

Towards Improving Predictive Statistical Learning Model Accuracy by Enhancing Learning Technique

Ali Algarni¹, Mahmoud Ragab^{2,3,4,*}, Wardah Alamri⁵ and Samih M. Mostafa⁶

¹Statistics Department, Faculty of Science, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

²Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

³Mathematics Department, Faculty of Science, Al-Azhar University, Naser City, 11884, Egypt

⁴Centre of Artificial Intelligence for Precision Medicines, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

⁵Educational Technology Department, Educational Graduate Studies Faculty, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

⁶Computer Science Department, Faculty of Computers and Information, South Valley University, Qena, 83523, Egypt

*Corresponding Author: Mahmoud Ragab. Email: mragab@kau.edu.sa

Received: 29 July 2021; Accepted: 30 August 2021

Abstract: The accuracy of the statistical learning model depends on the learning technique used which in turn depends on the dataset's values. In most research studies, the existence of missing values (MVs) is a vital problem. In addition, any dataset with MVs cannot be used for further analysis or with any data driven tool especially when the percentage of MVs are high. In this paper, the authors propose a novel algorithm for dealing with MVs depending on the feature selection (FS) of similarity classifier with fuzzy entropy measure. The proposed algorithm imputes MVs in cumulative order. The candidate feature to be manipulated is selected using similarity classifier with Parkash's fuzzy entropy measure. The predictive model to predict MVs within the candidate feature is the Bayesian Ridge Regression (BRR) technique. Furthermore, any imputed features will be incorporated within the BRR equation to impute the MVs in the next chosen incomplete feature. The proposed algorithm was compared against some practical state-of-the-art imputation methods by conducting an experiment on four medical datasets which were gathered from several databases repository with MVs generated from the three missingness mechanisms. The evaluation metrics of mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2 score) were used to measure the performance. The results exhibited that performance vary depending on the size of the dataset, amount of MVs and the missingness mechanism type. Moreover, compared to other methods, the results showed that the proposed method gives better accuracy and less error in most cases.

Keywords: Bayesian ridge regression; fuzzy entropy measure; feature selection; imputation; missing values; missingness mechanisms; similarity classifier; medical dataset



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

MVs are considered a critical problem that can occur in many scientific areas such as biological, psychological, or medical [1]. Commonly, many reasons may lead to the occurrence of MVs, for instance, wrong data entry, improper data collection, management of similar but not identical datasets and malfunctioning measurement equipment [2]. Machine learning (ML), big data and any data driven tool require high data quality which results in good analysis and outcomes. The existence of MVs within a dataset can result in problems, for instance, bad data analysis, reducing the research results obtained from such dataset and presenting amount of bias [3]. To this end, significant information is incorporated within MVs which should be manipulated before using the incomplete dataset with any data driven tool. Furthermore, many researches were done and novel algorithms were proposed to solve the problem of MVs, especially in medical data [4]. Nevertheless, several imputation algorithms may result in poor imputation and may fail in handling all MVs in the dataset. In addition, they may not deal with all missingness mechanisms. These shortcomings of these algorithms encouraged the authors to propose a novel algorithm introduced in this paper. The proposed algorithm utilizes the most significant feature to impute MVs in cumulative order. Besides MVs, FS also affects the ML model performance.

1.1 Feature Selection

High dimensionality data is problematic especially in circumstances when a dataset contains a few numbers of training instances and a large number of features. This type of data commonly exists in medicine where cost and time problems may limit the number of training observations, while the number of diseases increases through the years [5]. FS helps to overcome the problem of high dimensionality by selecting a subset of features that have a strong relationship with the target feature. In addition, in the existence of MVs the FS is considered a vital preprocessing step such as correlation, mutual information and fuzzy FS. Dropping features that hold a large number of MVs (*e.g.*, >50%) is an easy solution. But such a solution may result in bad analysis, losing the ability to recognize statistically significant variations and may also generates bias. Missingness mechanisms have a large effect on FS that's why before applying any FS technique missingness mechanisms need to be taken into consideration [3].

1.2 Missingness Mechanisms

Before introducing different methods for handling MVs, it is essential to present the different types of missingness mechanisms (*i.e.*, the reason for the occurrence of MVs in data). MVs are commonly classified to one of three MVs mechanisms [6]:

- *Missing Completely at Random (MCAR)*: This type of MVs mechanisms happens when the probability of the existence of MVs is independent from any other features in the data. From statistical perspective, MCAR can be stated as in Eq. (1) [1].

$$f(M|Y, \emptyset) = f(M|Y) \text{ for all } Y, \emptyset \quad (1)$$

where M and Y represent the missing and observed data respectively. The conditional probability is denoted by f and \emptyset represents the unknown parameter. An example of MCAR MVs occur when the measuring equipment stops working correctly [7].

- *Missing at Random (MAR)*: In this mechanism the relationship between MVs and other features existed in the dataset is a dependent relationship. In other words, the probability of the occurrence of MVs depends on observed values in other features and not on other MVs in the target feature [8]. MAR can be represented as in Eq. (2) [1].

$$f(M|Y, \emptyset) = f(M|Y_{obs}, \emptyset) \text{ for all } Y_{mis}, \emptyset \quad (2)$$

where Y_{mis} and Y_{obs} are the missing and observed values from Y respectively. From a medical perspective, this situation may occur when an experiment has not been accomplished because a feature within the dataset shows that the patient is a woman for example [7].

- *Missing Not at Random (MNAR)*: For this type, there is a dependent relationship between the MVs and the observed data. MNAR can be expressed using Eq. (3) [1].

$$f(M, Y|\theta, \emptyset) = f(Y|\theta)f(M|Y, \emptyset) \quad (3)$$

where θ (*i.e.*, parameter of the distribution Y). is estimated from the detected data The distribution of the missingness is denoted by \emptyset . An example of MNAR MVs, when people having a too high or too low income reject to reveal it [7].

1.3 Handling Missing Data

The simplest methods for handling MVs are traditional methods. Traditional methods can be deletion (*i.e.*, delete instances that hold MVs), mean or median (*i.e.*, replace the MVs with the mean or median of the feature that holds MVs) substitution [9]. Deletion can be case deletion or pairwise deletion. Case deletion (a.k.a., listwise) in which any instance holds MVs is dropped from analysis. In many statistical packages listwise is the default choice [10]. Pairwise deletion is considered a selective method, which tries to minimize the lost amount of data instances that occur in case of using the listwise method by including into the analysis the instances with *MVs*. In other words, pairwise deletion will drop only particular features with MVs from the analysis and use the remainder features with no *MVs*. The selection of features varies from analysis to another depending on the missingness. Using deletion methods results in reducing the data size [11].

The other methods that overcome the defects of deletion methods are called imputation methods. In imputation methods, predefined (mean, median, etc.) or estimated (using statistical methods, ML algorithms, etc.) value is used instead of MVs [12]. Imputation is classified into single and multiple imputation. In single imputation, MVs are imputed by a value one time. Though, single imputation does not require computational resources it can result in biased results [3]. In multiple imputation, m copies from the original dataset are generated. In each generated dataset, MVs are imputed using single imputation techniques. The final imputed dataset is the average analysis of the m imputed datasets [13,14]. ML algorithms can also be used to predict MVs depending on using the available information within the given dataset. Some examples of ML techniques that are used to predict MVs include linear regression, k-nearest neighbour (KNN), decision trees [1] and BRR. BRR is the predictive model used within the proposed algorithm in this paper to predict MVs, which can be expressed using Eq. (4) [15].

$$y \sim N(\mu, \alpha) \quad (4)$$

where:

$$\mu = \beta X = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_q x_q$$

$$\beta \sim N(0, \lambda^{-1} I_q)$$

$$\alpha \sim g(\alpha_1, \alpha_2)$$

$$\lambda \sim g(\lambda_1, \lambda_2)$$

The target feature is denoted by y which is distributed as a normal distribution characterized by mean $\mu = \beta X$ and variance α . $\beta = \{\beta_0, \beta_1, \beta_2, \dots, \beta_q\}$ denotes the unknown parameters and $X = \{x_1, x_2, \dots, x_q\}$

denotes the independent features. The number of independent features is represented by q . α and λ represent the regularization parameters which are assessed jointly while fitting the model through maximizing the log marginal likelihood and both of them are assumed to be distributed as gamma distribution. $\alpha_1, \alpha_2, \lambda_1, \lambda_2$ are hyper-parameters of the gamma prior distributions.

The rest of the paper is organized as follows: Section 2 presents a brief literature review about analysis of MVs . Sections 3 and 4 reveal the proposed algorithm and explains in detail the experimental setup, respectively. Section 5 is devoted to the presentation of the results and discussion while section 6 concludes this paper and exhibits some perspectives of future work.

2 Literature Review

Hot-Deck (HD) imputation is a popular choice for manipulating MVs in survey research. Hot deck technique finds a similar dataset and imputes MVs by substituting MVs with an observed value from this dataset. Although, this technique is easy to implement but it may be computationally [16]. The method that looks like hot-deck imputation but the data source and current data set must be different from each other is known as Cold-Deck imputation [17]. In many time-series and longitudinal data, one of the most common and used imputation methods is the Last Observation Carried Forward (LOCF). This method imputes each missing value using the last observed value from the same data [18]. The maximum likelihood method can also be used to manipulate MVs . The maximum likelihood assumes that the detected data is a sample taken from a multivariate normal distribution. After the estimation of the parameters using the available information, the MVs are imputed depending on the estimated parameters [19,20]. In regression imputation, the complete features are used to predict the MVs within the features that contain MVs . The predicted values is used to impute the MVs . Regression imputation keeps all data and hence overcomes the pairwise or listwise deletion and does not change the shape of the distribution. In regression imputation, no information is changed or added and the standard error is reduced, hence, little or no biased predictions are generated from the imputation stage [21]. Expectation-Maximization Imputation (EMI) is a kind of the maximum likelihood technique that can be used to manipulate MVs . EMI uses the values assessed by the use of maximum likelihood methods to impute MVs [22]. This method begins with the expectation step, through which the parameters (*e.g.*, means, covariances, and variances) are assessed, possibly by the use of listwise deletion. Predicting MVs is implemented after creation of a regression equation by the use of the estimated parameters. In the maximization step, the regression equations are used to impute MVs . By repeating the expectation and maximization steps until the covariance matrix for the successive iteration is almost the same as that for the previous one. When there is large amount of MVs , EMI method require long time to converge. EMI can result in biased parameter assessments, hence, the standard error is underestimated [21]. KNN imputation technique is considered as one of the most commonly used imputation techniques KNN detects between the complete instances the k most nearest neighbors of a missing data point. The MVs are then imputed with an average of the values of its neighbors in this point. The performance of KNN is extremely bounded especially when the percentage of MVs is high. A simple improvement for manipulating MVs using KNN lies in looking for incomplete neighbors (*i.e.*, act as donors) of an instance given that these neighbors are detected for the features missing instances. This method is known as incomplete case k -nearest neighbors imputation (ICKNNI). ICKNNI in somewhat considered a complex method [23]. Methods that manipulate MVs problems directly without the need of any deletion or imputation step have been developed. For example, logistic regression with MVs by using a Gaussian mixture model to assess the conditional density functions was performed by the authors in [24]. For clustering purposes, the Kernel Spectral Clustering (KSC) algorithm was proposed, which encodes as a set of supplemental soft constraints the partially detected features [25]. MLPimpute is a novel algorithm for handling MVs depending multilayer perceptron (MLP) networks was proposed. Although MLP exhibits a good accuracy

the relationship between data genes is not sufficient for the method [26]. An iterative learning method consists of fuzzy k-means and decision trees was used to manipulate MVs. When this iterative learning compared with KNNimpute it exhibits a better accuracy [27].

3 Proposed Algorithm

This section aims to introduce and elaborate the proposed algorithm in details. The next procedural steps help in clarifying the proposed algorithm.

- **Splitting Dataset:** The proposed algorithm gets a dataset D as input which incorporates MVs, then creates from D two subsets. The first set $X^{(comp)}$ holds all features with no MVs and the second set $X^{(mis)}$ holds all features with MVs. The target feature was assumed to be perfect feature (*i.e.*, does not hold MVs), thus $X^{(comp)}$ holds all perfect features besides the target feature y .
- **Feature Selection:** The proposed algorithm uses the FS of fuzzy entropy measure introduced by Parkash et al. given by Eq. (5) with the similarity-based classification [28]

$$S(\mathbf{x}, \mathbf{v}) = \left(\frac{1}{t} \sum_{r=1}^t w_r (1 - |x(f_r)^p - v(f_r)^p|)^{\frac{m}{p}} \right)^{\frac{1}{m}}, \mathbf{x}, \mathbf{v} \in [0, 1]^t \quad (5)$$

where t represents the number of features of varied types f_1, \dots, f_t that can be observed from the objects, the ideal vector $\mathbf{v}_i = (v_i(f_1), \dots, v_i(f_t))$ should be determined for every class i . $\mathbf{x} = (x(f_1), \dots, x(f_t))$ represents vectors which belong to known class. m is the power value that is obtained from the generalized mean from the generalized Lukasiewicz structure. The parameter p can be detected from the generalized Lukasiewicz structure. w_r is a weight parameter. The weights were set as one.

$$H(A; w) = \sum_{j=1}^n \left(\sin \frac{\pi \mu_A(x_j)}{2} + \sin \frac{\pi(1 - \mu_A(x_j))}{2} - 1 \right) \quad (6)$$

where H represents the fuzzy entropy. j represents the number of features and the fuzzy values are denoted by $\mu_A(x_j)$. A denotes the fuzzy set which is the maximum element of the ordering specified by H when $\mu_A(x) = 0.5$.

The proposed algorithm chooses the feature that exhibits the lowest fuzzy entropy, which gives a strong relationship with the output feature.

- **Imputation:** After the candidate feature $X_g^{(miss)}$ is being selected, the model is fitted using $X^{(comp)}$ as the input features and the candidate feature as target with the cumulative formula described in Eq. (7).

$$X_g^{(miss)} \sim N(\mu_g, \alpha_g) \quad (7)$$

where:

$$\mu_g = \beta_o + \sum_{i=1}^c \beta_i X_i^{(comp)} + \beta_{c+1} y + \sum_{imp=1}^{g-1} \beta_{imp+c+1} X_{imp}^{(miss)2}$$

$$\beta \sim N(0, \lambda_g^{-1} I_{imp+c+1})$$

$$\alpha_g \sim \Gamma(\alpha_{1g}, \alpha_{2g})$$

$$\lambda_g \sim \Gamma(\lambda_{1g}, \lambda_{2g})$$

where $g = 1, 2, \dots, m$. m is the number of features holding MVs and c is the number of perfect features.

- **Update datasets:** The selected feature dropped from $X^{(mis)}$ and after imputing MVs within the $X_g^{(miss)}$, the imputed feature $X_{imp}^{(mis)}$ is added to $X^{(comp)}$. Now $X^{(comp)}$ holds all perfect features, $X_{imp}^{(mis)}$ and y . A new $X_g^{(miss)}$ from $X^{(mis)}$ is chosen. The model is fitted with the cumulative formula with $X^{(comp)}$ as an input features and the new $X_g^{(miss)}$ as the target feature.

Repeat from step 2 of feature selection until $X^{(mis)}$ holds no features, at that moment return ($X^{(comp)}$) as the imputed dataset as described in the following algorithm.

The Proposed Algorithm

- 1: **Input:**
 - 2: D : a dataset with MVs containing n instances.
 - 3: **Output:**
 - 4: $D_{imputed}$: a dataset with all missing features imputed.
 - 5: **Definitions:**
 - 6: $X^{(comp)}$ Set of complete features.
 - 7: $X^{(mis)}$ Set of incomplete features.
 - 8: $X_{imp}^{(mis)}$ Imputed feature from $X^{(mis)}$.
 - 9: m Number of features containing MVs.
 - 10: $H[i]$ Fuzzy entropy measure with the similarity.
 - 11: **Begin**
 - 12: 1: Split D into $X^{(comp)}$ and $X^{(mis)}$
 - 13: 2: From $X^{(mis)}$ select $X_i^{(mis)}$ that exhibits $\min H[i]$ (using similarity classifier with Parkash's fuzzy entropy measure) $\#l \in \{1, \dots, m\}$.
 - 14: 3: While $X^{(mis)} \neq \phi$
 - 15: i: $g \leftarrow$ index of the candidate feature in $X^{(mis)}$.
 - 16: ii: Fit a Bayesian ridge regression model on $X^{(comp)}$ as independent features and $X_g^{(miss)}$ as dependent feature.
 - 17: iii: $X_{imp}^{(mis)} \leftarrow$ Impute the MVs in $X_g^{(miss)}$ with the fitted model.
 - 18: iv: Delete $X_g^{(miss)}$ from $X^{(mis)}$ and add $X_{imp}^{(mis)}$ to $X^{(comp)}$.
 - 19: End While
 - 20: 4: return $D_{imputed} \leftarrow X^{(comp)}$.
 - 21: **End**
-

4 Experimental Setup

4.1 Datasets

Usually, applying and comparing several imputation algorithms on diverse datasets *versus* the proposed algorithm will result in different imputation performances. Furthermore, this difference in imputation helps in judgement about the compared algorithms and the proposed one and also gives an insight about how the proposed algorithm will perform in future and in different situations. The focus in this paper is on medical datasets. The used datasets in this experiment were obtained from several data repositories and are freely access. [Tab. 1](#) gives an overview about the specifications of the datasets used in the experiment. In each dataset, the generation of MVs proportions, 10%, 20%, 30%, 40% and 50%, were performed using the `ampute` function from the R environment [29] for every missingness mechanism, MAR, MCAR and MNAR.

Five practical imputation algorithms were used in the experiment against the proposed algorithm. [Tab. 2](#) describes briefly the compared algorithms used in the experiment.

Table 1: The fundamental specifications the used datasets

Dataset name	#Ins.	#Features	#Class	Missingness mechanism		
				MAR	MCAR	MNAR
Dermatology [30]	366	33	6	10%, 20%, 30%, 40% and 50%		
Breast Cancer [31]	699	10	2			
Parkinsons [32]	197	23	2			
Pima Indians Diabetes [33]	768	8	2			

Table 2: The algorithms used in the comparison

Package (function name)	Description
autoimpute (stochastic) [34]	imputes MVs using the least squares methodology, then adds to the imputations a stochastic element.
autoimpute (nocb) [34]	imputes MVs by carrying next observation moving backward.
SimpleImputer (median) [34]	imputes MVs using the median for each feature.
impyute (EMI) [35]	handles MVs using Expectation Maximization Imputation.
impyute (random) [35]	imputes MVs using a randomly selected value from the same feature.

The experiments were conducted using a laptop with the following specification: Windows 10 OS, 4 GB memory, AMD A4-6210 APU with AMD Radeon R3 Graphics (1.80 GHz) processor, 500 GB HDD and Python (version 3.7) programming language and R (version 3.5.2).

4.2 Evaluation Metrics

Imputation performance can be measured using various metrics. This section exhibits an overview of most metrics used in the experimental implementation to measure the imputation performance; these metrics include MAE, RMSE, and R^2 score.

4.2.1 MAE and RMSE

MAE is used to calculate the average of the absolute differences between the predicted and true values. It gives an intuition about the magnitudes (absolute values) of the error in prediction, but does not offer any idea about the direction of the prediction (*i.e.*, under or over predicting) [36]. RMSE is much like the MAE in that it gives an idea of the magnitude of error. Furthermore, as the variance related to the error magnitudes distribution increases RMSE also increases and MAE is steady. Eqs. (8) and (9) describes MAE and RMSE respectively [3]

$$MAE = \frac{1}{n} \sum_{l=1}^n |y_l - \hat{y}_l| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{l=1}^n (y_l - \hat{y}_l)^2} \quad (9)$$

where the real and predicted values are denoted as y_l and \hat{y}_l of the l th observation respectively and the number of the observations is denoted as n .

4.2.2 R^2 Score

R^2 score, given by Eq. (10), gives an indication of the prediction's goodness of fit to the true values. From a statistical perspective, the R^2 score has been dubbed as the coefficient of determination [14].

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_l - \hat{y}_i)^2}{\sum_{i=1}^n (y_l - \bar{y})^2} \tag{10}$$

where:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_l$$

\bar{y} represents the mean of the detected data.

5 Results and Discussion

Figs. 1 to 3 present the improvement in performance, using RMSE, MAE and R^2 score, of the compared algorithms *versus* the proposed one. The performance evaluation of the proposed algorithm against the compared algorithms for each MVs percentage, 10%, 20%, 30%, 40% and 50%, generated from the missingness mechanisms, MAR, MCAR and MNAR, is presented in more details in Tabs. 3 to 6. The results exhibit that the performance differs from one algorithm to another depending on the dimension of the dataset, the missingness mechanism type, and the amount of MVs in the dataset. The computational complexity of both CBRL and CBRC is $O(n)$.

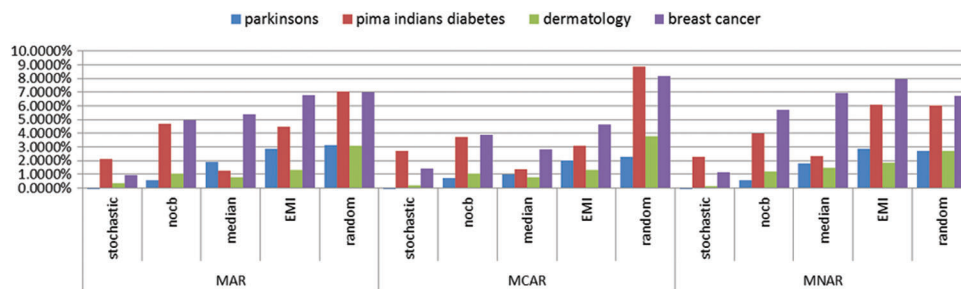


Figure 1: R^2 score improvement percentage of the proposed algorithm *versus* the compared algorithms

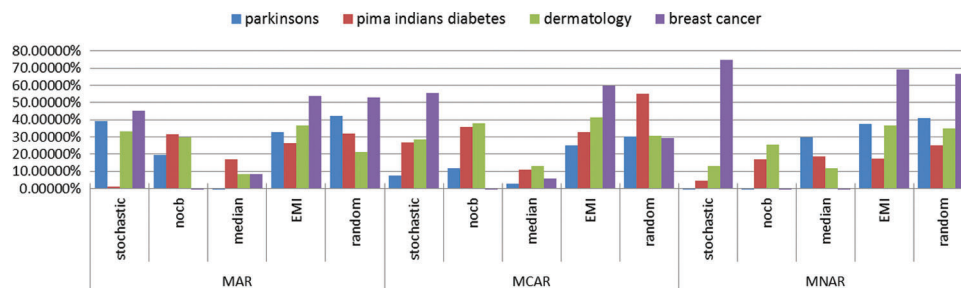


Figure 2: MAE improvement percentage of the proposed algorithm *versus* the compared algorithms

This section is subdivided into two subsections. The first section explains the accuracy analysis evaluated using R^2 score (higher values is better) and the second represents the error analysis evaluated using RMSE and MAE metrics (lower value is better).

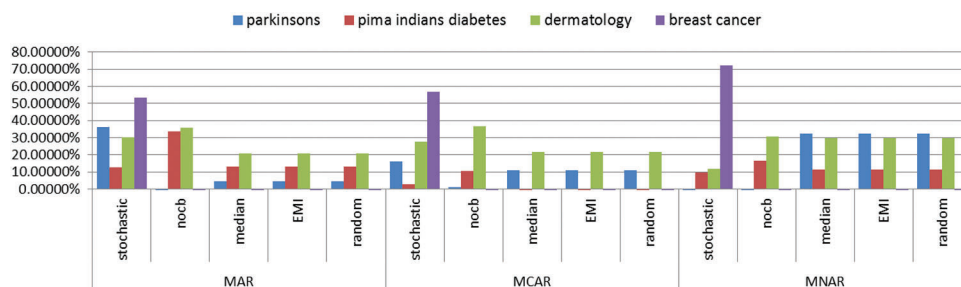


Figure 3: RMSE improvement percentage of the proposed algorithm *versus* the compared algorithms

Table 3: MAE, RMSE and R^2 score of the proposed algorithm against the compared algorithms (**breast cancer** dataset)

		Breast cancer											
Mech	% missing value	Proposed algorithm		Stochastic		NOCB		Median		EMI		Random	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
MAR	10	320.64	14398.76	407.99	16689.91	17.316	1215.105	373.904	16117.67	373.9	19940.49	373.90	16754.00
	20	443.09	14370.89	678.96	21339.27	235.09	12320.38	447.678	16186.82	447.68	41898.87	447.68	88603.04
	30	394.73	12750.13	973.25	25093.19	31.25	1475.361	421.783	14278.08	421.783	22493.84	421.78	15263.37
	40	983.17	21635.06	2532.58	50233.29	404.21	17943.89	901.744	23224.22	901.744	49081.14	901.74	22287.51
	50	834.06	17070.85	1792.33	33483.05	280.10	15684.06	670.810	17701.69	670.810	40735.57	670.81	27890.12
	Average	595.14	16045.14	1277.02	29367.74	193.59	9727.76	563.18	17501.69	563.18	34829.98	563.18	34159.61
	Improvement			0.53	0.45	-2.07	-0.65	-0.06	0.08	-0.06	0.54	-0.06	0.53
MCAR	10	522.92	15759.95	1335.46	40018.23	197.18	14155.30	509.28	17330.07	509.28	37906.66	509.28	17228.65
	20	541.59	11785.89	1335.61	32870.79	256.52	14398.2	378.15	10291.57	378.15	31369.78	378.15	27283.8
	30	791.10	18334.46	1979.26	46012.12	79.82	3571.15	808.66	20314.66	808.65	50625.01	808.66	18119.15
	40	1039.62	19614.44	1531.01	27097.39	354.61	17217.28	905.83	20613.91	905.83	55373.14	905.83	31896.57
	50	1296.27	24990.94	3480.9	57980.1	255.97	11428.92	1289.99	27442.8	1289.99	50641.55	1289.98	33227.33
	Average	838.3	18097.14	1932.45	40795.91	228.82	12154.17	778.38	19198.60	778.38	45183.23	778.38	25551.1
	Improvement			0.57	0.56	-2.66	-0.49	-0.08	0.06	-0.08	0.6	-0.08	0.292
MNAR	10	105.56	4262	490.45	20762.39	4.33	218.99	50.40	2157.73	50.40	18587.69	50.40	11489.77
	20	264.5	7141.54	652.4	20256.7	8.017	284.40	129.97	3845.5	129.97	28673.89	129.97	11829
	30	481.06	9917.79	1785.58	43896.77	19.02	699.58	354.49	7929.36	354.49	32101.63	354.49	16315.06
	40	735.63	13379.96	2861.79	53352.1	29.28	1005.46	603.55	11938.82	603.55	38242.58	603.55	87175.66
	50	812.28	13960.04	2837.58	54955.41	154.06	9768.59	647.49	13497.92	647.49	40577.24	647.5	18928.93
	Average	479.81	9732.27	1725.56	38644.85	42.94	2395.40	357.18	7873.87	357.18	31636.61	357.18	29147.68
	Improvement			0.72	0.75	-10.17	-3.063	-0.34	-0.24	-0.34	0.69	-0.34	0.67

(Continued)

Mech	% missing value	R ² score					
MAR	10	0.989	0.985	0.969	0.965	0.955	0.974
	20	0.984	0.980	0.966	0.947	0.933	0.923
	30	0.977	0.971	0.926	0.923	0.907	0.913
	40	0.978	0.962	0.903	0.912	0.902	0.907
	50	0.968	0.953	0.899	0.898	0.886	0.859
	Average	0.979	0.970	0.933	0.929	0.917	0.915
	Improvement	0.009	0.050	0.054	0.068	0.070	
MCAR	10	0.995	0.989	0.979	0.987	0.980	0.962
	20	0.994	0.985	0.970	0.978	0.965	0.953
	30	0.987	0.972	0.941	0.958	0.943	0.902
	40	0.991	0.979	0.947	0.960	0.936	0.899
	50	0.983	0.956	0.926	0.932	0.906	0.859
	Average	0.990	0.976	0.953	0.963	0.946	0.915
	Improvement	0.014	0.039	0.028	0.046	0.082	
MNAR	10	0.990	0.985	0.966	0.953	0.952	0.973
	20	0.976	0.971	0.932	0.925	0.917	0.947
	30	0.970	0.958	0.917	0.901	0.898	0.909
	40	0.973	0.953	0.913	0.903	0.886	0.884
	50	0.972	0.959	0.889	0.882	0.870	0.860
	Average	0.976	0.965	0.924	0.913	0.904	0.915
	Improvement	0.012	0.057	0.069	0.079	0.067	

Table 4: MAE, RMSE and R² score of the proposed algorithm against the compared algorithms (dermatology dataset)

Dermatology													
Mech	% missing value	Proposed algorithm		Stochastic		NOCB		Median		EMI		Random	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MAR	10	0.265	0.005	0.223	0.004	0.196	0.004	0.247	0.005	0.269	0.005	0.247	0.005
	20	0.166	0.003	0.371	0.006	0.320	0.006	0.208	0.004	0.490	0.004	0.218	0.004
	30	0.237	0.006	0.499	0.011	0.529	0.013	0.265	0.009	0.198	0.009	0.443	0.009
	40	0.194	0.006	0.252	0.009	0.287	0.012	0.250	0.009	0.350	0.009	0.194	0.009
	50	0.509	0.015	0.708	0.021	0.623	0.020	0.531	0.017	0.863	0.017	0.639	0.017
	Average	0.274	0.007	0.411	0.010	0.391	0.011	0.300	0.009	0.434	0.009	0.348	0.009
	Improvement			0.332	0.302	0.298	0.357	0.085	0.207	0.368	0.207	0.212	0.207

(Continued)

Table 4 (continued)

Dermatology													
Mech	% missing value	Proposed algorithm		Stochastic		NOCB		Median		EMI		Random	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MCAR	10	0.325	0.005	0.277	0.005	0.430	0.007	0.331	0.006	0.413	0.006	0.519	0.006
	20	0.166	0.004	0.487	0.010	0.366	0.008	0.269	0.007	0.399	0.007	0.311	0.007
	30	0.237	0.006	0.319	0.008	0.282	0.009	0.226	0.007	0.342	0.007	0.288	0.007
	40	0.155	0.006	0.138	0.008	0.242	0.010	0.188	0.008	0.434	0.008	0.193	0.008
	50	0.182	0.008	0.265	0.010	0.394	0.014	0.207	0.010	0.221	0.010	0.224	0.010
	Average	0.213	0.006	0.297	0.008	0.343	0.010	0.244	0.008	0.362	0.008	0.307	0.008
	Improvement			0.284	0.277	0.379	0.368	0.129	0.218	0.412	0.218	0.306	0.218
MNAR	10	0.158	0.004	0.173	0.005	0.236	0.006	0.180	0.005	0.310	0.005	0.331	0.005
	20	0.373	0.008	0.573	0.012	0.465	0.013	0.386	0.012	0.423	0.012	0.604	0.012
	30	0.271	0.008	0.349	0.010	0.412	0.011	0.329	0.012	0.409	0.012	0.335	0.012
	40	0.325	0.012	0.327	0.012	0.538	0.019	0.350	0.017	0.304	0.017	0.548	0.017
	50	0.372	0.014	0.303	0.013	0.367	0.016	0.456	0.019	0.923	0.019	0.485	0.019
	Average	0.300	0.009	0.345	0.010	0.403	0.013	0.340	0.013	0.474	0.013	0.461	0.013
	Improvement			0.131	0.119	0.257	0.307	0.119	0.296	0.367	0.296	0.349	0.296
Mech	% missing value	R ² score											
MAR	10	0.999		0.998		0.995		0.998		0.997		0.993	
	20	0.998		0.996		0.993		0.994		0.988		0.981	
	30	0.995		0.992		0.985		0.985		0.981		0.965	
	40	0.993		0.987		0.974		0.982		0.977		0.952	
	50	0.992		0.987		0.978		0.979		0.971		0.937	
	Average	0.995		0.992		0.985		0.988		0.983		0.966	
	Improvement			0.003		0.010		0.008		0.013		0.031	
MCAR	10	0.998		0.997		0.994		0.994		0.992		0.983	
	20	0.998		0.995		0.992		0.992		0.989		0.972	
	30	0.995		0.993		0.984		0.988		0.980		0.967	
	40	0.993		0.991		0.982		0.981		0.977		0.944	
	50	0.991		0.989		0.971		0.979		0.972		0.926	
	Average	0.995		0.993		0.985		0.987		0.982		0.958	
	Improvement			0.002		0.010		0.008		0.013		0.038	
MNAR	10	0.996		0.997		0.987		0.988		0.986		0.985	
	20	0.993		0.991		0.984		0.980		0.978		0.971	
	30	0.991		0.989		0.980		0.978		0.976		0.967	
	40	0.988		0.989		0.974		0.968		0.962		0.954	
	50	0.991		0.987		0.975		0.974		0.966		0.950	
	Average	0.992		0.991		0.980		0.977		0.974		0.965	
	Improvement			0.001		0.012		0.015		0.019		0.027	

Table 5: MAE, RMSE and R^2 score of the proposed algorithm against the compared algorithms (parkinsons dataset)

Parkinsons													
Mech	% missing value	Proposed algorithm		Stochastic		NOCB		Median		EMI		Random	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MAR	10	1.267	0.035	2.070	0.043	0.662	0.014	0.831	0.021	1.828	0.021	1.596	0.021
	20	1.774	0.064	1.972	0.057	6.857	0.137	1.884	0.070	1.927	0.070	2.103	0.070
	30	1.974	0.072	2.305	0.084	0.385	0.016	1.934	0.069	4.167	0.069	2.300	0.069
	40	2.656	0.115	6.924	0.278	1.850	0.066	2.178	0.103	3.131	0.103	7.423	0.103
	50	2.110	0.099	2.868	0.140	2.425	0.064	2.904	0.140	3.492	0.140	3.516	0.140
	Average	1.956	0.077	3.228	0.120	2.436	0.059	1.946	0.081	2.909	0.081	3.388	0.081
	Improvement			0.394	0.361	0.197	-0.297	-0.005	0.046	0.328	0.046	0.423	0.046
MCAR	10	1.085	0.024	1.734	0.042	0.223	0.005	0.951	0.025	2.464	0.025	2.018	0.025
	20	5.895	0.109	5.047	0.113	6.452	0.102	6.323	0.129	7.252	0.129	5.956	0.129
	30	1.493	0.056	3.553	0.127	0.904	0.022	1.766	0.076	2.639	0.076	6.722	0.076
	40	1.794	0.071	1.449	0.048	2.606	0.070	2.071	0.079	2.628	0.079	2.130	0.079
	50	7.332	0.240	7.272	0.268	9.779	0.306	7.027	0.254	8.483	0.254	8.404	0.254
	Average	3.520	0.100	3.811	0.119	3.993	0.101	3.628	0.113	4.693	0.113	5.046	0.113
	Improvement			0.076	0.163	0.119	0.011	0.030	0.111	0.250	0.111	0.302	0.111
MNAR	10	1.060	0.022	0.766	0.013	1.164	0.018	1.339	0.025	1.629	0.025	1.722	0.025
	20	1.042	0.027	1.801	0.047	0.443	0.012	2.130	0.061	1.701	0.061	3.160	0.061
	30	2.922	0.099	2.797	0.091	5.653	0.140	3.725	0.143	5.184	0.143	4.305	0.143
	40	4.320	0.165	4.111	0.142	2.008	0.065	4.957	0.180	5.823	0.180	6.201	0.180
	50	1.885	0.088	1.692	0.077	1.601	0.055	3.837	0.186	3.676	0.186	3.674	0.186
	Average	2.246	0.080	2.234	0.074	2.174	0.058	3.197	0.119	3.602	0.119	3.813	0.119
	Improvement			-0.005	-0.085	-0.033	-0.381	0.298	0.325	0.377	0.325	0.411	0.325
Mech	% missing value	R^2 score											
MAR	10	0.997		0.998		0.997		0.990		0.988		0.985	
	20	0.992		0.997		0.976		0.970		0.966		0.970	
	30	0.981		0.991		0.988		0.966		0.950		0.973	
	40	0.996		0.985		0.984		0.983		0.968		0.948	
	50	0.992		0.993		0.984		0.956		0.949		0.931	
	Average	0.992		0.993		0.986		0.973		0.964		0.961	
	Improvement			-0.001		0.005		0.019		0.028		0.031	
MCAR	10	1.000		0.999		0.999		0.997		0.994		0.994	
	20	0.994		0.995		0.989		0.987		0.981		0.974	
	30	0.996		0.997		0.990		0.984		0.974		0.973	
	40	0.996		0.997		0.992		0.989		0.975		0.980	
	50	0.987		0.986		0.967		0.968		0.950		0.940	
	Average	0.994		0.995		0.987		0.985		0.975		0.972	
	Improvement			-0.001		0.007		0.010		0.020		0.023	
MNAR	10	0.999		0.999		0.995		0.990		0.986		0.988	
	20	0.996		0.997		0.996		0.983		0.973		0.976	
	30	0.995		0.996		0.988		0.974		0.960		0.968	
	40	0.990		0.993		0.979		0.968		0.965		0.947	
	50	0.992		0.994		0.988		0.970		0.951		0.962	
	Average	0.995		0.996		0.989		0.977		0.967		0.968	
	Improvement			-0.001		0.005		0.018		0.028		0.027	

Table 6: MAE, RMSE and R² score of the proposed algorithm against the compared algorithms (Pima Indians Diabetes dataset)

Pima Indians Diabetes													
Mech	% missing value	Proposed algorithm		Stochastic		NOCB		Median		EMI		Random	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
MAR	10	0.379	9.700	0.332	7.291	0.455	11.393	0.448	11.316	0.448	11.753	0.448	10.818
	20	0.510	6.424	0.573	6.635	0.760	9.864	0.511	7.373	0.511	9.472	0.511	10.740
	30	0.558	7.566	0.735	9.264	0.746	8.699	0.644	8.909	0.644	9.968	0.644	10.516
	40	0.878	9.244	1.038	9.677	1.343	15.416	0.990	11.254	0.990	13.486	0.990	14.768
	50	0.959	9.039	1.088	9.502	1.640	15.887	1.185	11.715	1.185	12.472	1.185	14.980
	Average	0.656	8.395	0.753	8.474	0.989	12.252	0.756	10.113	0.756	11.430	0.756	12.364
	Improvement			0.128	0.009	0.336	0.315	0.131	0.170	0.131	0.266	0.131	0.321
MCAR	10	0.849	3.821	0.473	7.773	0.352	5.153	0.266	3.798	0.247	4.384	0.247	11.382
	20	0.874	7.494	0.688	9.948	0.670	10.193	0.430	7.684	0.504	11.153	0.504	13.846
	30	1.021	10.058	0.953	10.781	1.084	12.833	0.697	12.217	0.947	14.416	0.947	16.033
	40	0.845	7.480	1.012	9.752	1.130	12.879	0.702	7.826	0.702	9.056	0.702	18.766
	50	0.939	8.057	1.531	12.185	1.818	16.443	1.147	9.919	1.147	15.752	1.147	22.155
	Average	0.906	7.382	0.932	10.088	1.011	11.500	0.649	8.289	0.709	10.952	0.709	16.436
	Improvement			0.028	0.268	0.104	0.358	-0.396	0.109	-0.277	0.326	-0.277	0.551
MNAR	10	1.070	8.684	0.484	9.689	0.469	9.466	0.502	10.700	0.502	9.781	0.502	9.367
	20	1.112	14.583	0.696	13.146	0.883	15.400	0.790	16.248	0.790	14.527	0.790	13.169
	30	0.86	7.75	1.06	10.36	1.16	11.70	1.06	10.98	1.06	11.56	1.06	18.88
	40	0.84	14.04	1.33	12.99	1.38	16.51	1.32	16.58	1.32	19.15	1.32	16.53
	50	0.84	12.07	1.66	13.74	1.77	15.70	1.66	15.82	1.66	14.25	1.66	18.23
	Average	0.94	11.43	1.05	11.99	1.13	13.76	1.07	14.07	1.07	13.85	1.07	15.24
	Improvement			0.10	0.05	0.17	0.17	0.12	0.19	0.12	0.18	0.12	0.25

Mech	% missing value	R ² score					
MAR	10	0.98	0.98	0.97	0.98	0.96	0.96
	20	0.98	0.97	0.95	0.97	0.96	0.94
	30	0.97	0.94	0.93	0.96	0.94	0.92
	40	0.95	0.91	0.88	0.92	0.88	0.86
	50	0.95	0.92	0.88	0.93	0.88	0.84
	Average	0.97	0.95	0.92	0.95	0.92	0.90
	Improvement		0.02	0.05	0.01	0.04	0.07
MCAR	10	0.99	0.97	0.97	0.99	0.98	0.96
	20	0.98	0.96	0.96	0.98	0.96	0.93
	30	0.96	0.95	0.92	0.95	0.92	0.86
	40	0.96	0.92	0.90	0.95	0.93	0.85
	50	0.96	0.92	0.92	0.92	0.92	0.85
	Average	0.97	0.94	0.93	0.96	0.94	0.89
	Improvement		0.03	0.04	0.01	0.03	0.09
MNAR	10	0.98	0.97	0.96	0.96	0.95	0.96
	20	0.95	0.94	0.93	0.94	0.93	0.94
	30	0.95	0.93	0.91	0.93	0.88	0.87
	40	0.94	0.91	0.90	0.91	0.87	0.88
	50	0.93	0.89	0.87	0.90	0.85	0.82
	Average	0.95	0.93	0.91	0.93	0.90	0.90
	Improvement		0.02	0.04	0.02	0.06	0.06

5.1 Accuracy Analysis

This subsection exhibits that the proposed algorithm offers better accuracy *versus* the compared algorithms in many cases. The accuracy analysis is represented by calculating R^2 score. Fig. 1 exhibits the improvement percentage of R^2 score which is given by Eq. (10) for the proposed algorithm *versus* the compared algorithms. In what follows, the comparison of R^2 score is discussed in detail. In all missigness mechanisms, R^2 score of the proposed algorithm is better than noCb, median, EMI and random when applied on all datasets used in the experiment. In addition, R^2 score of the proposed algorithm is better than stochastic when applied on all used datasets but worse than stochastic when applied on parkinsons dataset in all missigness mechanisms.

5.2 Error Analysis

This subsection exhibits that the proposed algorithm gives lower error *versus* the compared algorithms in many cases. MAE and RMSE, given by Eqs. (8) and (9) respectively, are the metrics used for assessing the error in imputation. Figs. 2 and 3 exhibit the improvement percentage in evaluating both MAE and RMSE respectively.

In all missigness mechanisms, MAE given by the proposed algorithm is lower than MAE given by EMI and random when applied on all datasets used in the experiment. When the proposed algorithm is compared with stochastic in MAR and MCAR, it was observed that MAE granted by the proposed algorithm is the lowest in all used datasets. In MNAR, MAE of the proposed algorithm is better than stochastic in all used datasets except when applied on the parkinsons dataset. In MAR and MCAR, MAE of the proposed algorithm is better than noCb in all used datasets except when applied on the breast cancer dataset. In MNAR, MAE of the proposed algorithm is better than noCb in all used datasets except when applied on the breast cancer and parkinsons datasets. When the proposed algorithm is compared with median in MCAR it was observed that MAE of the proposed algorithm is better in all used datasets. In MAR, MAE of the proposed algorithm is better than median in all used datasets except when applied on the parkinsons dataset. In MNAR, MAE of the proposed algorithm is better than median in all used datasets except when applied on the breast cancer dataset.

In MAR and MCAR, RMSE of the proposed algorithm is better than stochastic in all used datasets. Also in MNAR, RMSE of the proposed algorithm is better than stochastic in all used datasets except when applied on the parkinsons dataset. When the proposed algorithm is compared with noCb in MAR and MNAR, it was noticed that RMSE of the proposed algorithm is better than noCb in all used datasets except when applied on the parkinsons and breast cancer datasets. When the proposed algorithm is compared with median, EMI and random in MAR and MNAR, it was observed that RMSE of the proposed algorithm is better in all used datasets. Also in MCAR, RMSE of the proposed algorithm is better than RMSE of median, EMI and random when applied in all datasets used in the experiment except in Pima Indians Diabetes and breast cancer datasets.

6 Conclusion

MVs are considered a critical problem in pattern recognition, ML and data mining applications. Many extensive studies have been performed for manipulating the problem of MVs especially in medical data. In addition to MVs, FS is the data preprocessing strategy which has been considered to be efficient when preparing data (specifically large volume data) in ML. It has been confirmed to be efficient and effective in handling high-dimensional data for any data dependent tool. Reducing the computational cost of modeling and the number of input features helps in improving the performance of the model.

In this paper, novel algorithm was proposed to manipulate MVs. The proposed algorithm depends on FS of similarity classifier with Parkash's fuzzy entropy measure to select the candidate feature and the BRR

model to predict MVs in the selected feature. Hence, the proposed algorithm consists mainly of two phases. In the first phase, the FS of similarity classifier with Parkash's fuzzy entropy is used to select features to be imputed one after one. In the second phase, the MVs in the selected feature are predicted using the BRR model. The first and second phases are repeated until the imputation of the whole dataset. The proposed algorithm is easy to implement and can deal with all MVs from any missingness mechanism. Furthermore, the proposed algorithm exhibits a good performance against the compared algorithms.

In future research, the proposed algorithm will be implemented on new medical datasets like pulmonary embolism data and cardiovascular disease. Furthermore, additional performance metrics will be taken into consideration such as the normalized root mean square error (NRMSE), statistical tests and predictive accuracy (PAC).

Acknowledgement: This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University (KAU), Jeddah, Saudi Arabia, under grant No. (PH: 13-130-1442). The authors, therefore acknowledge with thanks DSR for technical and financial support.

Funding Statement: This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University (KAU), Jeddah, Saudi Arabia, under grant No. (PH: 13-130-1442).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] M. S. Osman, A. M. Abu-Mahfouz and P. R. Page, "A Survey on data imputation techniques: Water distribution system as a use case," *IEEE Access*, vol. 6, pp. 63279–63291, 2018.
- [2] D. Li, H. Zhang, T. Li, A. Bouras, X. Yu *et al.*, "Hybrid missing value imputation algorithms using fuzzy c-means and vaguely quantified rough set," in *IEEE Transactions on Fuzzy Systems*. Early Acce, 2021. DOI 10.1109/TFUZZ.2021.3058643.
- [3] S. M. Mostafa, A. S. Eladimy, S. Hamad and H. Amano, "CBRG: A novel algorithm for handling missing data using bayesian ridge regression and feature selection based on gain ratio," *IEEE Access*, vol. 8, pp. 216969–216985, 2020.
- [4] X. Zhu, J. Wang, B. Sun, C. Ren, T. Yang *et al.*, "An efficient ensemble method for missing value imputation in microarray gene expression data," *BMC Bioinformatics*, vol. 22, no. 1, pp. 1–25, 2021.
- [5] D. I. Lewin, "Getting clinical about neural networks," *IEEE Intelligent Systems and their Applications*, vol. 15, no. 1, pp. 2–5, 2000.
- [6] A. N. Baraldi and C. K. Enders, "An introduction to modern missing data analyses," *Journal of School Psychology*, vol. 48, no. 1, pp. 5–37, 2010.
- [7] G. Doquire and M. Verleysen, "Feature selection with missing data using mutual information estimators," *Neurocomputing*, vol. 90, no. 2, pp. 3–11, 2012.
- [8] S. M. Mostafa, "Missing data imputation by the aid of features similarities," *Int. Journal of Big Data Management*, vol. 1, no. 1, pp. 81–103, 2020.
- [9] S. M. Mostafa, "Imputing missing values using cumulative linear regression," *CAAI Transactions on Intelligence Technology*, vol. 4, no. 3, pp. 182–200, 2019.
- [10] M. L. Yadav and B. Roychoudhury, "Handling missing values: A study of popular imputation packages in R," *Knowledge-Based Systems*, vol. 160, no. 9, pp. 104–118, 2018.
- [11] A. C. Acock, "Working with missing values," *Journal of Marriage and Family*, vol. 67, no. 4, pp. 1012–1028, 2005.
- [12] M. B. Albayati and A. M. Altamimi, "An empirical study for detecting fake facebook profiles using supervised mining techniques," *Informatica*, vol. 43, no. 1, pp. 77–86, 2019.

- [13] P. Madley-Dowd, R. Hughes, K. Tilling and J. Heron, "The proportion of missing data should not be used to guide decisions on multiple imputation," *Journal of Clinical Epidemiology*, vol. 110, no. 1, pp. 63–73, 2019.
- [14] S. M. Mostafa, A. S. Eladimy, S. Hamad and H. Amano, "CBRL and CBRC: Novel algorithms for improving missing value imputation accuracy based on bayesian ridge regression," *Symmetry (Basel)*, vol. 12, no. 10, pp. 1594, 2020.
- [15] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa *et al.*, "Scikit-learn: Machine learning without learning the machinery," *GetMobile: Mobile Computing and Communications*, vol. 19, no. 1, pp. 29–33, 2015.
- [16] P. L. Roth, "Missing data: A conceptual review for applied psychologists," *Personnel Psychology*, vol. 47, no. 3, pp. 537–560, 1994.
- [17] P. J. García-Laencina, J. L. Sancho-Gómez and A. R. Figueiras-Vidal, "Pattern classification with missing data: A review," *Neural Computing and Applications*, vol. 19, no. 2, pp. 263–282, 2010.
- [18] R. M. Hamer and P. M. Simpson, "Last observation carried forward versus mixed models in the analysis of psychiatric clinical trials," *American Journal of Psychiatry*, vol. 166, no. 6, pp. 639–641, 2009.
- [19] R. J. Little and D. B. Rubin, "Maximum likelihood for general patterns of missing data: Introduction and theory with ignorable nonresponse," in *Statistical Analysis with Missing Data*. Second Edition, John Wiley & Sons, Inc., Hoboken, New Jersey, USA, pp. 164–189, 2002.
- [20] K. M. Lang and T. D. Little, "Principled missing data treatments," *Prevention Science*, vol. 19, no. 3, pp. 284–294, 2018.
- [21] H. Kang, "The prevention and handling of the missing data," *Korean Journal of Anesthesiology*, vol. 64, no. 5, pp. 402–406, 2013.
- [22] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
- [23] J. Van Hulse and T. M. Khoshgoftaar, "Incomplete-case nearest neighbor imputation in software measurement data," *Information Sciences*, vol. 259, no. 2, pp. 596–610, 2014.
- [24] D. Williams, X. Liao, Y. Xue and L. Carin, "Incomplete-data classification using logistic regression," in *Proc. of the 22nd Int. Conf. on Machine Learning*, Association for Computing Machinery, New York, NY, United States, pp. 972–979, 2005.
- [25] K. Wagstaff, "Clustering with missing values: No imputation required," in *Classification, Clustering, and Data Mining Applications*. Berlin, Heidelberg: Springer, pp. 649–658, 2004.
- [26] E. L. Silva-Ramírez, R. Pino-Mejías, M. López-Coello and M. D. Cubiles-de-la-Vega, "Missing value imputation on missing completely at random data using multilayer perceptrons," *Neural Networks*, vol. 24, no. 1, pp. 121–129, 2011.
- [27] S. Nikfalazar, C. H. Yeh, S. Bedingfield and H. A. Khorshidi, "Missing data imputation using decision trees and fuzzy clustering with iterative learning," *Knowledge and Information Systems*, vol. 62, no. 6, pp. 2419–2437, 2020.
- [28] P. Luukka, "Feature selection using fuzzy entropy measures with similarity classifier," *Expert Systems with Applications*, vol. 38, no. 4, pp. 4600–4607, 2011.
- [29] S. Rianne, L. Peter, B. Jaap and V. Gerko, "Generate missing values with ampute," 2017, [Online]. Available: https://rianneschouten.github.io/mice_ampute/vignette/ampute.html.
- [30] M. D. Nilsel Ilter and H. A. Guvenir, "Dermatology," [Online]. 2021. Available: <https://archive.ics.uci.edu/ml/datasets/dermatology>.
- [31] Wi. H. Wolberg, "Breast cancer wisconsin," [Online]. 2021. Available: [https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(original\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)).
- [32] Max Little, "Parkinsons data set," [Online]. 2021. Available: <https://archive.ics.uci.edu/ml/datasets/parkinsons>.
- [33] R. A. Rossi and Nesreen K. Ahmed, "Pima Indians Diabetes," [Online]. 2021. Available: <http://networkrepository.com/pima-indians-diabetes.php>.
- [34] J. Kearney and S. Barkat, "Autoimpute," [Online]. 2021. Available: <https://autoimpute.readthedocs.io/en/latest/>.
- [35] E. Law, "Impyute," [Online]. 2021. Available: <https://impyute.readthedocs.io/en/latest/>.
- [36] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005.