IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 25, 2022, accepted February 10, 2022, date of publication February 15, 2022, date of current version February 28, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3151892

An Improved Imputation Method for Accurate Prediction of Imputed Dataset Based Radon Time Series

ADIL ASLAM MIR[1,2], FATIH VEHBI ÇELEBI[1], MUHAMMAD RAFIQUE[3], LAL HUSSAIN[2,4], AHMED S. ALMASOUD[5], MASOUD ALAJMI[6], (Member, IEEE), FAHD N. AL-WESABI[7], AND ANWER MUSTAFA HILAL[8]

[1]Department of Computer Engineering, Ankara Yıldırım Beyazıt University, 06010 Ayvalı, Keçiören, Ankara, Turkey
[2]Department of Computer Science and Information Technology, The University of Azad Jammu and Kashmir, King Abdullah Campus, Chatter Kalas, Muzaffarabad, Azad Kashmir 13100, Pakistan
[3]Department of Physics, The University of Azad Jammu and Kashmir, King Abdullah Campus, Chatter Kalas, Muzaffarabad, Azad Kashmir 13100, Pakistan
[4]Department of Computer Science and IT, The University of Azad Jammu and Kashmir, Neelum Campus, Athmuqam, Azad Kashmir 13230, Pakistan
[5]Department of Information Systems, College of Computer and Information Sciences, Prince Sultan University, Riyadh 12435, Saudi Arabia
[6]Department of Computer Engineering, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia
[7]Department of Computer Science, College of Science and Art at Mahayil, King Khalid University, Abha 62529, Saudi Arabia
[8]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia

Corresponding authors: Lal Hussain (lall_hussain2008@live.com) and Fahd N. Al-Wesabi (falwesabi@kku.edu.sa)

Taif University Researchers Supporting Program (project number: TURSP-2020/195), Taif University, Saudi Arabia, supported this research. The authors are grateful to King Khalid University's Deanship of Scientific Research for financing this research under grant number (RGP 1/14/43).The authors would like to express their gratitude to Prince Sultan University for paying the publication's Article Processing Charges (APC). The data used in the current study is a part of the research conducted for the project grant no: 6453/AJK/NRPU/R&D/HEC/2016 against the NRPU project executed by one of the co-authors, MR.

ABSTRACT This article primarily focuses on the performance evaluation of a new methodology, imputation by feature importance (IBFI), to serve its imputed dataset in further regression scenarios when dealing with soil radon gas concentration (SRGC) time-series data. The time-series data have been collected spanning over fourteen(14) months period, which included four seismic events, and have been used for experimentation. The imputation by feature importance (IBFI) has been experimented and obtained results are found more efficient in the imputation of missing patterns in investigated time series when compared to traditionally used imputation methods viz. mean, median, mode, predictive mean matching (PMM), and hot-deck imputation.The IBFI methodology has been used in a variety of settings, such as data missing not at random (MNAR), missing completely at random (MCAR), and missing at random (MAR), with missingness percentages ranging from 10% to 30%. In this study, the imputed datasets, 9 for each imputation method, have been used further to predict the attribute of interest (radon concentration (RN)) keeping others as independent attributes such as thoron, temperature, relative humidity, and pressure time series. Support vector machine (SVM) with linear kernel has been used as a learning algorithm and its performance was evaluated based on the fact that how efficient and unbiased values were imputed. Statistical performance evaluation measures viz. root mean squared log error (RMSLE), root mean square error (RMSE), mean squared error (MSE),and mean absolute percentage error (MAPE) have been calculated for the assessment of performance. The findings of our study show that the IBFI imputed dataset has provided a better-fitted model. The model generation and predictions upon IBFI imputed time series result in more accurate predictions when compared to mean, median, mode, PMM, and hot-deck imputed time series. Furthermore, PMM and median imputed time series also perform closer to the IBFI imputed time series.

INDEX TERMS Predictive mean matching, missingness, radon concentration, support vector machine, imputation, IBFI.

I. INTRODUCTION

Radon gas ^{222}Rn poses health threats to human health and is an immediate decay product of radium ^{226}Ra [1]. The presence of ^{226}Ra is ubiquitous and found in trace amounts in soils and rocks.^{222}Rn, a noble gas, is transported from its place of origin to the surface of the earth and its motion is subjected to geological structures and meteorological factors. It reaches to surface of the earth and exhales within

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

and outside the closed house environment. Along with the characteristics of the building the exhaled radon creates high levels of indoor radon concentrations [2]–[5]. It is found in water, air, and soil, and it concentrates in the environment and buildings in a variety of ways based on numerous geological, chemical, climatic, and other temporally variable elements [2], [3], [6]–[13]. Despite the carcinogenic nature of radon, it has many useful applications including its use as a precursor to the earthquake [14]–[24].

For prediction and forecasting purposes, numerous studies have been carried out by employing different methodologies [25], [26]. Different geophysical and seismological activities occur beneath the surface throughout the earthquake preparation phase. One of the precursors deep down the earth is soil radon gas that is witnessed of anomalous behavior before occurrences of several earthquakes.A variety of research has been conducted around the world in this area, concentrating on earthquake prediction based on anomalous radon gas behavior in the atmosphere, soil, and water [25], [27]–[30]. Furthermore, meteorological variables such as temperature, rainfall, and pressure, among others, influence radon emission dynamics, with typical features persisting for a period. In this regard, numerous studies had been carried out by exploiting different computational intelligence models to understand the correlation between soil radon gas concentration and different meteorological parameters [11], [31]–[33]. Radon and thoron time series are subject to nonlinear processes and extracting some meaningful information from such series is not an easy task and needs the use of modern computational techniques. Detrended fluctuation analysis (DFA), detrended cross-correlation analysis (DCCA), and multifractal detrended fluctuation analysis (MF-DFA) of soil radon (^{222}Rn) and thoron(^{220}Rn) time series have been used to find long-range correlations and characterization of correlated data of more than one non-stationary time series and to examine the scaling and multifractal features of radon and thoron time series [29], [34].

Missing patterns in the time series data are often encountered by many researchers during their scientific experimentations and result in unreliable predictions or modeling if these missing patterns are not properly imputed. The correct and unbiased imputations improve the performance of the dataset for further analysis and experimentation. There occurs a variety of circumstances that leads to the missingness of data. This includes machine malfunctioning, human error, routine maintenance, etc. [35]. The missingness of the data can be classified according to the means through which it is generated [37]. These missing data can be classified as MAR, MCAR, and MNAR when, the missingness of a data point is not related to other missing data but with the observed data, the probability for the missingness is the same for all cases, the hypothetical value determines if a data point is missing, or the cause of missingness is related to the other features in the data, respectively. To impute these missing values, usually simple and straightforward methods are used which include mean, median, mode, missing-indicator methods for example, but results in severely biased estimates and makes it inefficient for further analysis [36], [37]. In addition to it, multiple imputation methods also exist which results in more accurate imputation than other existing conventional methods [38]–[41]. Moreover, a methodology was proposed, imputation by feature importance (IBFI), which iteratively imputes the missing patterns in the data by taking feature importance to dynamically select the best attribute to impute first [42]. The methodology can envelop any machine learning algorithm e.g. Random Forest, naïve Bayes as a base learner method for imputation. Furthermore, to make it more efficient, the learning models have been stored and utilized those models in the subsequent iterations. The reusability of the previously trained model reduces computation time. The detailed understanding of imputation by feature importance (IBFI) is presented in the methodology section.

This study is the progressive stage of the previous work, imputation by feature importance (IBFI), which had been done for the reconstruction of missing patterns in soil radon gas concentration (SRGC) data and has been published elsewhere [42].As stated that the imputed values in a dataset play an important role in further analyses and experimentation. In this regard, the performance evaluation of imputation by feature importance (IBFI), to serve its imputed dataset for further regression scenarios is studied when predicting radon concentration from other meteorological attributes. Imputation by feature importance (IBFI) is applied to reconstruct the missing patterns in soil radon gas concentration (SRGC) data at different missingness scenarios. Using the R package ''mice'' missing data was artificially introduced into the dataset in different missingness scenarios across 10 to 30% [43]. In this paper, the imputed datasets (9 for each imputation method) by IBFI are used further in the regression scenario. For the prediction of radon concentration, the support vector machine (SVM) with the linear kernel is employed as a learning method. The accurate prediction of soil radon gas concentration relies on the accuracy and unbiasedness of the imputed patterns in the soil radon gas concentration (SRGC) dataset. To evaluate the prediction model's performance, the mean absolute percentage error (MAPE), root mean square error (RMSE), mean squared error (MSE) and mean squared log error (RMSLE) are calculated.

II. MATERIAL AND METHODS

This section describes the statistical aspects of the soil radon gas time-series dataset. Furthermore, detailed information about the methodology is also provided in terms of missing values introduction and their imputation by IBFI and other imputation methods. The simulation plan for the prediction of soil radon concentration is presented and its concrete details are also provided. The working procedure of imputation by feature importance(IBFI) for imputation of missing patterns is also discussed. The mathematical formulation of the performance metrics used in this study is also provided.

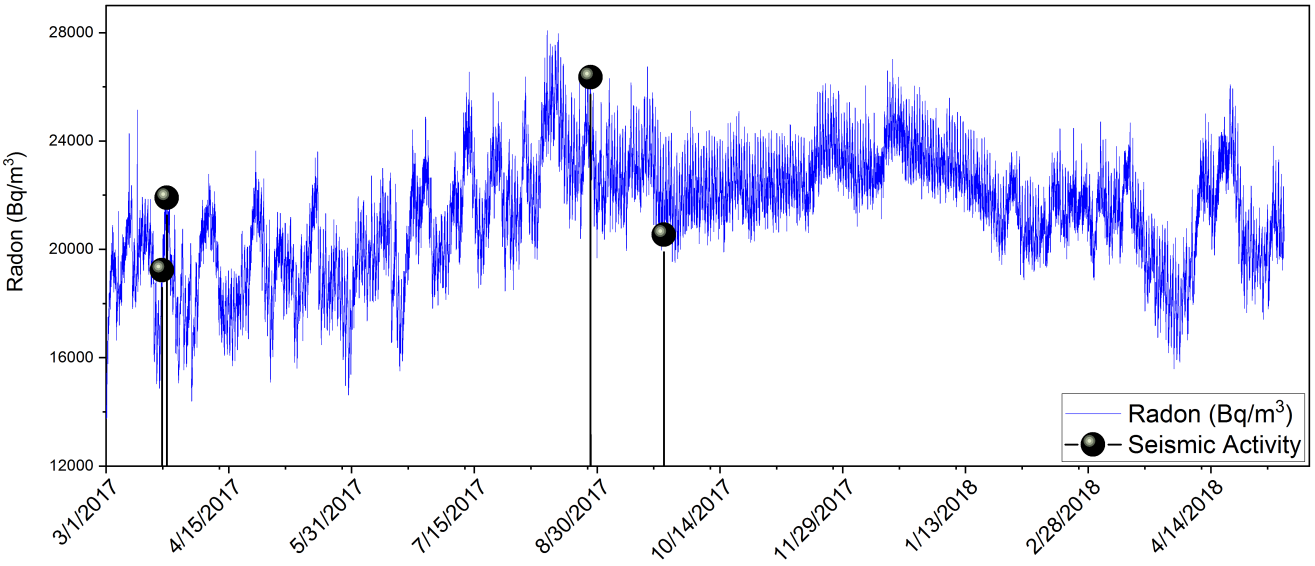

FIGURE 1. The observed soil radon concentration time series data with four earthquakes (Seismic Activities).

A. DATA DESCRIPTION

On the fault line near Muzaffarabad, a city in the Pakistani part of Kashmir, the soil radon gas time series was obtained.To record continuous measurements of radon, thoron, temperature, humidity, and pressure, a humidity-insensitive radon and thoron monitor (SARAD RTM 1688-2, Nuclear Instruments, Germany) had been used at the latitude and longitude of 34.39621 and 73.47347 respectively. For more than 1 year, data is recorded at the interval of 40 minutes and results in 36 samples every 24 hours. Moreover, the resulting data and additional details of instrumentation are reported elsewhere [25], [26], [28]. respectively. The studied data consists of 15692 radon valid observations along with other attributes such as thoron (Bq/m^3), temperature (^0C), relative humidity, and pressure(mbar) ranging from 1^{st} of March 2017 till 11^{th} of May 2018. During the study period, four earthquakes occurred, presented in Figure 1 as black bubbles (21 and 23 March 2017, 27 August 2017, and 23 September 2017). Different statistical measures are calculated for radon concentration (see Figure 2) and other independent features (thoron, temperature, relative humidity, and pressure) (see Table 1). Considering observed radon concentration, the whole period has a minimum concentration of 13743 Bq/m^3, maximum concentration 28085 Bq/m^3, mean concentration 21364 Bq/m^3, and median of 21569Bq/m^3. Moreover, from the statistics shown in Figure 1, the p-value of <0.005 calculated from the Anderson-Darling normality test [44] indicates that there is enough evidence to say that the series is not normally distributed. The detailed statistical summary of other independent attributes such as thoron, temperature, relative humidity, and pressure are presented in Table 1. During the study period, the thoron time series data have a maximum concentration of 16182Bq/m^3, minimum of 1495Bq/m^3 when there was no seismic activity

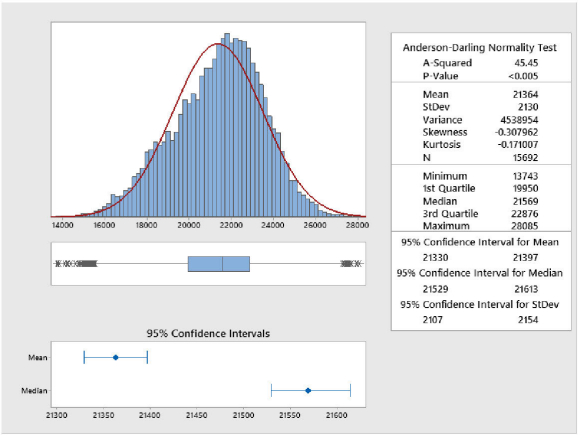

FIGURE 2. Summary of the statistics calculated from soil radon concentration (Bq/m^3) time-series data.

observed. On the other hand, during the time of seismic activities, the minimum and maximum observed thoron concentrations were 1677 Bq/m^3 and 3734 Bq/m^3 respectively. Moreover, the deviation from the mean for temperature, relative humidity, and pressure are higher with the values of 8.097, 13.196 and 4.93 respectively considering normal time series data was observed whereas lower values of standard deviation are observed for seismic activity data except for thoron.

B. PROPOSED SIMULATION AND ANALYSIS PLAN

The complete simulation and analysis plan for the current investigation is shown in Figure 4. To assess the efficiency of imputation methods regarding how much its imputed datasets perform in further analyses, the current study utilizes the imputed datasets by IBFI, mean, median, mode, PMM, and

TABLE 1. Statistical measures calculated from thoron, temperature, relative humidity, and pressure time-series data.

Variable	Activity	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Thoron (Bq/m^3)	NSA	2514.9	3.08	384.0	1495	2247	2489	2760	16182
	SA	2559.8	34.5	413.9	1677	2222.3	2537	2922.8	3734
Temperature(^0c)	NSA	22.475	0.0649	8.097	4	16	23	28.50	42.50
	SA	23.628	0.553	6.638	14.50	18	23.5	28.375	36.50
Relative Humidity	NSA	77.875	0.106	13.196	34	70	81	88	101
	SA	78.847	0.776	9.310	56	72	81	86	92
Pressure (mbar)	NSA	928.26	0.0395	4.93	914	925	929	932	943
	SA	928.42	0.390	4.68	920	924.25	928	932	936

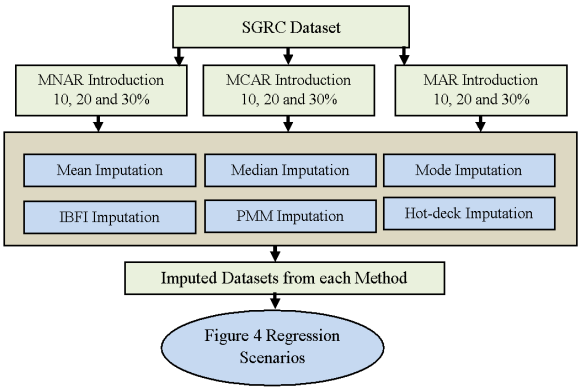

FIGURE 3. Missing values introduction and its imputation by different imputation methods.

hot-deck imputation methods in different missingness scenarios as shown in Figure 3. For this work, the imputed datasets (9 for each imputation method) are used further to predict the soil radon gas concentration from other independent attributes. For experimentation, presented in Figure 4, the imputed dataset is divided into two parts i.e. non-seismic activity data (NSAD) and seismic activity data (SAD). Non-seismic activity data consists of those samples when no earthquake was reported whilst seismic activity data (SAD) consists of samples when there was an earthquake. Because of the unusual behavior of radon before and after the earthquake, research studies have been conducted in the past to identify a certain range of window sizes to predict radon concentration. [25], [32], [45]–[48]. In this paper, the data is partitioned by keeping the window size of 5, which is 5 days before and after the seismic activity or an earthquake. We tested with two distinct settings, setting 1 and setting 2, to predict the radon concentration during different seismic events, as shown in figure 4. As stated above, the experimented data contains four seismic activities which occur during the data recording period. Setting 1 incorporates seismic activity (SA) 1, 2, and 4 with non-seismic activity data (NSAD) to produce a training set, with seismic activity (SA) 3 serving as a test set to assess the performance. In setting 2, seismic activity (SA) 1, 2, and 3 are merged with non-seismic activity data (NSAD) to constitute the training set, with seismic activity (SA) 4 serving as a test set for performance evaluation. Furthermore, the training set is subjected to a support vector machine (SVM) with a linear kernel, yielding a machine learning model. The test set is further passed to the fitted model and predicts the radon concentration. To assess the performance of the fitted model which is trained on different imputed datasets, the different performance metrics are calculated such as RMSE, RMSLE, MAPE, and MSE to estimate the error between actual and predicted radon concentration.

C. IMPUTATION BY FEATURE IMPORTANCE (IBFI) METHOD

Imputation by feature importance (IBFI) is an imputation method that iteratively imputes missing patterns in data using feature importance. It can envelop any machine learning algorithm as a base learning algorithm to impute missing data. The imputation process starts by first splitting the dataset into two parts i.e. impure and pure data as shown in Figure 5. The pure data (PD) consists of those samples from the whole dataset where each sample has available values for all its attributes or features whilst impure data (ID) is constituted by those samples which have one or more values missing that need to be imputed. IBFI provides decision-making on the response variable to choose the best available predictor variables at run-time, resulting in efficient machine learning model development for missing data imputation. Suppose, we have different attributes in a dataset such as $Atrr_1, Atrr_2, \ldots.., Atrr_n$. In a machine learning context, if missing values occur in $Atrr_1$, the attributes $Atrr_2, \ldots.., Atrr_n$ can be used to train any machine learning model. Further, the trained model can be used to forecast $Atrr_1$ value. For the case discussed above, it works efficiently but in the cases where more than one value is missing in the samples and the attributes have strong dependencies among each other, makes the task of the imputation process more challenging. Consider we have 5 attributes in a dataset such as $Atrr_1, Atrr_2, Atrr_3, Atrr_4, Atrr_5$ and the missing values observed in the attributes $Atrr_1 and Atrr_5$. On the other hand, when predicting attributes of interest, we've discovered that certain attributes have a high feature importance when compared to other attributes such as $Atrr_1 and Atrr_5$. Moreover, $Atrr_1 and Atrr_5$ have feature importance values in descending order of $Atrr_5, Atrr_3, Atrr_4, Atrr_2$ and $Atrr_2, Atrr_4, Atrr_1, Atrr_3$ respectively. Conventionally, in the scenarios where $Atrr_1$ is missing, $Atrr_2, Atrr_3, Atrr_4, Atrr_5$ is used to train a machine learning model and for $Atrr_5$, the attributes $Atrr_1, Atrr_2, Atrr_3, Atrr_4$ is used for training and

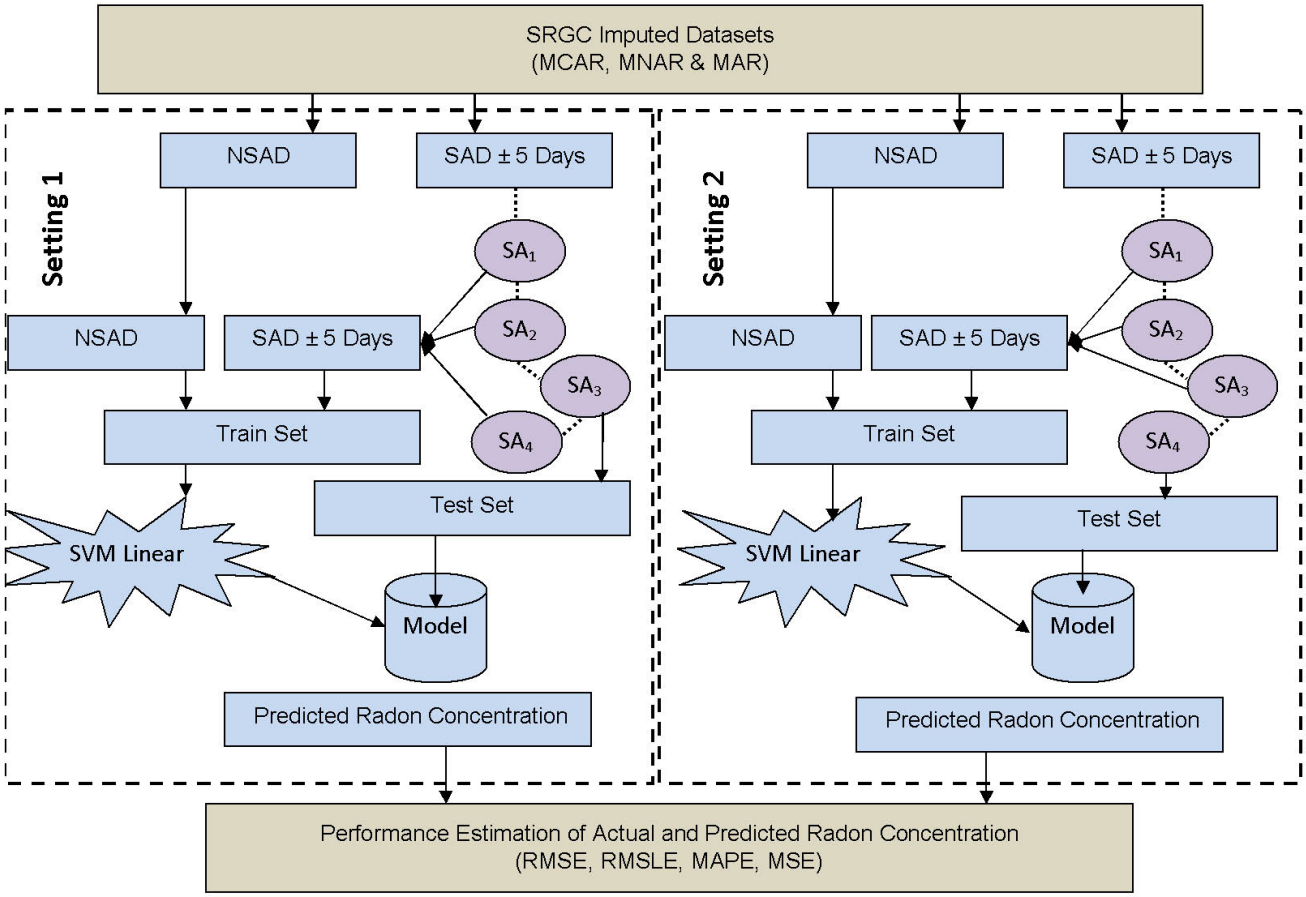

FIGURE 4. Simulation plan of the study.

finally, these fitted models can be used to impute the values in the samples where $Atrr_1 and Atrr_5$ are missing. In this case, we have only 3 attributes, $Atrr_2, Atrr_3, Atrr_4$ for training. To better impute the missing values for $Atrr_1 and Atrr_5$ using machine learning methods, the feature importance vectors shows that $Atrr_5$ is more important when predicting the value of $Atrr_1$, and $Atrr_2$ is the important one when coming to the prediction of the value of $Atrr_5$. IBFI utilizes that fact and decides to impute the value of $Atrr_5$ at first after training from available attributes. Moreover, the imputed value of $Atrr_5$ is further used to predict the value of $Atrr_1$. The decision of selection of aviable predictor features for certain response features at runtime makes IBFI better for imputing missing patterns by enveloping any machine learning method. Imputation by feature importance (IBFI) is an imputation method that iteratively imputes missing data using feature importance. It can envelop any machine learning algorithm as a base learning algorithm to impute missing data. The imputation process starts by first splitting the dataset into two parts i.e. impure and pure data as shown in Figure 5. The pure data (PD) consists of those samples from the whole dataset where each sample has available values for all its attributes or features whilst impure data (ID) is constituted by those samples which have one or more values missing that need to be imputed. IBFI provides decision-making on the best available predictor variables for different response variables at run-time, resulting in efficient machine learning model creation for missing data imputation. In IBFI, the feature importance matrix (FIM) is responsible for the order in which the missing features are imputed. The feature importance matrix is constructed by computing the variable importance for individual attributes in the dataset. This is done by taking each attribute as a response while others as predictors. These feature importance values for individual attributes are arranged in descending order. As presented in figure 5, the IBFI process needs some termination criterion to stop the imputation process, rejection threshold is selected. The rejection threshold determines the extent up to which the number of missing values is imputed per sample in the impure dataset. For the dataset having 5 attributes, the rejection threshold of 2 means that the samples having more than 2 missing values are discarded by the IBFI. Furthermore, by storing the models that are fitted throughout successive iterations, the IBFI methodology utilized these models in subsequent iterations. Models are saved in memory in such a way that if A_1 is a dependent feature and F_2 and F_3 are independent

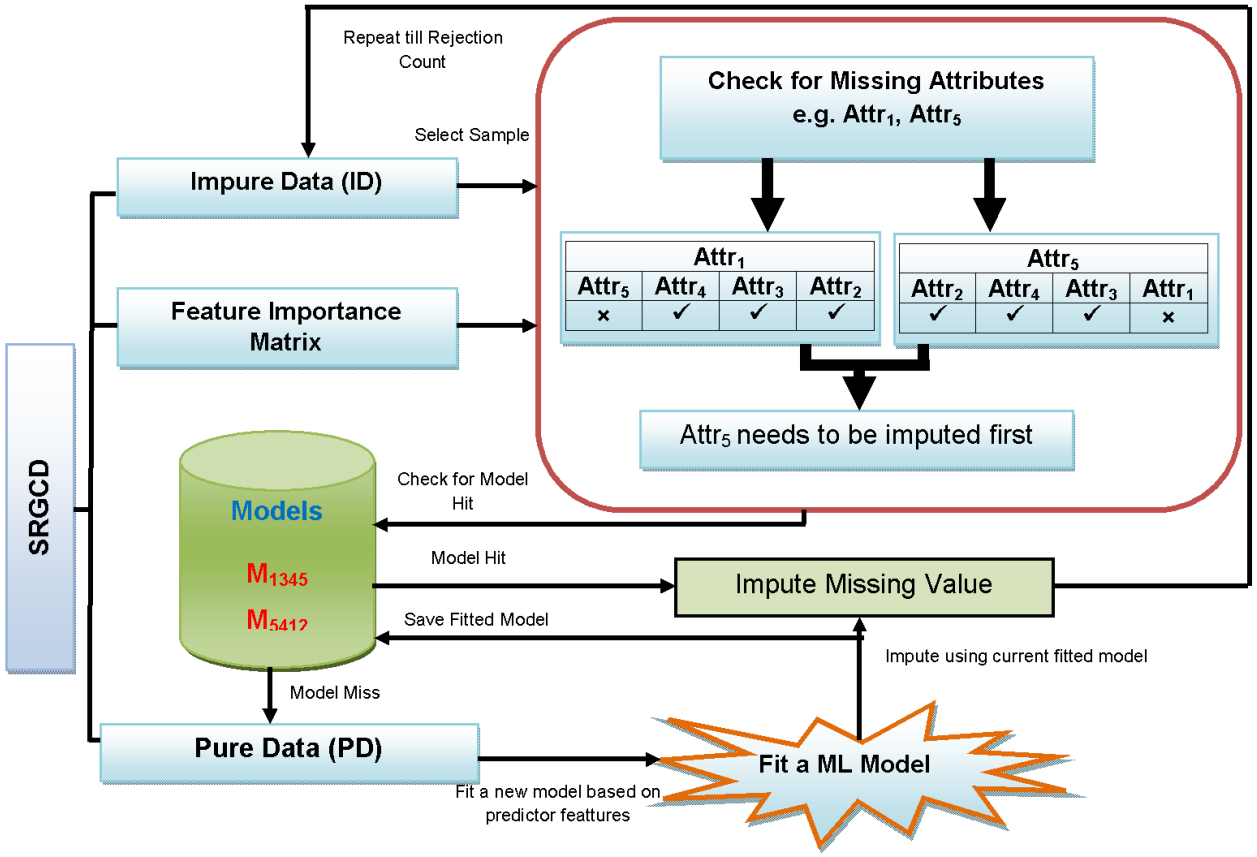

FIGURE 5. Imputation by Feature Importance (IBFI).

features, the model is saved as $Model_{123}$. In later iterations, for example, missing at three features is decreased to missing at two features, and A_1 must be trained again using A_2 and A_3; rather than training another model, the same model $Model_{123}$ will be used to impute the value for F_1.

III. PERFORMANCE MEASURE

In this study, different commonly used performance metrics are computed to analyze the effectiveness of the imputed dataset in predicting radon concentration (RN). The error between actual and predicted radon concentration is computed viz. root mean square error (RMSE), root mean squared log error (RMSLE), mean squared error (MSE) and mean absolute percentage error (MAPE). The root mean square error (RMSE) is a commonly used metric for the evaluation of performance that has been applied to a variety of fields of research where prediction models are of main concern [25], [49], [50]. It is more susceptible to outliers since a considerable divergence between actual and anticipated values has a significant impact on its value. The RMSE can be calculated using the following formula:

$$\text{RMSE} = \sqrt{\frac{1}{V}\sum_{n=1}^{V}(X_n - Y_n)^2}$$
where V represents total number of samples (1)

Because the presence of an outlier might cause the error term to go up while computing RMSE, RMSLE can scale down the outliers and nullify their influence. The RMSLE may be calculated using the following equation:

$$\text{RMSLE} = \sqrt{\frac{1}{V}\sum_{n=1}^{V}(log(X_n + 1) - log(Y_n + 1))^2}$$
where V represents total number of samples (2)

In the cases when the values are higher in number and have an excessive effect of the large differences between predicted and actual values, RMSLE is mostly used in these scenarios. Moreover, the MAPE is also a frequently used performance metric which is used to assess the accurateness of the prediction model, computed from:

$$\text{MAPE} = \frac{1}{V}\sum_{n=1}^{V}\left|\frac{Y_n - X_n}{X_n}\right| \quad (3)$$

The average absolute percentage error is referred to as MAPE. MAPE's scale independence and ease of interpretation are the two qualities that make it popular and helpful [51]. It has certain downsides in addition to its benefits, such as undefined or endless values when the actual values are zero or close to zero. Actual values with a magnitude smaller than one resulted in a greater percentage value for the

TABLE 2. The RMSE values for IBFI and other imputation methods (Mean, Median, Mode, PMM, and Hotdeck) for missingness of 10, 20, and 30% of type MCAR, MNAR, and MAR imputed datasets for predicting radon concentration keeping Setting 1.

	MCAR 10%	MCAR 20%	MCAR 30%	MNAR 10%	MNAR 20%	MNAR 30%	MAR 10%	MAR 20%	MAR 30%
IBFI	1776	1763.3	1657.8	1730.3	1705.8	1699.5	1757.3	1719.1	1760
Mean	1852.5	1919.6	1677.9	1881.2	1915.6	1860.7	1880.3	1891.9	1972.5
Median	1848.2	1908.4	1909	1872.1	1897.4	1824.6	1873.9	1864.6	1935.3
Mode	1928.5	2008.5	1804	1941.5	2044.2	1890.5	1973.9	1962.5	2058.9
PMM	1790.2	1833.6	1719.1	1814.1	1823.3	1822.7	1778.9	1780.3	1806.3
Hot-deck	1888.5	2029.7	1863.3	1994.9	2033.1	2045.1	1956.9	2084.9	2111.5

MAPE, whereas actual zero values resulted in infinite MAPE values [52]. Furthermore, Mean Squared Error (MSE) is a performance statistic that estimates the closeness of the predicted and actual values and is calculated using the following formula:

$$\text{MSE} = \frac{1}{V}\sum_{n=1}^{V}(Y_n - X_n)^2 \tag{4}$$

More precisely, it's the average square difference between the actual and predicted value. The lower the MSE score, the better the prediction model fits the data.

IV. RESULT AND DISCUSSION

When predicting the radon concentration (RN), the RMSE statistics for all imputed datasets (10, 20, and 30% MCAR, MNAR, and MAR) from methods such as IBFI, mean, median, mode, PMM, and hot-deck imputation with setting 1 are shown in Table 2. The learning based on the IBFI imputed datasets in MCAR, MNAR, and MAR-based missingness scenarios, the IBFI imputed dataset performs better than other imputed datasets and results in minimum observed RMSE value. For IBFI imputed dataset, the minimum RMSE value of 1657.8 is observed for MCAR 30%, which is less when compared to hot-deck imputation with the RMSE value of 1863.3 in the same missingness scenario. In the case of MNAR (10 to 30%), for IBFI imputed datasets, the RMSE value observed is less than 1730 across all missingness percentages while PMM performs relatively closer to IBFI with the maximum RMSE value of 1823. The highest RMSE value of 2045.1 is observed in MNAR 30% for the hot-deck imputed dataset. A similar pattern was observed for PMM based imputed datasets in other missingness scenarios such as MCAR 20, MNAR, and MAR 10 to 30% where the difference of RMSE value from IBFI ranged from 21.6 to 123.2 which is very less when compared to other imputed datasets. In MCAR 30%, the mean imputed dataset performs better in predicting radon concentration (RN) with the least difference of RMSE value from IBFI which is 20.1, when compared to other methods, ranging from 61.3 to 251.2. From all the statistics above, it is concluded that IBFI imputed dataset provides more accurate imputations than other imputed datasets such as mean, median, mode, PMM, and hot-deck having the least RMSE values when compared.

The RMSE statistics for IBFI and other imputed datasets such as mean, median, mode, PMM, and hot-deck in MCAR, MNAR, and MAR 10 to 30% missing data for predicting radon concentration (RN) from other environmental attributes keeping setting 2 are presented in Table 3. The IBFI based imputed dataset performs best among others for training and results in a more accurate prediction of radon concentration (RN) from the fact that its RMSE is very much less when compared to mean, median, mode, PMM, and hot-deck imputation datasets. In MCAR 10 to 30%, the minimum and maximum RMSE values of 1141.1 and 1166.3 respectively for IBFI imputation, which is less when compared to other imputation methods such as hot-deck imputation with the maximum RMSE value of 1454.1. A similar pattern is observed in MNAR and MAR 10 to 30% datasets having the least RMSE value compared to other imputed datasets. As far as the other imputed datasets are of concern, in setting 2, median and PMM based imputed datasets performs closer to IBFI based imputed dataset. In MCAR 20, 30, and MAR 10% based imputed datasets, PMM performs closer to the IBFI imputed dataset with the difference of RMSE from IBFI of 47, 65.7, and 17.3 respectively. On the other hand, the median imputed dataset performs closer to IBFI with the difference ranging from 65.5 to 120.4 for MCAR 20,30%, MNAR 10 to 30%, and MAR 20,30%. For the statistics discussed above for setting 1 and setting 2, it is concluded that IBFI based imputed dataset performs better than other imputed datasets. In setting 1, PMM imputed dataset performs better than other imputed datasets apart from IBFI imputed dataset. In setting 2, PMM and median-based imputed dataset perform very closer to IBFI imputed dataset when predicting radon concentration keeping other attributes as predictor attributes such as thoron, temperature, relative humidity, and pressure. Figure 6 (a,b) show the results when the MSE statistic across the variable radon concentration (RN) is normalized to the average for MCAR, MNAR, and MAR 10 to 30 percent for setting 1 and 2. To better interpret the results from the analysis, MSE statistics are decimally scaled. It can be observed in Figures 6a and 6b regarding setting1 and setting 2, IBFI imputed dataset is superior for all the cases of MCAR, MNAR, and MAR 10 to 30% of missingness. For IBFI based imputed datasets in setting 1 and 2, the MSE value ranged from 0.291 to 0.308 and 0.132 to 0.137 respectively, which is very less (decimal scaled) when compared to other imputed datasets for the prediction of radon concentration (RN). For setting 1, PMM performs very closer to IBFI in all degrees of missingness with very little difference of MSE value when

TABLE 3. The RMSE values for IBFI and other imputation methods (Mean, Median, Mode, PMM, and Hotdeck) for missingness of 10, 20, and 30% of type MCAR, MNAR, and MAR imputed datasets for predicting radon concentration keeping Setting 2.

	MCAR 10%	MCAR 20%	MCAR 30%	MNAR 10%	MNAR 20%	MNAR 30%	MAR 10%	MAR 20%	MAR 30%
IBFI	1151.2	1166.3	1141.1	1144.3	1188.6	1170.5	1148.8	1158.7	1187.1
Mean	1219.6	1247.8	1255.1	1216.8	1296.4	1318.6	1204.4	1273.1	1284.7
Median	1216.7	1240.8	1244.8	1211.9	1272.8	1290.9	1203.1	1257.7	1260.3
Mode	1343.2	1313.2	1384.9	1283.4	1409.6	1368.6	1375	1414.7	1376.6
PMM	1239	1213.3	1206.8	1215.5	1315.2	1331	1166.1	1282.7	1379.7
Hot-deck	1380.5	1401.7	1454.1	1375.6	1555.3	1545.9	1409.4	1526.7	1615.6

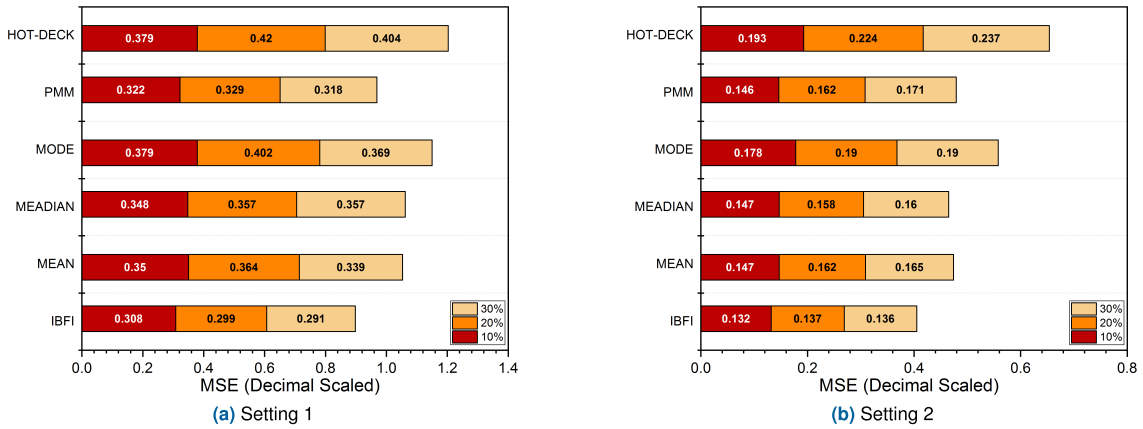

FIGURE 6. IBFI compared with other methods (mean, median, mode, PMM, Hotdeck) normalized to the average statistic for all methods showing average MSE of 10%, 20%, and 30% for MCAR, MNAR, and MAR missingness keeping a) setting 1 and b) setting 2.

compared to IBFI imputed dataset of 0.014, 0.03, and 0.027. Moreover, in setting 2, median and pmm based imputed datasets both perform closer to IBFI imputed datasets such as for 10% of missingness in the average of MCAR, MNAR, and MAR, there is a 10.61% increase in MSE value from IBFI while in 20 and 30% of missingness, median performs closer with the percentage increase of MSE value of 15.33% and 17.65% respectively. A similar pattern was observed in Tables 2 and 3 from which it is concluded that PMM and median imputed dataset performs closer to IBFI imputed dataset. Moreover, in all types and degrees of missingness, IBFI imputed dataset performs better than other imputed datasets by mean, median, mode, pmm, and hot-deck. Similar performance statistics are observed in Figures 7 and 8. In figure 7, the root mean squared log error is presented which is calculated on average for setting 1 and setting 2 in MCAR (black bubble), MNAR (blue bubble), and MAR (red bubble) across the degree of 10 to 30% while figure 8 presents the average MAPE for setting 1 and setting 2 in MCAR, MNAR and MAR scenarios across the same degree of missingness. From Figure 7, it is further concluded that IBFI imputed dataset provides the best fit for the prediction of radon concentration (RN) with lower RMSLE for all the types and degrees of missingness. However, PMM and median imputed dataset performs closer to each other. In Figure 8, the performance of the fitted model for the prediction of radon concentration using different imputed datasets of IBFI,

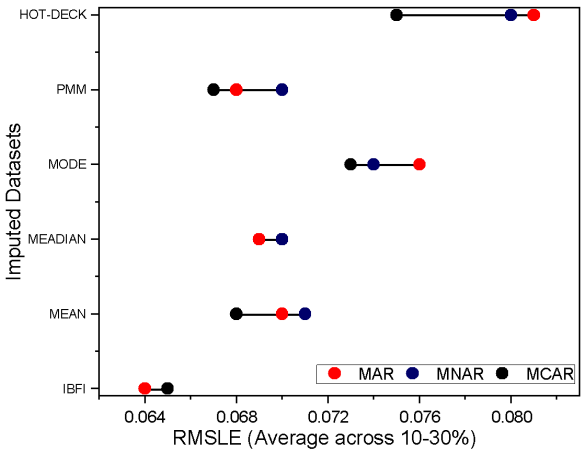

FIGURE 7. IBFI compared with other methods (mean, median, mode, PMM, Hotdeck) normalized to the average statistic for all methods showing average RMSLE of 10%, 20%, and 30% for MCAR, MNAR, and MAR missingness across setting 1 and 2.

mean, median, mode, PMM, and hot-deck is measured in terms of MAPE which is the average value of all degrees of missingness across setting 1 and setting 2. The IBFI imputed dataset results in better prediction accuracy for the prediction of radon concentration (RN) with the least MAPE value of 0.050 when compared to model fitting on other imputed datasets. In the MCAR scenario, mean, median, and PMM

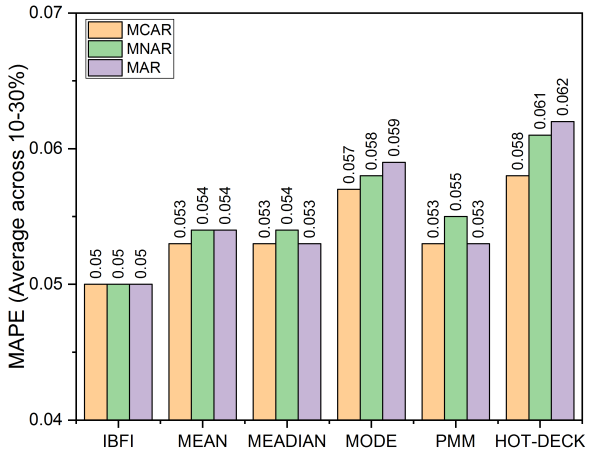

FIGURE 8. IBFI compared with other methods (mean, median, mode, PMM, Hotdeck) normalized to the average statistic for all methods showing average MAPE of 10%, 20%, and 30% for MCAR, MNAR, and MAR missingness across setting 1 and 2.

performs equivalent to each other while PMM and median perform equals to each other in the MAR scenario with the MAPE of 0.053.

A. COMPARISON WITH EXISTING LITERATURE

In this section, we have compared the simulation plan of this study with other recent research work regarding soil radon gas concentration data. The comparison is done by comparing and contrasting the proposed methodology for data preprocessing, data splitting for training and testing purposes, and performance evaluation metrics. For the accurate prediction of soil radon gas concentration data, a methodology named delegated regressor was proposed based on a delegation framework [25]. Before training the models, the original soil radon gas time series data was partitioned into two subsets called seismic and non-seismic datasets. These partitions were made by incorporating time windows. After training the models using non-seismic data, the soil radon gas concentration data from seismic data was predicted. The testing results reveals that the delegated regressor methodology achieves the least RMSE score when compared to other prediction models. Furthermore, a methodology was proposed by Mir *et al.* [27]classifies soil radon gas time series data into seismic and non-seismic by employing stacking for classification and an automatic anomaly indication tool as a post-processing method. The predictions from first-level learners along with class labels in the stacking framework were further passed to the meta-classifier for training. For test data, the classifications made by the second level learner were passed to the automatic anomaly indication function to classify the series into seismically active or in-active. The automatic anomaly indication function calculates the percentage of indications for anomaly and classifies the coming series into seismic when this indication percentage gets equal to or higher than the threshold. In another study by Tareen *et al.* [28], boxplots are employed to detect specific patterns in the soil radon gas concentration time series data. These patterns were observed in the time series because of different geological activities before the occurrence of earthquakes. Tareen *et al.* [26] experimented with different computational intelligent techniques for analyzing anomalous behavior in soil radon gas. This study concludes that the anomaly in soil radon gas is mainly caused the noise and seismic activity. In comparison to recent studies, this one focuses primarily on the filling of missing patterns in soil radon gas concentration time series data. The main objective of this paper is to experiment with a new methodology, imputation by feature importance (IBFI), for serving its imputed dataset in further experimentation of soil radon gas concentration dataset. This paper concludes that IBFI based imputed datasets could be better served for further regression scenarios.

V. CONCLUSION

Missing patterns in the real-time series data often occur due to several possible reasons as discussed above in different sections. Because missing values in the data can produce bias in the forecasting model, imputations of these missing values in the data are critical for further analysis. In this article i.e. imputation by feature importance (IBFI) has been used against other imputation methods for serving its imputed dataset for further prediction and forecasting scenarios. To analyze the performance of these imputation methods, this work has utilized the imputed datasets by IBFI and other imputation methods. The imputation was done by first introducing missing patterns in the data at different missingness scenarios such as missing completely at random (MCAR), missing not at random (MNAR), and missing at random (MAR) across the missingness percentage of 10 to 30%. The missing data is reconstructed and it was concluded that imputation by feature importance (IBFI) efficiently imputed the missing patterns in soil radon gas concentration (SRGC) time-series data in all types and degrees of missingness. Furthermore, the imputed datasets from IBFI and other imputation methods are used to forecast the radon concentration (RN) from other environmental attributes present in the imputed dataset. These imputed datasets are 9 for each imputation (3 for each missingness type) method with a total sum of 54. The experimentation is carried out in two different settings such as setting 1 and 2 which is the effort to incorporate the different seismic activities for the fitted model evaluation. Findings of the study show that IBFI imputed dataset results in a better-fitted machine learning model and predicts the radon concentration of the test set with less error when compared to the fitted model with other imputed datasets of mean, median, mode, PMM, and hotdeck. Moreover, PMM imputed dataset performs closer to IBFI in setting 1 while median and PMM performs very closer to IBFI imputed dataset in setting 2. The performance of the IBFI imputed dataset is based upon the ability of IBFI to choose the best predictor variable for different response variables for the better and unbiased reconstruction of missing patterns.

ACKNOWLEDGMENT

Taif University Researchers Supporting Program (project number: TURSP-2020/195), Taif University, Saudi Arabia, supported this research. The authors are grateful to King Khalid University's Deanship of Scientific Research for financing this research under grant number (RGP 1/14/43).The authors would like to express their gratitude to Prince Sultan University for paying the publication's Article Processing Charges (APC). The data used in the current study is a part of the research conducted for the project grant no: 6453/ AJK/NRPU/R&D/HEC/2016 against the NRPU project executed by one of the co-authors, MR.

REFERENCES

[1] A. H. Alomari, M. A. Saleh, S. Hashim, A. Alsayaheen, and I. Abdeldin, "Activity concentrations of ^{226}Ra, ^{228}Ra, ^{222}Rn and their health impact in the groundwater of Jordan," *J. Radioanal. Nucl. Chem.*, vol. 322, no. 2, pp. 305–318, 2019.

[2] K. J. Kearfott, R. L. Metzger, K. R. Kraft, and K. E. Holbert, "Mitigation of elevated indoor radon gas resulting from underground air return usage," *Health Phys.*, vol. 63, no. 6, pp. 674–680, Dec. 1992.

[3] K. J. Kearfott, "Preliminary experiences with 222Rn gas in Arizona homes," *Health Phys.*, vol. 56, no. 2, pp. 169–179, Feb. 1989.

[4] M. Rafique, S. Qayyum, S. U. Rahman, and M. Matiullh, "The influence of geology on indoor radon concentrations in Neelum valley Azad Kashmir, Pakistan," *Indoor Built Environ.*, vol. 21, no. 5, pp. 718–726, Oct. 2012.

[5] D. Xie, M. Liao, H. Wang, and K. J. Kearfott, "A study of diurnal and short-term variations of indoor radon concentrations at the University of Michigan, USA and their correlations with environmental factors," *Indoor Built Environ.*, vol. 26, no. 8, pp. 1051–1061, Oct. 2017.

[6] D. Banks, B. Frengstad, A. K. Midtgård, J. R. Krog, and T. Strand, "The chemistry of Norwegian groundwaters: I. The distribution of radon, major and minor elements in 1604 crystalline bedrock groundwaters," *Sci. Total Environ.*, vol. 222, nos. 1–2, pp. 71–91, Oct. 1998.

[7] E. Levintal, M. I. Dragila, H. Zafrir, and N. Weisbrod, "The role of atmospheric conditions in CO_2 and radon emissions from an abandoned water well," *Sci. Total Environ.*, vol. 722, Jun. 2020, Art. no. 137857.

[8] M. T. Olguin, N. Segovia, E. Tamez, M. Alcántara, and S. Bulbulian, "Radon concentration levels in ground water from Toluca, Mexico," *Sci. Total Environ.*, vols. 130–131, pp. 43–50, Mar. 1993.

[9] F. Perrier, P. Richon, and J.-C. Sabroux, "Temporal variations of radon concentration in the saturated soil of Alpine grassland: The role of groundwater flow," *Sci. Total Environ.*, vol. 407, no. 7, pp. 2361–2371, Mar. 2009.

[10] M. A. Shenber, "Radon (^{222}Rn) short-lived decay products and their temporal variation in surface air in Tripoli," *Sci. Total Environ.*, vol. 119, pp. 243–251, Jun. 1992.

[11] A. V. Sundal, V. Valen, O. Soldal, and T. Strand, "The influence of meteorological parameters on soil radon levels in permeable glacial sediments," *Sci. Total Environ.*, vol. 389, nos. 2–3, pp. 418–428, Jan. 2008.

[12] D. Xie, H. Wang, and K. J. Kearfott, "Modeling and experimental validation of the dispersion of ^{222}Rn released from a uranium mine ventilation shaft," *Atmos. Environ.*, vol. 60, pp. 453–459, Dec. 2012.

[13] M. Yitshak-Sade, A. J. Blomberg, A. Zanobetti, J. D. Schwartz, B. A. Coull, I. Kloog, F. Dominici, and P. Koutrakis, "County-level radon exposure and all-cause mortality risk among medicare beneficiaries," *Environ. Int.*, vol. 130, Sep. 2019, Art. no. 104865.

[14] F. Ambrosino, L. Thinová, M. Briestenský, and C. Sabbarese, "Analysis of radon time series recorded in Slovak and Czech caves for the detection of anomalies due to seismic phenomena," *Radiat. Protection Dosimetry*, vol. 186, nos. 2–3, pp. 428–432, Dec. 2019.

[15] B. R. Arora, A. Kumar, V. Walia, T. F. Yang, C.-C. Fu, T.-K. Liu, K.-L. Wen, and C.-H. Chen, "Assesment of the response of the meteorological/hydrological parameters on the soil gas radon emission at Hsinchu, northern Taiwan: A prerequisite to identify earthquake precursors," *J. Asian Earth Sci.*, vol. 149, pp. 49–63, Nov. 2017.

[16] A. Barkat, A. Ali, N. Siddique, A. Alam, M. Wasim, and T. Iqbal, "Radon as an earthquake precursor in and around northern Pakistan: A case study," *Geochem. J.*, vol. 51, no. 4, pp. 337–346, 2017.

[17] V. M. Choubey, N. Kumar, and B. R. Arora, "Precursory signatures in the radon and geohydrological borehole data for M4.9 Kharsali earthquake of Garhwal Himalaya," *Sci. Total Environ.*, vol. 407, no. 22, pp. 5877–5883, Nov. 2009.

[18] R. G. M. Crockett, G. K. Gillmore, P. S. Phillips, A. R. Denman, and C. J. Groves-Kirkby, "Radon anomalies preceding earthquakes which occurred in the UK in summer and autumn 2002," *Sci. Total Environ.*, vol. 364, nos. 1–3, pp. 138–148, Jul. 2006.

[19] D. Ghosh, A. Deb, and R. Sengupta, "Anomalous radon emission as precursor of earthquake," *J. Appl. Geophys.*, vol. 69, no. 2, pp. 67–81, Oct. 2009.

[20] E. Hauksson, "Radon content of groundwater as an earthquake precursor: Evaluation of worldwide data and physical basis," *J. Geophys. Res., Solid Earth*, vol. 86, no. B10, pp. 9397–9410, Oct. 1981.

[21] G. Igarashi, S. Saeki, N. Takahata, K. Sumikawa, S. Tasaka, Y. Sasaki, M. Takahashi, and Y. Sano, "Ground-water radon anomaly before the Kobe earthquake in Japan," *Science*, vol. 269, no. 5220, pp. 60–61, Jul. 1995.

[22] K. Kawabata, T. Sato, H. A. Takahashi, F. Tsunomori, T. Hosono, M. Takahashi, and Y. Kitamura, "Changes in groundwater radon concentrations caused by the 2016 Kumamoto earthquake," *J. Hydrol.*, vol. 584, May 2020, Art. no. 124712.

[23] H. S. Virk and B. Singh, "Radon recording of Uttarkashi earthquake," *Geophys. Res. Lett.*, vol. 21, no. 8, pp. 737–740, Apr. 1994.

[24] H. Woith, "Radon earthquake precursor: A short review," *Eur. Phys. J. Special Topics*, vol. 224, no. 4, pp. 611–627, May 2015.

[25] M. Rafique, A. D. K. Tareen, A. A. Mir, M. S. A. Nadeem, K. M. Asim, and K. J. Kearfott, "Delegated regressor, a robust approach for automated anomaly detection in the soil radon time series data," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, Dec. 2020.

[26] A. D. K. Tareen, K. M. Asim, K. J. Kearfott, M. Rafique, M. S. A. Nadeem, T. Iqbal, and S. U. Rahman, "Automated anomalous behaviour detection in soil radon gas prior to earthquakes using computational intelligence techniques," *J. Environ. Radioactivity*, vol. 203, pp. 48–54, Jul. 2019.

[27] A. A. Mir, F. V. Çelebi, M. Rafique, M. R. I. Faruque, M. U. Khandaker, K. J. Kearfott, and P. Ahmad, "Anomaly classification for earthquake prediction in radon time series data using stacking and automatic anomaly indication function," *Pure Appl. Geophys.*, vol. 178, pp. 1593–1607, May 2021.

[28] A. D. K. Tareen, M. S. A. Nadeem, K. J. Kearfott, K. Abbas, M. A. Khawaja, and M. Rafique, "Descriptive analysis and earthquake prediction using boxplot interpretation of soil radon time series data," *Appl. Radiat. Isot.*, vol. 154, Dec. 2019, Art. no. 108861.

[29] M. Rafique, J. Iqbal, K. J. Lone, K. J. Kearfott, S. U. Rahman, and L. Hussain, "Multifractal detrended fluctuation analysis of soil radon (^{222}Rn) and thoron (^{220}Rn) time series," *J. Radioanal. Nucl. Chem.*, vol. 328, no. 1, pp. 425–434, Apr. 2021.

[30] V. I. Ulomov and B. Mavashev, "A precursor of a strong tectonic earthquake," *Doklady Akademii Nauk*, vol. 176, no. 2, pp. 319–321, 1967.

[31] M. E. Kitto, "Interrelationship of indoor radon concentrations, soil-gas flux, and meteorological parameters," *J. Radioanal. Nucl. Chem.*, vol. 264, no. 2, pp. 381–385, May 2005.

[32] R. C. Ramola, Y. Prasad, G. Prasad, S. Kumar, and V. M. Choubey, "Soil-gas radon as seismotectonic indicator in Garhwal Himalaya," *Appl. Radiat. Isot.*, vol. 66, no. 10, pp. 1523–1530, Oct. 2008.

[33] M. Singh, R. Ramola, S. Singh, and H. Virk, "The influence of meteorological parameters on soil gas radon," *J. Assoc. Explor. Geophys.*, vol. 9, no. 2, pp. 85–90, 1988.

[34] J. Iqbal, K. J. Lone, L. Hussain, and M. Rafique, "Detrended cross correlation analysis (DCCA) of radon, thoron, temperature and pressure time series data," *Phys. Scripta*, vol. 95, no. 8, Jul. 2020, Art. no. 085213.

[35] N. A. Zakaria and N. M. Noor, "Imputation methods for filling missing data in urban air pollution data formalaysia," *Urbanism Arhitectura Constructii*, vol. 9, no. 2, p. 159, 2018.

[36] R. J. Little, "Regression with missing *X*'s: A review," *J. Amer. Stat. Assoc.*, vol. 87, no. 420, pp. 1227–1237, 1992.

[37] S. Greenland and W. D. Finkle, "A critical look at methods for handling missing covariates in epidemiologic regression analyses," *Amer. J. Epidemiol.*, vol. 142, no. 12, pp. 1255–1264, Dec. 1995.

[38] Y. C. Yuan, "Multiple imputation for missing data: Concepts and new development," in *Proc. 25th Annu. SAS Users Group Int. Conf.*, 2000, vol. 267, no. 11, pp. 1–11.

[39] M. Fichman and J. N. Cummings, "Multiple imputation for missing data: Making the most of what you know," *Org. Res. Methods*, vol. 6, no. 3, pp. 282–308, Jul. 2003.
[40] P. Li, E. A. Stuart, and D. B. Allison, "Multiple imputation: A flexible tool for handling missing data," *J. Amer. Med. Assoc.*, vol. 314, no. 18, pp. 1966–1967, 2015.
[41] G. L. Schlomer, S. Bauman, and N. A. Card, "Best practices for missing data management in counseling psychology," *J. Counseling Psychol.*, vol. 57, no. 1, p. 1, 2010.
[42] A. A. Mir, K. J. Kearfott, F. V. Çelebi, and M. Rafique, "Imputation by feature importance (IBFI): A methodology to envelop machine learning method for imputing missing patterns in time series data," *PLoS ONE*, vol. 17, no. 1, Jan. 2022, Art. no. e0262131.
[43] B. Sv and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *J. Stat. Softw.*, vol. 45, no. 3, pp. 1–67, 2011.
[44] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *J. Amer. Stat. Assoc.*, vol. 49, no. 268, pp. 765–769, 1954.
[45] S. A. Pulinets, V. A. Alekseev, A. D. Legen'ka, and V. V. Khegai, "Radon and metallic aerosols emanation before strong earthquakes and their role in atmosphere and ionosphere modification," *Adv. Space Res.*, vol. 20, no. 11, pp. 2173–2176, Jan. 1997.
[46] S. Pulinets and K. Boyarchuk, *Ionospheric Precursors of Earthquakes*. Springer, Aug. 2004.
[47] H. Virk, A. K. Sharma, and V. Walia, "Correlation of alpha-logger radon data with microseismicity in NW Himalaya," *Current Sci.*, vol. 72, no. 9, pp. 656–663, 1997.
[48] R. Ramola, M. Singh, A. Sandhu, S. Singh, and H. Virk, "The use of radon as an earthquake precursor," *Int. J. Radiat. Appl. Instrum. E.*, vol. 4, no. 2, pp. 275–287, 1990.
[49] H. A. Afan, A. El-Shafie, Z. M. Yaseen, M. M. Hameed, W. H. M. W. Mohtar, and A. Hussain, "ANN based sediment prediction model utilizing different input scenarios," *Water Resour. Manage.*, vol. 29, no. 4, pp. 1231–1245, Mar. 2015.
[50] S. Saeed, L. Hussain, I. A. Awan, and A. Idris, "Comparative analysis of different statistical methods for prediction of $PM_{2.5}$ and PM_{10} concentrations in advance for several hours," *Int. J. Comput. Sci. Netw. Secur.*, vol. 17, no. 11, pp. 45–52, 2017.
[51] R. F. Byrne, "Beyond traditional time-series: Using demand sensing to improve forecasts in volatile times," *J. Bus. Forecasting*, vol. 31, no. 2, pp. 13–19, 2012.
[52] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int. J. Forecasting*, vol. 32, no. 3, pp. 669–679, 2016.

ADIL ASLAM MIR received the B.S. and M.Phil. degrees in computer sciences from The University of Azad Jammu and Kashmir, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with Ankara Yıldırım Beyazıt University, Turkey. He is employed as a Research Associate with the Department of Computer Science and Information Technology, The University of Azad Jammu and Kashmir. His research interests include machine learning-based decision support systems (DSSs), data mining, and artificial intelligence.

FATIH VEHBI ÇELEBI received the B.Sc. degree in electrical-electronics engineering from Middle East Technical University, in 1988, the M.Sc. degree in electrical-electronics engineering from Gaziantep University, in 1996, and the Ph.D. degree in electrical-electronics engineering from Erciyes University, in 2002. He is currently a full-time Professor and the Dean of the Faculty of Engineering and Applied Sciences, Ankara Yıldırım Beyazıt University (AYBU). He has published so many scientific papers and his current research interests include cyber security, artificial intelligence, machine learning, and optoelectronics.

MUHAMMAD RAFIQUE received the Ph.D. degree in computational physics from the Department of Physics and Applied Mathematics, Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan, in 2005. He worked as a Visiting Faculty Member and a Postdoctoral Fellow at the Nuclear Engineering and Radiological Science Department, University of Michigan, Ann Arbor, USA, in 2014. He is currently working as a Professor of physics at the Department of Physics, The University of Azad Jammu and Kashmir, Muzaffarabad. He has also served as the Head of the Department of Physics, the Director for Quality Enhancement, and the Director for ORIC. As an additional charge, he is serving as the Director for advanced studies at The University of Azad Jammu and Kashmir. He has published more than 100 research articles in international and national journals of repute. He has also produced more than 35 M.Phil./M.S. and five Ph.D. students as a Supervisor and a Co-Supervisor. His research interests include reactor physics, radiation physics, computational physics and mathematics, geophysics, and medical physics.

LAL HUSSAIN received the M.S. degree (Hons.) in communication and networks from Iqra University, Islamabad, Pakistan, in 2012, and the Ph.D. degree from the Department of Computer Science and Information Technology, The University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan, in February 2016. He worked as a Visiting Ph.D. Researcher at Lancaster University, U.K., for six months, under the HEC International Research Initiative Program and worked under the supervision of Dr. Aneta Stefanovska, a Professor of biomedical physics at the Physics Department, Lancaster University, from 2014 to 2015. He recently completed one-year postdoctoral fellowship at the Montefiore Medical Center and the Albert Einstein College of Medicine, New York, USA, under the supervision of Dr. Tim Q. Duong, a Professor and the Vice Chair of MRI Research. He also worked at the Duong Laboratory, Stony Brook University, USA, in different on-going projects with Dr. Duong, from January 2020 to March 2020. He is currently an Assistant Professor at the Department of Computer Science and IT, The University of Azad Jammu and Kashmir. Recently, he is ranked in 1% top world scientists list of 2021 by Elsevier based on research record. He is the author of more than 50 publications of highly reputed peer-reviewed and impact fact journals as the principal author. He completed various funded projects as a PI and a Co-PI from Ignite, ICT Pakistan, and the University of Jeddah and Saudi Electronic University, Saudi Arabia. He has presented various talks at Pakistan, U.K., Peru, and USA. His research interests include developing and optimizing AI tools, including machine learning, deep learning and neural networks algorithms, feature extraction and selection methods, information-theoretic methods, time-frequency representation methods, and cross frequency coupling to predict the disease severity, progression, survival, and recurrence. His area of interests include biomedical signal and image processing problems, including prostate cancer, breast cancer, lung cancer, brain tumor, covid-19 lung infection with different modalities (i.e., MRI, CT, and X-Ray), and brain dynamics and diseases (i.e., autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), and Alzheimer's Disease). He received the Gold Medal for his M.S. degree from Iqra University.

AHMED S. ALMASOUD received the highest degree from the University of Technology at Sydney. He has been working with Prince Sultan University (PSU), Riyadh, Saudi Arabia, since 2014, where he is currently an Assistant Professor with the College of Computer and Information Sciences. He has published original articles in the finest journals in the area of his studies. His research interests include (but not limited to) artificial intelligence, machine learning, security architecture, and the Internet of Things.

MASOUD ALAJMI (Member, IEEE) received the B.S. degree in electrical engineering from the King Fahd University of Petroleum and Minerals (KFUPM), in 2004, and the M.S. degree in electrical engineering and the Ph.D. degree in electrical and computer engineering from Western Michigan University, Kalamazoo, MI, USA, in 2010 and 2016, respectively. He has over four years of experience in industry. For three months, he worked at Zamel and Turbag Consulting Engineers, Al-Khobar, Saudi Arabia, as an Electrical Engineer. After that, he joined Saudi Electricity Company (SEC), Abha, Saudi Arabia, where he worked as a Pre-Commissioning Engineer, from 2004 to 2008. During that period, he completed many training programs in the technical and administrative fields at well-known institutes. He was also assigned to be the Commissioning Leader for many projects in Saudi Arabia. He was assigned to be the SEC representative to supervise factory acceptance tests for Siemens Company, Berlin, Germany, in 2007; and in 2008, for Hyundai Heavy Industries Company Ltd., Ulsan, South Korea. From 2012 to 2015, he also worked as a Teaching Assistant with the Electrical and Computer Engineering Department, Western Michigan University, where he received the 2014–2015 Graduate Teaching Effectiveness Award for excellent teaching skills. He is currently an Associate Professor with the Computer Engineering Department, Taif University, Taif, Saudi Arabia. He is involved in various technical committees and is also a coauthor of about 30 articles in international journals and conference proceedings. His research interests include signal processing, biomedical image processing, image encryption, watermarking, steganography, data hiding, machine learning, and smart grids and renewable energy.

FAHD N. AL-WESABI received the B.S. degree in computer science from the University of Science and Technology, Sana'a, Yemen, in 2006, the M.S. degree in computer information systems from the Arabic Academy for Banking and Financial Sciences, Sana'a Branch, Yemen, in 2009, and the Ph.D. degree in computer science from SRTM University, India, in 2015. From 2006 to 2009, he was a Research Assistant and from 2010 to 2015, he was a Lecturer with the Faculty of Engineering, University of Science and Technology. From 2015 to 2018, he was an Assistant Professor with the Faculty of Computer and Information Technology, Sana'a University, Yemen. Since October 2018, he has been an Assistant Professor with the Computer Science Department, King Khalid University, Saudi Arabia. He is the author of ten books, more than 80 articles, and many funded research projects. His research interests include AI, the IoT, smart cities, machine learning, biomedical, software engineering, applied soft computing, information security, and enterprise systems.

ANWER MUSTAFA HILAL received the Ph.D. degree from Omdurman Islamic University, Khartoum, Sudan, in 2017, for his thesis titled A Semantic Data Mining Model for Exploring the Holy Quran. He is currently an Assistant Professor of computer science with the Department of Computer and Self Development, Prince Sattam Bin Abdulaziz University. His research interests include data mining, text mining and mobile, and web development.