Assessing the impact of missing value mechanisms on anomaly detection in healthcare wearable data

Afonso M. S. Lima[1]

[1]Institute of Mathematical and Computer Science (ICMC)
University of São Paulo (USP)
São Carlos – SP – Brazil

afonso.msl@proton.me

Abstract. *Remote health monitoring using wearable devices has transformed healthcare by enabling continuous observation and early intervention. However, these systems frequently suffer from missing data, which can introduce bias and impair clinical decision-making. The uncertainty surrounding the cause of missing values further complicates data analysis. This paper investigates the impact of different missing data mechanisms (i.e., MCAR, MAR, and MNAR) in the healthcare wearable data anomaly detection task. Using heart rate and step count data from patients with respiratory illnesses, we assess the performance of an anomaly detection method under varying missingness conditions. Our findings demonstrate that more complex mechanisms (MAR and MNAR) significantly degrade detection performance, even at low missing rates, highlighting the importance of developing robust imputation strategies tailored to the nature of missingness.*

1. Introduction

Remote health monitoring using wearable devices, such as smartwatches, holds great promise in healthcare by allowing for continuous monitoring of individuals, facilitating early detection, preventive care, and timely interventions [Canali et al. 2022, Getzen et al. 2023]. However, this approach is significantly affected by missing values, as it is common that continuously generated wearable data presents a percentage of missingness [Psychogyios et al. 2023], which can yield biased findings and mistaken treatment decisions [Isgut et al. 2022].

Furthermore, there is uncertainty regarding the cause of the missing values, i.e., the missing mechanism, as they can arise not only from a sensor failure at an arbitrary time but also from their relation to other measurements and the missing value itself [Ren et al. 2023]. Making the correct assumption about the missing mechanism helps choose the best methods for handling missing values for a given domain [Emmanuel et al. 2021, Ren et al. 2023, Lima and Sousa 2024]

This paper aims to assess the impact of missing values in healthcare wearable data anomaly detection by considering the three main missing data mechanisms: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR) [Emmanuel et al. 2021, Ren et al. 2023]. We conducted experiments

Acknowledgments This research was supported by CAPES (Brazilian Coordination for Improvement of Higher Level Personnel) and CNPq (Brazilian National Council for Supporting Research).

using heart rate and step data from 27 patients with respiratory diseases to detect pre-symptomatic cases of positive COVID-19 [Mishra et al. 2020]. Our findings demonstrated a minimum decrease of 10% in almost all metrics, even with a low missing rate of 5%. Furthermore, it indicates that more complex mechanisms, i.e., MAR and MNAR, further degrade detection performance when compared to MCAR, even at low missing rates of 5% and 10%. These results highlight the importance of developing robust imputation strategies tailored to the nature of missingness.

2. Background

Wearable devices are becoming increasingly common in the healthcare field, providing functionalities such as monitoring, screening, detection, and prediction [Canali et al. 2022]. For instance, Mishra *et al.* (2020) analyzed physiological and activity data from 32 individuals infected with COVID-19, and, using retrospective smartwatch data, they show that 63% of the COVID-19 cases could have been detected before symptom onset in real time.

However, there is a growing recognition that ubiquitous missing data in healthcare, even when analyzed using powerful statistical and machine learning algorithms, can yield biased findings and unfair treatment decisions [Lin et al. 2020]. For instance, Getzen *et al.* (2023) assessed missing data impact on electronic health records (i.e., the digital version of a patient's paper medical chart) and found that missing data have greater negative impact on the performance of disease prediction models in groups that tend to have less access to healthcare, or seek less healthcare.

For wearable data, in addition to random sensor faults, these missing values may occur due to inconsistent collection periods (e.g., varying compliance and wearing behavior) [Mishra et al. 2020, Lin et al. 2020]. Thus, different scenarios of missing values can correspond to distinct missing data mechanisms. Table 1 summarizes each mechanism along with a practical healthcare example. Correctly identifying the missing mechanism may lead to better treatment of missing values [Emmanuel et al. 2021, Ren et al. 2023]. For instance, with a MCAR assumption, simpler methods like mean imputation can be used since missing values are randomly distributed. For MAR, correlations between missing and non-missing attributes may help estimate missing values. In the case of MNAR, understanding the patterns in missing value distributions can guide more effective imputation, such as training models on specific value ranges.

Table 1. Missing mechanisms summary and healthcare example

Missing Mechanism	Definition	Example
Missing Completely At Random (MCAR)	The missing cause is unrelated to other observed or unobserved values.	The wearable device failed, and the value was not generated.
Missing At Random (MAR)	The missing cause depends on other observed attributes.	Patient removed their wearable sensor in sleep hours ("heartrate" depends on "datetime").
Missing Not at Random (MNAR)	The missing cause is related to specific information about the own missing value that is not present in the dataset.	The wearable sensor is not sensitive enough to low values, setting it to zero. ("heartrate" depends on itself).

3. Method

To assess and measure the impact of missing values presence in healthcare wearable data mining task, we simulated missing values, considering each missing mechanics characteristics, in a healthcare dataset obtained from wearable devices of COVID-19 positive patients, presented in the work of Mishra *et al.* (2020). Figure 1 summarizes the missing values assessment process.

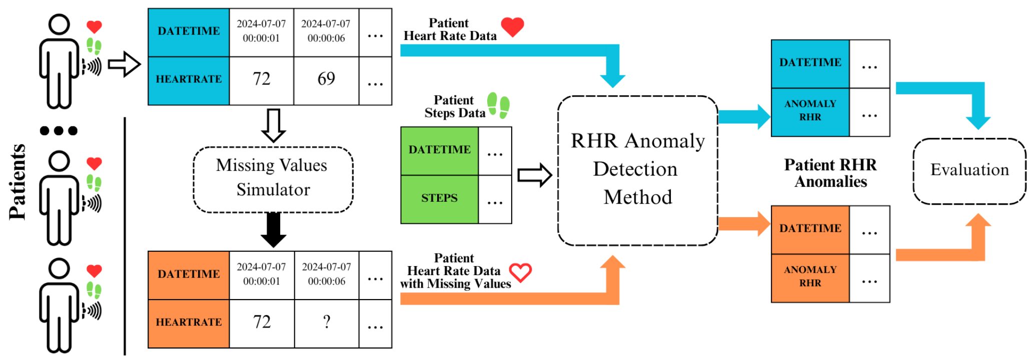

Figure 1. Missing values assessment process

These wearable heart rate data consists of a timestamp (e.g., “2024-07-07 00:00:01”, “2024-07-07 00:00:06”, ...) and a respective heart rate measurement (e.g., “72”, “69”, ...) from 27 patients with reproducible results from the original paper. The timestamp interval between each heart rate measurement is irregular for all patients. Most of them were measured between a few seconds, but hours-long intervals were common, representing the patient’s wearing behavior (i.e., not wearing the device). These characteristics are expected in real-time wearable data processing scenarios. This heart rate data, along with the patient’s steps, are used to calculate the Resting Heart Rate (RHR) metric. This metric refers to heart rate measurements taken when the patient is at rest, i.e, when the step value is equal to zero.

To artifically introduce missingness, we used the mdatagen Python library [Mangussi et al. 2024] developed upon the work of Santos *et al.* (2019). It contains procedures to simulate the MCAR, MAR, and MNAR missing mechanisms. For our experiment, as specific missing mechanism settings, we chose the “random” strategy for MCAR, where the missing location are randomly selected in the feature “heartrate”. For MAR, we selected the “lowest” strategy looking up to the hour part of “datetime” feature, i.e., we are choosing the early hours of the day to select the “heartrate” value location. The reasoning for this decision is that during the early hours of the day, when the patient is likely resting, there are the most RHR measurements, which may significantly impact the anomaly detection task. For MNAR, a threshold must be defined to select value locations from a feature, ranging from 0 to 1. A threshold of 0 picks the lowest values, while 1 selects the highest. In our experiment, we set the threshold to 0 to focus on the lowest heart rate values, as low readings may occur when the patient is resting, affecting the RHR results.

For each mechanism, we considered the following missingness rates: 5%, 10%, 15%, 20%, and 25% for each patient. We opted for low missing rates to measure the

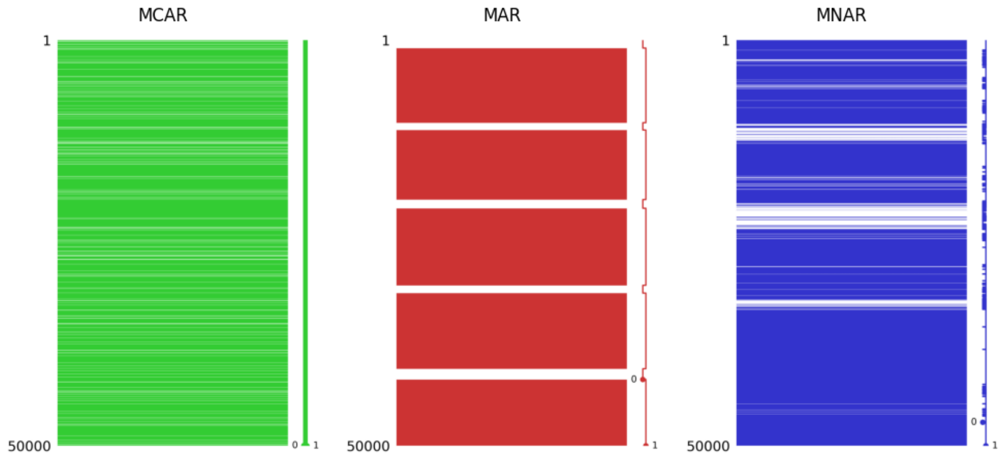

Figure 2. Missing values density with 10% missing rate for each mechanism regarding the first 50000 heart rate measurements from patient “AAXAA7Z”

impact on less extreme scenarios. For MCAR and MNAR, we simulated missing values 10 times for each missing rate and each patient. This is necessary to avoid bias, as MCAR randomly selects the locations of missing values. In contrast, MNAR randomly selects them when the minimum heart rate values are repeated and the missing rate is exceeded. For both, we considered the mean result of the 10 times.

As an example, Figure 2 presents, for each mechanism, the missing values density in the first 50000 heart rate measurements (approximately, five days) from a patient labeled as “AAXAA7Z”, with 10% missing rate. Blank lines indicate a missing value location in the dataset. The MCAR distribution of blank lines does not show areas of density for missingness, as the missing values are randomly distributed. In contrast, MAR and MNAR exhibit a higher concentration of missing values in specific periods, i.e., the earlier hours of the day. This is expected for MAR configuration, as our setting defines correlation between missing values with the earlier hour of the day. For MNAR, most of the dense regions of missing values are also present close to the earlier hours, as heart rate measurement in sleep hours is lower.

Next, we applied the RHR Offline Anomaly Detection method [Mishra et al. 2020]. Using this method, elevated RHR time intervals based on the standardized residuals were detected and identified. The RHR anomalies obtained from the COVID-19 wearables dataset were validated alongside continuous patient monitoring. Thus, we define the original set of anomalies as our ground-truth.

For evaluation, the original set of RHR anomalies in these patients was compared with those detected using heart rate data with simulated missing values, allowing for quantitative measurement of the results. We considered an anomaly detected at the same timestamp as a true positive, making it possible to calculate the main performance metrics, i.e., accuracy, precision, recall and F1 Score. For reproducibility, the patients’ heart rate measurement datasets with simulated missing values and the code used in this experiment are available in a public repository on GitHub[1].

[1] Available at: https://github.com/afonsoMatheus/Wearables-Offline-Analysis

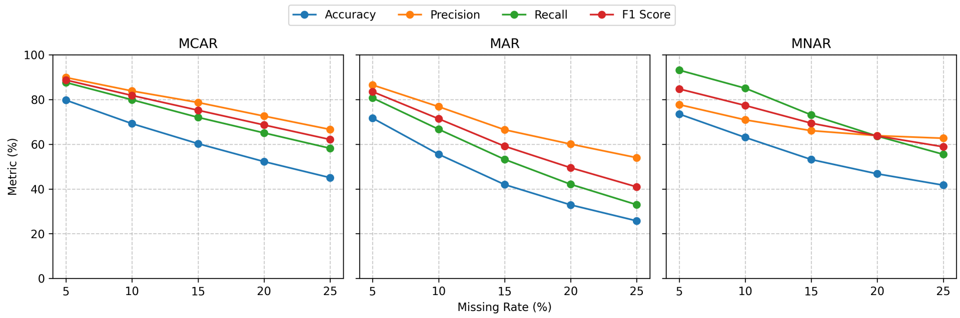

Figure 3. Performance results of the 27 patients for each missing rate and missing mechanism

4. Results

Figure 3 shows the performance results considering all patients. Overall, the presence of missing values negatively affected all metrics, leading to poorer anomaly detection outcomes, even at the lowest missing rate of 5%. Among the metrics, accuracy exhibited the most substantial decline. When examining each mechanism, the MAR mechanism demonstrated the greatest performance drop across all missing rates, suggesting that missing values during the early hours of the day have a significant impact on the RHR metric. In contrast, the MNAR mechanism, with the lowest heart rate configuration, affected overall performance to a lesser extent than MAR, although its precision was the lowest among all mechanisms.

To better understand the performance impact of different mechanisms, Figure 4 illustrates the relationship between recall and precision results. For the MCAR and MAR mechanisms, both precision and recall remained similar across varying missing rates. However, the MAR mechanism showed a greater decline in both metrics compared to MCAR; for instance, at a 15% missing rate in MAR, the results were worse than those seen at a 25% missing rate in MCAR. In contrast, MNAR low missing rates resulted in

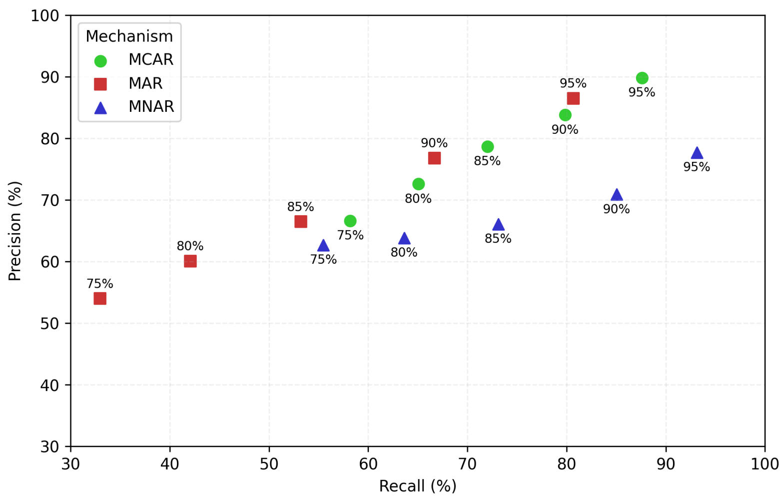

Figure 4. Recall X Precision results of the 27 patients for each missing rate and missing mechanism. The labels in each dot represent the percentage of the remaining original dataset (i.e., 100 - Missing Rate)

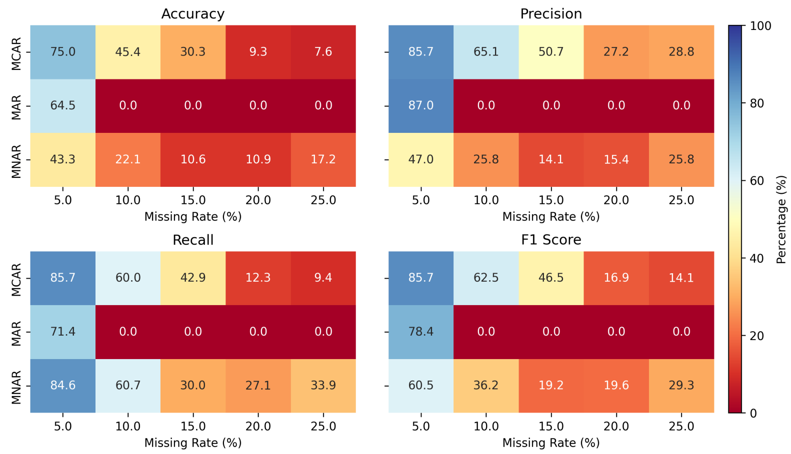

Figure 5. Heatmap with the performance results of patient labeled “AIFDJZB” for each missing rate and missing mechanism

better recall than other mechanisms, although it exhibited the poorest precision overall. This suggests that the MNAR mechanism does not significantly impede the identification of the correct anomaly timestamp, as shown by the high recall scores. However, the low precision indicates an overestimation of the anomaly period, leading to more false positives. This may affect the detection of patients’ anomaly periods, leading to inaccurate assessments of when the RHR significantly deviates from normal levels, potentially hindering early disease detection.

Furthermore, while the overall results indicate that the performance is unbiased, individual patients may be affected differently by the configuration of the missing mechanism. For example, Figure 5 shows the performance heatmap of a patient identified as “AIFDJZB.” In this case, starting from a 10% missing rate under the MAR mechanism, all metrics returned 0% results, meaning that no anomalies were detected. Upon further analysis, we found that the MAR configuration, which caused values to be missing during the early hours of the day, led to the absence of critical heart rate data necessary for identifying patients’ anomalies, even with only a 10% missing rate. This case highlights the importance of each heart rate measurement and why it cannot be overlooked, motivating the use of more sophisticated imputation methods tailored to the MAR mechanism.

5. Conclusion and Future Work

This paper evaluated the impact of missing values in healthcare wearable data anomaly detection by considering three main missing data mechanisms: MCAR, MAR and MNAR. The results indicate that each missing mechanism influences the anomaly detection task differently. Specifically, the more complex mechanisms, MAR and MNAR, significantly impacted the target performance metrics, even at low missing data rates. This finding highlights the need for more robust methods for handling missing values that can effectively address the characteristics of these mechanisms.

In future work, we plan to further assess the impact of missing values in health monitoring by using multiple wearable datasets that involve different mining tasks. Additionally, we aim to apply more evaluation approaches to conduct a deeper analysis of the results and better understand how missing values affect the outcomes.

References

Canali, S., Schiaffonati, V., and Aliverti, A. (2022). Challenges and recommendations for wearable devices in digital health: Data quality, interoperability, health equity, fairness. *PLOS Digital Health*, 1(10):e0000104.

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., and Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big data*, 8:1–37.

Getzen, E., Ungar, L., Mowery, D., Jiang, X., and Long, Q. (2023). Mining for equitable health: Assessing the impact of missing data in electronic health records. *Journal of biomedical informatics*, 139:104269.

Isgut, M., Gloster, L., Choi, K., Venugopalan, J., and Wang, M. D. (2022). Systematic review of advanced ai methods for improving healthcare data quality in post covid-19 era. *IEEE Reviews in Biomedical Engineering*, 16:53–69.

Lima, A. S. and Sousa, E. (2024). Handling missing values in data streams: An overview. In *Anais do XXXIX Simpósio Brasileiro de Bancos de Dados*, pages 750–756, Florianópolis, SC, Brasil. SBC.

Lin, S., Wu, X., Martinez, G., and Chawla, N. V. (2020). Filling missing values on wearable-sensory time series data. In *Proceedings of the 2020 SIAM International Conference on Data Mining*, pages 46–54. SIAM.

Mangussi, A. D., Santos, M. S., Lopes, F. L., Pereira, R. C., Lorena, A. C., and Abreu, P. H. (2024). mdatagen: A python library for generating missing data. https://arthurmangussi.github.io/pymdatagen/.

Mishra, T., Wang, M., Metwally, A. A., Bogu, G. K., Brooks, A. W., Bahmani, A., Alavi, A., Celli, A., Higgs, E., Dagan-Rosenfeld, O., et al. (2020). Pre-symptomatic detection of covid-19 from smartwatch data. *Nature biomedical engineering*, 4(12):1208–1220.

Psychogyios, K., Ilias, L., Ntanos, C., and Askounis, D. (2023). Missing value imputation methods for electronic health records. *IEEE Access*, 11:21562–21574.

Ren, L., Wang, T., Seklouli, A. S., Zhang, H., and Bouras, A. (2023). A review on missing values for main challenges and methods. *Information Systems*, page 102268.

Santos, M. S., Pereira, R. C., Costa, A. F., Soares, J. P., Santos, J., and Abreu, P. H. (2019). Generating synthetic missing data: A review by missing mechanism. *IEEE Access*, 7:11651–11667.