



Missing Health Data Pattern Matching Technique for Continuous Remote Patient Monitoring

Teena Arora, Venki Balasubramanian, and Andrew Stranieri^(✉)

Federation University Australia, Ballarat, Australia
teenaarora@students.federation.edu.au, {v.balasubramanian,
a.stranieri}@federation.edu.au

Abstract. Remote patient monitoring (RPM) has been gaining popularity recently. However, health data acquisition is a significant challenge associated with patient monitoring. In continuous RPM, health data acquisition may miss health data during transmission. Missing data compromises the quality and reliability of patient risk assessment. Several studies suggested techniques for analyzing missing data; however, many are unsuitable for RPM. These techniques neglect the variability of missing data and provide biased results with imputation. Therefore, a holistic approach must consider the correlation and variability of the various vitals and avoid biased imputation. This paper proposes a coherent computation pattern-matching technique to identify and predict missing data patterns. The performance of the proposed approach is evaluated using data collected from a field trial. Results show that the technique can effectively identify and predict missing patterns.

Keywords: Missing health data · Pattern-matching technique · Remote patient monitoring · IoMT · Smart healthcare

1 Introduction

In recent years, wireless vital sign sensors, commonly known as the Internet of Medical Things (IoMT), have broadly extended the boundaries of remote patient monitoring (RPM) [1, 2]. These sensors remotely and continuously measure a patient's five vital signs, such as oxygen saturation (SpO₂), blood pressure (BP), body temperature (Temp), respiratory rate (RR), and heart rate (HR), which allows for early identification of any abnormalities or deterioration in the patient's conditions [3]. Continuously monitored patient data and its extracted features can be used for decision-making, early clinical event prediction models [4], and automated risk analysis [5, 6].

The modified early warning score (MEWS) is widely used in hospital wards for nurses to assess a patient's condition and raise alerts [7]. First introduced as an early warning score (EWS) [8], it was replaced with the modified early warning score (MEWS) [9, 10]. Weighted scores from 0 to 3 for each five vital signs parameter datasets are assigned

from thresholds on raw values; single vital parameter scores are summed to generate a MEWS score [10, 11], as shown in Table 1. Several variations to patient deterioration using MEWS and other approaches, including single parameter, multiple parameters, the aggregate score, and combination system methodologies, have been reviewed [11]; however, approaches commonly assume the minimal value of the parameter if there is no value for a particular parameter. Therefore, missing data refers to any parameter value with no recorded data in the five vital signs parameter dataset., Much research has focused on dealing with missing data [12, 13]; however, missing data in continuous vital signs in RPM is an unexplored area. The number of parameter data observed continuously in RPM makes the database highly susceptible to missing data occurrence, and its analysis is a significant challenge. Missing data is generally categorized into three categories: Missing completely at random (MCAR), Missing at random (MAR), Missing not at random (MNAR) [13].

Table 1. Modified Early Warning Score [9, 10]

Vital Signs	3	2	1	0	1	2	3
Systolic Blood Pressure(BP) (mmHg)	<70	71–80	81–100	101–199	–	>=200	–
Heart-Rate (HR)(bpm)	–	<40	41–50	51–100	101–110	111–129	>=130
Respiratory Rate(RR) (bpm)	–	<9	–	9–14	15–20	21–29	>=30
Temperature (Temp) (degree C)	–	<35	–	35–38.4	–	>=38.5	–
Oxygen Saturation (SpO ₂)	<85	85–89	90–93	>94	–	–	–

The amount of data missing in any clinical setting depends upon various causes, such as delayed data transmission, environmental issues, connectivity issues due to a network, sensor malfunction, power loss, and sensor detachment from the skin. Consequently, missing or poor-quality data will impact the evaluation of vital sign abnormalities, clinical event detection algorithms, or the risk analysis of the patient [14].

Extensive work exists on analyzing missing data in the literature [13, 15]. A general regression neural network (GRNN) and successive geometric transformation model (SGTM) were applied in a study to handle the missing clinical data [16]. Missing data may frequently occur in RPM monitoring, and the amount of recorded data variations makes it challenging to analyze as these algorithms need significant data for training [17]. In addition, these techniques often ignore unique patterns in data and correlations between the observed vitals.

This study proposes a similar pattern match to continuously provide missing values imputation based on a temporal association of vitals in a sliding window approach. Patterned early modified warning score (PMEWS) is introduced in the research as a pre-processing feature for finding pattern matches for the missing vital signs values. This pattern-matching technique helps identify the trends and patterns in the data even when the variability of the missing data is significant. This approach uses the required vital signs data from a clinical trial from our RPM real-time implementation [18]. The main contribution of this study is summarized below:

- We propose a pattern-matching algorithm to impute the missing sensor data from different sensor streams. Our algorithm creates the patterns of vital signs values appearing at a particular time to impute missing data using the pre-processing feature PMEWS.
- Our novel algorithm can detect and predict similar matching patterns for the missing values.
- The proposed approach utilizes the MAR and MCAR to simulate incomplete data proportions of 10%, 20%, and 30%. Then, to check the accuracy of our proposed algorithm, the actual dataset without omission is verified with our suggested matching of similar patterns for the missing values.

This paper is organized as follows: Sect. 2 outlines the related work of this research. Section 3 describes our proposed solution. Section 4 describes our results, and Sect. 5 concludes the paper.

2 Related Work

Imputation is the recovery-based approach to overcome the limitation of complete case analysis if the missing data falls under the MCAR category [14]. Imputation is used to replace the missing data values using single mean imputation, conditional mean imputation, and last observation carries forward (LOCF), multiple imputations (MI), and full information maximum likelihood (FIML) [19]. Single-mean imputation, where the missing data is imputed by the mean of the recorded value of that attribute, has a limitation that affects the dataset's variation. In comparison, conditional mean imputation intensifies the multivariate relationship in the data. Due to these limitations of mean and conditional imputation, these are recommended for baseline measurements in randomized trials [20]. LOCF is similar to single imputation, which replaces the missing values with the last observed value. MI has three phases: the imputation phase, the analysis phase, and the pooling phase [21]. MI generates n replacements for missing values. Then these completed datasets generated with n imputed values are analyzed and integrated into the final output. MI applies to both MAR and MCAR. FIML best suits the MAR context as it considers only the observed data and ignores non-response data [22]. These imputation methods are inappropriate if the missing data falls under the MNAR category.

Random Forest is inherently a MI approach, and it finds the mean of the data from numerous unpruned classification or regression trees for missForest [23, 24]. Bayesian ridge regression used the probability distributor when designing the linear regression, which enables the automated process for the missing data [25]. A cluster's variable mean, or mode is used to impute the missing values in the hot/cold deck [26]. Using the Euclidean distance, the K-nearest neighbour imputes the missing values with a similar one by defining the similarity between two values [26].

Normalization, formatting, and synchronization of the data are also necessary [27]. Normalizing the recorded data variables is essential as various scoring systems, such as the sequential organ failure assessment (SOFA) and MEWS, were registered at different frequencies [28]. Removing abnormal values from the data is also described in the literature as a part of the pre-processing [29]. One typical statistical analysis, the complete case, removes all the rows with the missing values and studies all accounts with the exclusive data [30]. This method has a disadvantage: it will reduce the amount of valuable data, producing biased results. Von Russum [12] analyzes techniques such as linear interpolation, spline interpolation, last observation, mean-forward, and cluster-based imputation for missing values. Imputation using these resulted in early warning score (EWS) miscalculations ranging from 1% to 8%. These methods produced more biased results than oversized observation windows [12]. A deep learning-based protocol for accurately predicting the missing data is also proposed in the literature [31]. Expectation Maximization (EM) [32] is a popular algorithm for imputing the missing data. One disadvantage of EM is that the recorded data should be highly correlative to get more reliable information [33].

Existing approaches perform well in imputation accuracy, though their efficiency is greatly affected by various factors such as computational time, complexity, and correlation of parameters. Our proposed pattern-matching imputation approach considers factors such as the correlation of the parameters, which helps to reduce the biased imputation and computational time can be set using the sliding window to show efficient performance in handling the missing values.

3 Pattern-Matching for Missing Values

The structure of our method is divided into two stages: Initially, the data is pre-processed using the patterns created using the MEWS [34] discussed in Sect. 3.1. Pre-processing the data helps ensure an accurate assessment of the patient's condition. Pre-processing includes removing erroneous values, no recorded vital values at a time, abnormal values, and high-frequency noise [35]. In the second stage, a pattern match is found for the missing values in an observation window. These two components work together to provide a high similarity match for the missing data.

Vital signs data for parameters systolic blood pressure, blood oxygen, body temperature, respiratory rate, and heart rate streams at a discrete time for each parameter using the wearable sensors recorded from patients in a general ward [18] were used in the current study. Publicly available datasets were not used to ensure data is simulated real data and were not curated in any other ways. The imputation method considers the data’s temporal context and more accurately predicts the missing values using a sliding window. A sliding window has been used to find patterns for a particular data segment [21, 22]. A six-hour prediction window for clinical deterioration is considered ideal [36]. The sliding window is predefined in our algorithm.

A sliding window analyses a subset of data within the process window of fixed length, often with a minimum overlap of one data point and shifting with the window increment time in the process window. The notations used for the pattern-match algorithm are the number of vitals to be recorded (T_{MV}), Sliding window time (t), String patterned array ($S_P[i]$), Variables ($Str1, Str2$), Counter for the matched character ($CS_P[k]$). Figure 1 shows the sliding window increment and overlaps in the process window. $T10, T20, T30, T40, T50,$ and $T60$ denote the time in Fig. 1. The sliding window (sw_{δ}^i) consists of different time slots $t_{(\delta+n)}^i$ shown in Eq. (1).

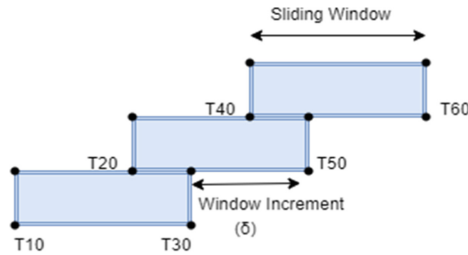


Fig. 1. Sliding window with the overlap and window increment

The process window length (L), where ($L \in N$) N is the set of all the natural numbers, and the slide window increment (δ), where ($1 < = \delta < L$), can be decided by the physician, as shown in Fig. 1. If urgent attention is needed, a shorter window increment time can deliver more frequent information on the patient’s condition. On the other hand, a longer window increment can be employed to lower the computation burden and boost the analysis accuracy if the patient’s condition is stable and not urgent. A significant observation window with personalized decision-making and imputation produced positive results for missing values on maternal health data [14]. The window number (i) where ($i \in N$). Notation ($t_{(\delta+n)}^i$) will be written as TS for calculations in the rest of the paper, as shown in Eq. (2).

$$sw_{\delta}^i = (t_{(\delta+0)}^i, t_{(\delta+1)}^i, t_{(\delta+2)}^i \dots \dots \dots t_{(\delta+n)}^i) \tag{1}$$

$$TS = (t_{(\delta+n)}^i) \tag{2}$$

The physician will decide the start time (T) and the time slot (TS) for the observation. The slide window increment (δ) value will be valid for the time slot (TS) and can be changed for the next process window (L). “Patterned modified early warning score” (PMEWS) is described in the next section. The PMEWS forms the basis for similar pattern matches for missing vitals values. The method is described in the next section after PMEWS.

3.1 Pattern Modified Early Warning Score (PMEWS)

Our method for handling missing values in our data set is to make a pattern of the values from the recorded set, replacing the missing values with letters or symbols. As shown in Table 2, recorded raw, vital values at any data points are converted into their respective MEWS score to form the pattern, where ‘N’ is used for the null values. Another significant benefit of using the pattern with the letters is that it avoids using any previously registered threshold value that could affect the prediction of the patient’s risk assessment by providing a biased estimation. In our method, patterns can be recorded in any format.

An ordered sequence of MEWS values that make a pattern is recorded from the patient’s observed data and is labelled as the PMEWS. The pattern array best suits sensor signal discrepancies because of null values and discrete-time readings from different sensors. This step is the second most vital step in raw data transformation.

Table 2. Pattern Formation

	HR	BP	SpO ₂	RR	Temp	Pattern
Raw-vital values	99	–	96	22	–	
MEWS	0	N	0	2	N	0N02N

For sliding windows (sw_{δ}^i) consisting of time slots (TS), threshold data values of patients’ vitals are recorded after processing them into their relevant MEWS score (M), and the resulting pattern is stored in an integer array $P[TS]$. Equation (3) shows an example of the pattern at the time slot (TS). MHR, MBP, MT, MSPO₂, and MRR are the MEWS scores derived from the raw, vital data for heart rate, blood pressure, temperature, oxygen saturation and respiratory rate, respectively.

$$P[TS] = [MHR, MBP, MT, MSPO_2, MRR] \quad (3)$$

The following section describes the pattern-match algorithm to identify similar patterns for missing values of vital signs.

3.2 Pattern-Matching Algorithm

Our main aim for this algorithm is to match the pattern with the missing value with the closest matching pattern that appeared in the sliding window. When the match is found, imputing the missing value parameter with the found match values is performed. For example, in Table 3, the raw data is recorded for the heart rate (HR), temperature (T), blood pressure (BP), oxygen saturation (SpO₂), and respiratory rate (RR) and relevant patterns are created for the time one minute to four minutes. HR is missing at minute four with the pattern ‘NNN20’ and will get pattern-match at minute one ‘0NN20’ as the highest number of symbols matched and imputed with the HR value of 99.

Table 3. Pattern Match

Time (minutes)	HR	T	BP	RR	SpO ₂	Pattern	Symbol match
1	99	–	–	22	96	0NN20	4
2	97	–	141	25	99	0N020	3
3	102	–	–	34	99	1NN30	2
4	*	–	–	28	99	NNN20	

After pre-processing, the first step is to initialize the variables T, L, t, T_{MV}, δ , sw_{δ}^i , and TS. The pattern $P[TS]$ is generated from MEWS for the recorded raw values. Loop to traverse the string pattern $S_P[i]$, from the start of a sliding window, $i = T$ to TS is set. If there is a pattern $S_P[i + 1]$ with no null values, the counter i is incremented for the following index. Otherwise, another inner loop is set for the string pattern array with the index $j = T$ to i to ensure all patterns are considered to match null data values. To ensure a match, two variables, str1 and str2, are set with $S_P[j]$ and $S_P[i + 1]$, respectively. A call to the function count passes these two string variables to check if str1 contains str2. For a matching pattern character case, the count is maintained for the pattern and stored in counter array $CS_P[k]$. The counter array $CS_P[k]$ from $j = T$ to TS is checked for the condition if the number of character matches is highest for the number of vitals observed, as shown in Table 4. Relative patterns and count values are displayed as suggestions for outcomes. Timeslot TS is updated with the sliding window increment time δ for traversing the sliding window loop. This algorithm is an iterative process in the sliding window to ensure all patterns are considered.

The proposed algorithm that accomplishes our pattern matching is shown as follows:

Start

Input

Process window length (L), start time (T), number of vitals to be recorded (T_{MV}), window increment (δ), sliding window time (t), Where:

sliding window $sw_{\delta}^t = (t_{(\delta+0)}^i, t_{(\delta+1)}^i, t_{(\delta+2)}^i, \dots, t_{(\delta+n)}^i)$,

time slot $TS = (t_{(\delta+n)}^i)$

While ($TS \leq L$) **do**

Record values of vitals (HR, Temp, BP, RR, and SpO₂)

// Assign MEWS to vitals, where null assigns "N"

Save $P[TS] = [MHR, MBP, MT, MSPO2, MRR]$

Copy recorded patterns in $S_P[]$

for each pattern $P[TS]$ in $S_P[]$ from $i=T$ to $i \leq TS$ **do**

if $S_P[i+1] \neq 'N'$ **then** $i++$ **_**

else assign:

for (int $j=T$; $j \leq i$; $j++$) **do**

$str1 = S_P[j]$

$str2 = S_P[i+1]$

count ($str1, str2$)

for each count c for pattern in array $CS_P[k]$

from $j = T$ to TS **do**

$CS_P[k] = c$, end for

for each count in $CS_P[k]$ from $j = T$ to TS **do**

If ($CS_P[k]$ is the highest match parameter count) **then**

print relevant pattern $P[TS]$ and count from $CS_P[k]$

Else print no suitable match found

end if, end for

end if, end for

Increment $TS = t + \delta$, end While

function count (string $str1, string str2$)

initialize counter $c = 0$, index $j = 0$

for each i in $str1.length()$ **do**

If

($str2.find(str1[i]) \geq 0$ (AND) $j == str1.find(str1[i])$) **then**

$c++$, $j++$

return c ,

end if, end for

end function, Stop

4 Result and Discussion

To assess the plausibility of our technique, we used the dataset collected from the clinical trial [18]. The trial dataset has the raw values for the vital parameters Temp, BP, SpO₂, RR, HR, and the time for each record, as shown in Table 3. For this study, only two parameters, SpO₂ and HR, were imputed. The dataset was split into additional sets of various record lengths. MI and EM are commonly used to impute missing values; our proposed method is compared against these. Normalization is an essential step in pre-processing our data, as it helps scale the data and make it more suitable for analysis by reducing the biased outcome. Normalization is performed using the MEWS. Starting with the entire collection of values, we removed various percentages of the SpO₂ and HR to test the precision of our technique. The values of the missing parameters in the dataset were then imputed using our pattern-match imputation, EM and MI in the same sliding window. The highest match for parameters is selected for the imputation by our algorithm. For comparison, we calculated the root mean squared error (RMSE) for the imputed value with the starting values from the whole dataset. Table 4 presents results obtained from four different datasets using different imputation methods.

For dataset D1:

- SpO₂: Pattern-match has the lowest RMSE of 0.82, outperforming EM and MI.
- HR: Pattern-match has the lowest RMSE of 3.9, outperforming EM and MI.

For dataset D2:

- SpO₂: Pattern-match has the lowest RMSE of 2.03, outperforming EM and MI.
- HR: Pattern-match has an RMSE of 13.9, higher than EM and MI.

For dataset D3:

- SpO₂: Pattern-match has an RMSE of 2.4, while EM has the lowest.
- HR: Pattern-match has the lowest RMSE of 3.9, followed by EM and MI.

For dataset D4:

- SpO₂: Pattern-match has an RMSE of 2.5, while EM has the lowest.
- HR: Pattern-match has the lowest RMSE of 7.3, followed by EM and MI.

Figures 2 and 3 show the variation of RMSE of two of our used parameters, SpO₂ and heart rate (HR). In the context of missing data, the sliding window approach and the PMEWS have proven very useful as the algorithm demonstrates comparable results. The algorithm considers every pattern in the sliding window before the missing pattern and uses these patterns to find a similar match. Then it slides the window across the process window to cover all the patterns that appeared in that timeframe.

Table 4. Results of the Datasets.

Dataset	Total Records	Parameter observed	Observed Records	Method	RMSE
D1	756	SpO ₂	45	Pattern-Match	0.82
				EM	2.01
				MI	2.52
D1	756	HR	27	Pattern-Match	3.9
				EM	9.6
				MI	10.2
D2	1000	SpO ₂	88	Pattern-Match	2.03
				EM	2.1
				MI	2.94
D2	1000	HR	26	Pattern-Match	13.9
				EM	20.6
				MI	24.3
D3	490	SpO ₂	25	Pattern-Match	2.4
				EM	1.2
				MI	2.04
D3	490	HR	20	Pattern-Match	3.9
				EM	4.3
				MI	7.4
D4	932	SpO ₂	84	Pattern-Match	2.5
				EM	1.39
				MI	2.06
D4	932	HR	28	Pattern-Match	7.3
				EM	7.8
				MI	14.04

The proposed algorithm's accuracy was satisfactory, which suggested that this approach is practical for finding a similar effective match for the missing vital parameter values in medical datasets generated in RPM settings. It was observed that most of the similar patterns in data appeared after the normalization phase of pre-processing of the data. The algorithm can leverage these patterns to find identical matches for the missing vital values, which can improve the accuracy of the prediction. These findings have implications for the researchers, physicians, and practitioners who work with the vital sign data set with the missing values in RPM for patient risk assessment analysis.

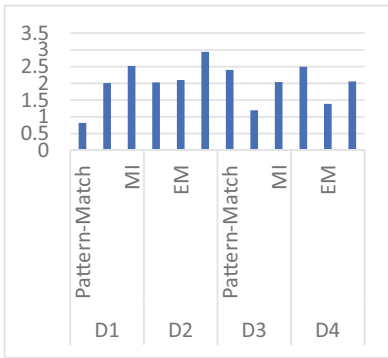


Fig. 2. Root Mean Square Error for SpO₂

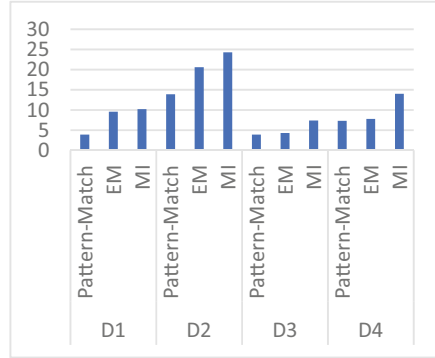


Fig. 3. Root Mean Square Error for HR

5 Conclusion

This study presents a method for pattern-matching the missing values in the RPM vital sign dataset using the PMEWS and the sliding window. This method offers the identification of similar patterns in the data, and using the PMEWS, this technique is further improved by leveraging these patterns. Compared to other approaches, this technique considers the temporal context of the missing values, which can be particularly important in the medical domain where the timings of the measurements can have critical implications for patient care. This method also has the potential to revolutionize how continuous RPM data is processed and analyzed, leading to more efficient healthcare delivery. Future research could explore other pre-processing techniques to improve the pattern match further. Overall, the proposed method can improve the accuracy of the medical data analysis for decision-making, specifically for the missing vitals values.

References

1. Posthuma, L., et al.: Remote wireless vital signs monitoring on the ward for early detection of deteriorating patients: a case series. *Int. J. Nurs. Stud.* **104**, 103515 (2020)
2. Jacob, S., et al.: AI and IoT-enabled smart exoskeleton system for rehabilitation of paralyzed people in connected communities. *IEEE Access* **9**, 80340–80350 (2021)
3. Areia, C., et al.: The impact of wearable continuous vital sign monitoring on deterioration detection and clinical outcomes in hospitalised patients: a systematic review and meta-analysis. *Crit. Care* **25**(1), 1–17 (2021)

4. Allami, R., et al.: A count data model for heart rate variability forecasting and premature ventricular contraction detection. *SIViP* **11**(8), 1427–1435 (2017)
5. Arora, T., Balasubramanian, V., Mai, S.: Prioritization of clinical alarms using semantic features of vital signs in remote patient monitoring. In: *Australasian Computer Science Week 2022*, pp. 242–245 (2022)
6. Balasubramanian, V.: Critical time parameters for evaluation of body area wireless sensor networks in a healthcare monitoring application. In: *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)* (2014)
7. Mitsunaga, T., et al.: Comparison of the national early warning score (NEWS) and the modified early warning score (MEWS) for predicting admission and in-hospital mortality in elderly patients in the pre-hospital setting and in the emergency department. *PeerJ* **7**, e6947 (2019)
8. Petersen, J.A.: Early warning score. *Dan Med J.* **65**(2) (2018)
9. Johnson, S., Shenoy, A.J.J.C.M.T.: Modified early warning score: does it warn enough. **2**(2), 14 (2017)
10. Subbe, C.P., et al.: Validation of a modified early warning score in medical admissions. *QJM: Int. J. Med.* **94**(10), 521–526 (2001)
11. da Costa, C.A., et al.: Internet of health things: toward intelligent vital signs monitoring in hospital wards. *Artif. Intell. Med.* **89**, 61–69 (2018)
12. van Rossum, M.C., et al.: Missing data imputation techniques for wireless continuous vital signs monitoring. *J. Clin. Monit. Comput.* (2023). <https://doi.org/10.1007/s10877-023-00975-w>
13. Little, R.J., Rubin, D.B.: *Statistical Analysis with Missing Data*. vol. 793, Wiley, Hoboken (2019)
14. Azimi, I., et al.: Missing data resilient decision-making for healthcare IoT through personalization: a case study on maternal health. *Futur. Gener. Comput. Syst.* **96**, 297–308 (2019)
15. Baraldi, A.N., Enders, C.K.: An introduction to modern missing data analyses. *J. Sch. Psychol.* **48**(1), 5–37 (2010)
16. Tkachenko, R., et al.: An approach towards increasing prediction accuracy for the recovery of missing IoT data based on the GRNN-SGTM ensemble. *Sensors* **20**(9), 2625 (2020)
17. Weenk, M., et al.: Continuous monitoring of vital signs using wearable devices on the general ward: pilot study. *JMIR Mhealth Uhealth* **5**(7), e7208 (2017)
18. Balasubramanian, V., et al.: A secured real-time IoMT application for monitoring isolated COVID-19 patients using edge computing. In: *2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, IEEE (2021)
19. Lang, K.M., Little, T.D.: Principled missing data treatments. *Prev. Sci.* **19**(3), 284–294 (2018)
20. Austin, P.C., et al.: Missing data in clinical research: a tutorial on multiple imputation. *Can. J. Cardiol.* **37**(9), 1322–1331 (2021)
21. Rosato, R., et al.: Missing data in longitudinal studies: comparison of multiple imputation methods in a real clinical setting. *J. Eval. Clin. Pract.* **27**(1), 34–41 (2021)

22. Willis, D.E., et al.: COVID-19 vaccine hesitancy: race/ethnicity, trust, and fear. *Clin. Transl. Sci.* **14**(6), 2200–2207 (2021)
23. Stekhoven, D.J., Bühlmann, P.: MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**(1), 112–118 (2011)
24. Nijman, S.W.J., et al.: Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. *J. Clin. Epidemiol.* **142**, 218–229 (2022)
25. Mostafa, S.M., et al.: CBRG: a novel algorithm for handling missing data using bayesian ridge regression and feature selection based on gain ratio. *IEEE Access* **8**, 216969–216985 (2020)
26. Wafaa Mustafa Hameed, N.A.A.: Comparison of seventeen missing value imputation techniques. *J. Hunan Univ. Natural Sci.* **49**(7) (2022)
27. Banaee, H., Ahmed, M.U., Loutfi, A.: Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors* **13**(12), 17472–17500 (2013)
28. Mukherjee, A., Pal, A., Misra, P.: Data analytics in ubiquitous sensor-based health information systems. In: 2012 Sixth International Conference on Next Generation Mobile Applications, Services and Technologies (2012)
29. Mao, Y., et al.: An integrated data mining approach to real-time clinical monitoring and deterioration warning. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2012)
30. Bihri, H., Hsaini, S., Nejjari, R., Azzouzi, S., Charaf, M.E.H.: Missing data analysis in the healthcare field: COVID-19 case study. In: Ben Ahmed, M., Teodorescu, H.N.L., Mazri, T., Subashini, P., Boudhir, A.A. (eds.) *Networking, Intelligent Systems and Security. Smart Innovation, Systems and Technologies*, vol. 237, pp 873–884. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-3637-0_61
31. Kök, İ., Özdemir, S.: DeepMDP: a novel deep-learning-based missing data prediction protocol for IoT. *IEEE Internet Things J.* **8**(1), 232–243 (2021)
32. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **39**(1), 1–22 (1977)
33. Agbo, B., Qin, Y., Hill, R.: Best fit missing value imputation (BFMVI) algorithm for incomplete data in the internet of things. In: *IoTBDs* (2020)
34. Ahn, J.H., et al.: Predictive powers of the Modified early warning score and the national early warning score in general ward patients who activated the medical emergency team. *PLoS ONE* **15**(5), e0233078 (2020)
35. Sow, D.M., Turaga, D.S., Schmidt, J.M.: Mining of sensor data in healthcare: a survey. In: *Managing and Mining Sensor Data*, pp. 459–504 (2013). https://doi.org/10.1007/978-1-4614-6309-2_14
36. Kia, A., et al.: MEWS++: enhancing the prediction of clinical deterioration in admitted patients through a machine learning model. *J. Clin. Med.* **9**(2), 343 (2020)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

