
Dickin 2019). However, WaSH condition is worsening in Phagi tehsil as most of the panchayats have less than 30% female participation in sanitation facilities decision-making (Fig. 3c). It can be understood from this survey that the existence of toilets, water supply in toilets, and high literacy rate could lead to an increase in toilet usage. Women in rural areas of India are usually less involved with decisions on spending for water and sanitation facilities compared to men (Routray et al. 2017). Therefore, household decision-making has a great influence on the outcomes of WASH interventions (Dery et al. 2020).

Several studies have highlighted the importance of hand-washing in the reduction of fecal–oral disease transmission paths (Cairncross 2003; Fewtrell et al. 2005; Herbst et al. 2009). The hygiene conditions have a huge impact on health; hence, soap hand-wash data was collected through the survey to understand the scenario in the villages of Phagi tehsil. It is found that age plays a significant role in soap usage for hand-washing after defecation and before meals. The results indicate that respondents below 20 years washed hands with soap after defecation (69.2%), whereas people above 60 years are least to use soap for hand-washing (32.6%). People in the surveyed villages have the poor habit of not washing hands before meals (Table 5). It was found that only 59.2% of respondents in the age group of 15–40 years washed their hands before meals. The survey results highlight that hand-washing with soap after defecation and before meals is common among those less than 40 years aged people due to education and awareness about good hygiene practices (Banda et al. 2007).

### WaSH risk areas

The sub-index calculation and risk areas categorization of panchayats based on the WaSH index were evaluated for further improvement in the study area. Water sub-index results reveal that only Renwal and Mohanpura panchayat was under the good category with 0.81 scores based on primary and secondary data; however, twelve panchayats scored 0.6 to 0.8 (Fig. 4a). For the sanitation sub-index calculation, six indicators were used to assess the existing sanitation practices in the study area using the household survey data. The results reveal that sanitation condition is worst in the Mandawari panchayat which scored only 0.18 (Fig. 4b), followed by Mendwas (0.20), Rotwara (0.28), Chanda Kalan (0.33), and Gohindi (0.36). In Madawari panchayat, out of sixteen respondents, only three had the toilet facility at home and used it. The low scores show that open defecation practices are still prevalent in the panchayats.

Eight panchayats of the study area viz. Nimeda, Madhoraipur, Phagi, Dabich Chittora, and Renwal scored more than 0.8, indicating very good sanitation practices adopted by villagers with no or very less open defecation. The hygiene sub-index was assessed based on four indicators, and results reveal that in Chanda Kalan panchayat, the hand-washing using soap after defecation (score 0.18) and before the meal (score 0.09) is not much practiced by villagers. Out of thirty-two panchayats, only five panchayats such as Phagi, Renwal, Dabich, Mohabbatpura, and Mohanpura scored 0.6 to 0.8 (Fig. 4c), reflecting the awareness among villagers related to cleanliness and hygiene.

The aggregation of three sub-indices was used for the estimation of the WaSH index for all the 32 panchayats of the Phagi tehsil (Fig. 4d). The results indicate that four panchayats viz. Mandawari, Mendwas, Chanda Kalan, and Rotwara have the worst conditions with WaSH scores of 0.34, 0.35, 0.39, and 0.38, respectively. However, four panchayats viz. Renwal, Phagi, Dabich, and Mohanpura show no risk conditions with WaSH scores higher than 0.8.

### WaSH index prediction using SVMR models

The household survey findings show that WaSH conditions in rural areas are determined by different factors like gender inequality, education, and water supply in toilets. In this research, machine learning technique such as SVMR was used to find the correlation and for predicting the WaSH index. In SVMR, the WaSH index was used as the dependent variable, whereas the eight variables (age, literacy, level of education, economic status, participation of females in sanitation facilities decision-making, open defecation practice, water supply in the toilet, and groundwater level) constituted the set of independent variables. Three different models of SVMR, PLS, S-SVR, and LS-SVR were used to predict the WaSH index and to understand the variables associated with the WaSH index. The data were divided into the learning

| Age (years) | Use soap after defecation (n = 182), n (%) | Soap not used after defecation (n = 137), n (%) | Use soap before meal (n = 79), n (%) | Soap not used before meal (n = 240), n (%) |
|-------------|------------------------------------------|-----------------------------------------------|-----------------------------------|---------------------------------------|
| 15–20       | 27 (69.2)                                 | 12 (30.8)                                     | 10 (25.6)                         | 29 (74.4)                             |
| 21–40       | 93 (65)                                   | 50 (35)                                       | 48 (33.6)                         | 95 (66.4)                             |
| 41–60       | 48 (51)                                   | 46 (49)                                       | 16 (17)                           | 78 (83)                               |
| > 60        | 14 (32.6)                                 | 29 (67.4)                                     | 5 (11.6)                          | 38 (88.4)                             |
process (70% of the dataset) and the testing process (the remaining 30%). The data for both learning and testing were selected randomly from a total of 32 panchayats to avoid estimation biases. RBF kernel is used to provide good results under the assumption of general smoothness (Wang et al. 2018; Yahya et al. 2019) among the kernels such as linear, sigmoid, polynomial, and RBF. The tenfold cross-validation process was repeated twenty times to derive the SVMR model parameters.

The actual and expected values for the WaSH index were calculated using different models, as shown in Fig. 5. The appropriate model was identified based on minimum values of RMSE, high $R^2$, and CC for assessment of the WaSH index. Table 6 displays a comparative analysis of the predictive performances of the SVMR models.

LS-SVR is capable of producing the most accurate estimates of the WaSH index with the coefficient of determination (0.902 and 0.877) and root means square error (0.041 and 0.05) in the training and testing stages, respectively. It can be seen that both models (PLS and LS-SVR) are capable of predicting the WaSH index accurately, but LS-SVR outperformed PLS. Figure 6 shows the actual and predicted values of the WaSH index at the panchayat level. It is found from the SVMR analysis that WaSH index shows positive correlation with open defecation ($r=0.94$), water supply in toilet ($r=0.92$), participation of females in sanitation facilities decision-making ($r=0.53$), followed by literacy rate ($r=0.33$) and economic status ($r=0.27$). The findings of this study emphasize the importance of water supply in toilets, literacy level, and participation of females in decision-making to WaSH Index scores (Hirai et al. 2016; WaterAid 2017).

**Conclusion**

This study assessed and quantified the domestic WaSH conditions in the rural areas of Rajasthan state in India at the panchayat level using GPS-based household survey data and
advanced computing techniques. The survey data reflect that factors such as the presence of toilets at home, water supply in toilets, high literacy rate, and female participation in sanitation facilities decision-making could help to reduce open defecation and improve hygiene practices in villages.

In order to understand and evaluate the practices with regard to existing water, sanitation, and defecation practices, a suitable index is developed for the spatial assessment of WaSH conditions. The WaSH risk areas were also identified for further improvement and ease of management by planners.

**Table 6** Comparison of SVMR models

| Model            | Training | Test    |
|------------------|----------|---------|
|                 | CC  | $R^2$ | RMSE | CC  | $R^2$ | RMSE |
| PLS              | 0.898 | 0.898 | 0.042 | 0.869 | 0.862 | 0.053 |
| Standard SVR     | 0.866 | 0.727 | 0.068 | 0.841 | 0.699 | 0.078 |
| LS-SVR           | 0.904 | 0.902 | 0.041 | 0.882 | 0.877 | 0.050 |

**Fig. 5** Scatter plots of actual and predicted values of WaSH index using different models.
at the local level to reduce the gaps in toilet ownership and usage. The integration of GIS and soft computing methods permitted a more in-depth examination of WaSH and behavioural determinants. Three different models of SVMR viz. PLS, S-SVR, and LS-SVR were used to predict the WaSH index and to understand the variables associated with the WaSH index. SVMR results reveal a strong correlation of the WaSH index with open defecation and water supply in toilets. The survey data also elucidate that most of the respondents consider open field defecation very economical as this eliminates the need to maintain toilets and because of the scarcity of water. Therefore, education and awareness campaigns on health and hygiene are essential to improve the WaSH condition in rural areas.

Acknowledgements This research was funded by the Department of Science & Technology – KIRAN Division, Ministry of Science and Technology, Government of India, under the Women Scientist scheme (WOS-B) project vide reference number DST/WOS-B/2018/1575/ETD/Ankita.

Author contribution All authors contributed to this manuscript. Dr. Ankita P. Dadhich conceived and designed the analysis, performed the analysis, and wrote the manuscript. Dr. Pran Nath Dadhich conceived and collected data and provided inputs in writing the manuscript. Dr. Rohit Goyal provided inputs in data analysis.

Funding This work was supported by the Department of Science & Technology, Government of India (grant number DST/WOS-B/2018/1575/ETD/Ankita) under Women Scientist Scheme (WOS-B). Author Ankita P. Dadhich has received research support from the Department of Science & Technology.

Data availability It will be provided as per the requirement of the editor/reviewer.

Declarations

Ethics approval Not applicable.

Consent to participate This manuscript, in part or in full, has not been submitted or published anywhere. The manuscript will not be submitted elsewhere until the editorial/review process is completed.

Consent for publication Hereby all authors agree to the publication of this manuscript in ESPR.

Competing interests The authors declare no competing interests.

References

Adimalla N, Wu J (2019) Groundwater quality and associated health risks in a semi-arid region of south India: implication to sustainable groundwater management. Hum Ecol Risk Assess 25(1–2):191–216

Anuradha R, Dutta R, Raja JD, Lawrence D, Timsi J, Sivaprakasam P (2017) Role of community in Swachh Bharat Mission. Their knowledge, attitude and practices of sanitary latrine usage in rural areas, Tamil Nadu. Indian J Community Med 42(2):107–110

Banda K, Sarkar R, Gopal S, Govindarajan J, Harijan BB, Jeyakumar MB (2007) Water handling, sanitation and defecation practices in rural southern India: a knowledge, attitude and practices study. Trans R Soc Trop Med Hyg 101:1124–1130

Barnard S, Routray P, Majorin F, Peletz R, Boisson S, Sinha A, Clason T (2013) Impact of Indian total sanitation campaign on latrine coverage and use: a cross-sectional study in Orissa three years following programme implementation. PLoS One 8(8):e71438. https://doi.org/10.1371/journal.pone.0071438

Bisung E, Dickin S (2019) Concept mapping: engaging stakeholders to identify factors that contribute to empowerment in the water and sanitation sector in West Africa. SSM-Population Health 9:100490

Cairncross S (2003) Handwashing with soap—a new way to prevent ARIs? Trop Med Int Health 8(8):677–679

Census (2011) Primary census abstracts, registrar general of India, ministry of home affairs, government of India. URL: https://www.census2011.co.in/census/district/435-jaipur.html. Accessed Dec 2019

Chan EY, Tong KH, Dubois C, Mc Donnell K, Kim JH, Hung KK, Kwok KO (2021) Narrative review of primary preventive interventions against water-borne diseases: scientific evidence of...
Yahya ASA, Ahmed AN, Othman FB, Ibrahim RK, Afan HA, El-Shafie A, Fai CM, Hossain MS, Ehteram M (2019) Water quality prediction model based support vector machine model for ungauged river catchment under dual scenarios. Water 11(6):1231

Yoon H, Kim Y, Ha K, Lee SH, Kim GP (2017) Comparative evaluation of ANN- and SVM-time series models for predicting freshwater-saltwater interface fluctuations. Water 9(5):323

Yu X, Li Y, Xi M, Kong F, Pang M, Yu Z (2019) Ecological vulnerability analysis of Beidagang National Park, China. Front Earth Sci 13:385–397. https://doi.org/10.1007/s11707-018-0726-8

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.