Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 7 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i07.61659

PAPER

Decision Framework Focused on Missing-Data Alarms in Healthcare

Elma Zanaj,
Lorena Balliu (✉),
Anita Xhemali,
Gledis Basha

Polytechnic University of Tirana, Tirana, Albania

lorena.balliu@fti.edu.al

ABSTRACT

In healthcare, missing data is common. Missing information can arise for various reasons, such as timing and synchronization issues, incorrect reporting, and connection loss or incorrect data-flow configuration. Time plays a crucial role in longitudinal clinical data. Many studies have focused on minimizing the root mean squared error (RMSE) of imputed missing data. However, in clinical practice, the primary goal is to make safe decisions by minimizing the risk of missing key events. The goal is not to reconstruct accurate missing data but to determine if the existing data is sufficient to support the detection of critical events. The predictive model combines a probabilistic event definition, empirical risk, calibrated thresholds for risk-targeted threshold calibration, a lightweight counterfactual glucose sampler, and the expected value of sample information (EVSI). The EVSI policy does not always request a confirmatory measurement. It is used only when the expected reduction in decision burden is higher than the additional delay. Results show an improved balance of risk and burden when compared to both non-query scenarios and simpler baseline querying strategies. The contribution of this work is not a new state-of-the-art data imputation system but a clinically interpretable decision layer for alarm management under missing data scenarios.

KEYWORDS

missing data, healthcare time series, risk-targeted calibration, expected value of sample information (EVSI), decision support, glucose alarms

1 INTRODUCTION

The most prominent studies on imputation for time series include different types of models. Some of the most well-known are attention-based models such as Self-Attention-based Imputation for Time Series (SAITS), diffusion-based models such as Conditional Score-based Diffusion Models for Imputation (CSDI), and clinically informed architectures such as Conditional Self-Attention Imputation for Healthcare Time Series (CSAI) [1]–[5]. Based on the literature, there is a need for

Zanaj, E., Balliu, L., Xhemali, A., Basha, G. (2026). Decision Framework Focused on Missing-Data Alarms in Healthcare. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(7), pp. 207–223. https://doi.org/10.3991/ijoe.v22i07.61659

Article submitted 2026-03-23. Revision uploaded 2026-05-06. Final acceptance 2026-05-13.

more clinically appropriate evaluation methods [6]–[10]. Good imputation accuracy does not always lead to improved downstream predictions [8], [9]. Reconstruction of missing data with high quality does not necessarily result in the highest decision value. A system may produce acceptable imputation information, but it does not always contribute, by default, to a better alarm decision for clinical events.

Based on the above, we focus our study not on highly accurate reconstruction but on decision sufficiency. The main question raised by our study is whether the available information is enough to support a safe decision or if it is worth requesting additional information. Moreover, when is it worth accepting additional delays and burdens to obtain extra information? The core idea is that missing healthcare data should not be considered a simple reconstruction problem.

We propose a decision-focused framework for short-term glucose alarms under missing data, designed to support safer clinical decisions under incomplete information. Our framework is based on a modular design. It acknowledges that imputation and downstream prediction do not need to be strictly matched [9], [10]. The system has three main components. The first is a base predictor that operates under partial information from an observed trajectory. The second is the calibration layer. The output of the base predictor is used to calibrate the threshold for the risk of missing critical events. The third is the Value-Aware Escalation Policy. It determines whether to abstain from making a prediction or to wait for future confirmation data. In situations of missing values, it can be more efficient to reduce the number of missed critical events than to not query at all or to query randomly.

The EVSI-based query policy is designed to trigger queries only when the expected reduction in downstream decision penalties is sufficiently large, thus distinguishing between high-value and low-value escalations. We evaluate two EVSI-based policies (Expected Value of Sample Information Positive policy (EVSI_POS) and Expected Value of Sample Information Budget-Binding policy (EVSI_BIND)), which are described in more detail in Section 3. In the representative 30-minute delay setting, EVSI_POS reduces strict false-negative rates compared with the no-query baseline in Missing Completely at Random (MCAR) and Missing At Random (MAR) scenarios, while showing limited additional benefit under Missing Not at Random (MNAR)_G250. The improvement is reflected in lower strict false negative rates, at the cost of a modest increase in decision burden. EVSI_BIND provides a more conservative alternative. It trades a higher burden for slightly stronger risk reduction.

The main contributions of this study relate to a framework focused on decision-making when dealing with missing glucose data for alarm generation. It includes an empirical risk-target calibration layer that identifies safer thresholds for the system; a delayed measurement budgeted policy based on the EVSI; and experimental evaluation of a framework comparison study including MCAR, MAR, and MNAR missing scenarios [11], along with ablation studies.

This paper is organized into six main sections. After the introduction, Section 2 summarizes existing works in the imputation of time series, downstream evaluation, and active decision-making, connecting within the context of our study. Section 3 describes the proposed methodology. It explains the modular structure of the framework, along with mathematical formulation, alert calibration layer, and EVSI-based policy. Section 4 discusses the main results, including baseline comparisons, robustness analysis, sensitivity analysis, runtime overhead, and ablation studies. Section 5 concludes the study. Finally, Section 6 outlines future work.

2 RELATED WORK

Time series imputation has seen significant growth in recent years. Particular advantages have been provided by the use of deep learning-based models. Bidirectional Recurrent Imputation for Time Series (BRITS) is a successful imputation method [3]. It uses bidirectional recurrent dynamics for time-series imputation. SAITS and CSDI utilize self-attention and diffusion-based mechanisms to capture both temporal relationships and cross-feature relationships [1]–[4]. CSAI is another approach that highlights the importance of imputing missing data. It operates in a way that preserves the structural characteristics of real clinical data [5]. In addition to advanced imputation models, special attention has also been given to the treatment of missing values in traditional statistical models. The authors [12] proposed an LR model in conjunction with an appropriate penalty for high-dimensional data and implemented multiple imputations to both improve classification accuracy and identify the most relevant features.

Good reconstruction does not necessarily lead to a good performance for the downstream predictions [6]–[10]. Evaluations reported in TSI-Bench suggest that root mean squared error (RMSE) and mean absolute error (MAE) alone are insufficient for assessing imputation quality. How imputation supports the final decision is equally important [7]–[10]. Furthermore, it is necessary to determine whether the available information adequately supports the decision-making process. Another related area of research is that of active information acquisition. Research on Risk-Averse Active Sensing has explored whether the cost and time required to acquire additional information are justified. These studies also examine how such costs and delays affect the final decision [13]–[15].

The Expected Value of Sample Information is used to determine if it is worth acquiring additional information before making a decision [16]–[18]. These studies are based on the assumption that, often, computationally expensive methods are not worth it in cases of real-time monitoring. This work presents a unified decision-focused framework for glucose alarms in the presence of missing values over a short period. The novelty of this work lies in combining probabilistic alarm prediction with risk-targeted calibration and EVSI-based delayed querying under budget and delay constraints. The empirically calibrated thresholds are constructed from an expectation of risk. We use EVSI for query delays under subject budget and time constraints. Our study aims not only to improve the quality of reconstructed data but also to enhance the safety of decisions made under missing data.

Recent IoT-based healthcare studies provide important background for real-time monitoring, mobile diabetes management, and the broader engineering foundations of IoT systems in healthcare; however, they mainly focus on data collection, application design, or IoT infrastructure rather than decision-making when measurements are missing or delayed.

[19] presented a Smart Healthcare Monitoring System that uses the Internet of Things (IoT) to collect patient data to allow for real-time observation. However, they did not discuss what happens with missing measurements when an alarm must still be generated. [20] developed a mobile application to monitor and manage Type 2 Diabetes Mellitus in real time. Their findings demonstrate relevance to the diabetes-monitoring context of our study. [21] defined IoT engineering as a

multidisciplinary emerging area that encompasses various technical aspects such as sensing, networking, big data analytics, cloud services, security, and AI, which can be applied to fields like healthcare.

Our prior studies [22]–[24] have also focused on the healthcare area. This work builds on that and extends this direction by focusing not only on providing estimates of missing data but also on determining if there is enough information to make a confident decision.

3 METHODOLOGY

The decision-focused framework for missing glucose data aims to reduce missed critical events under a limited operational burden (query and delay). At each time step t, the model receives a partially observed multivariate record with data on glucose levels, heart rate, the number of steps taken, total caloric intake, carbohydrate intake, bolus insulin administered, and basal rates of insulin. The decision is chosen from one of three possible alternatives: ALARM, NO ALARM, and ABSTAIN. A query represents a request for a delayed confirmatory glucose measurement. Temporary abstention allows for updating the decision once the delayed values are received.

3.1 Framework

This framework is built around one question: Does the benefit of delayed confirmation outweigh its cost? To provide an answer, the framework estimates the EVSI based on an approximation of expected downstream costs of not having this information. It compares this to expected downstream costs in which the missing glucose value is replaced with samples and added to expected additional burdens from sampling, abstaining, and delayed testing. A query is issued only when there is sufficient statistical evidence that it is worthwhile to do so. The EVSI_POS policy queries only when the expected value is positive. EVSI_BIND is a more conservative variant; it respects a strict query budget.

Figure 1 provides an overview of the layers and sequential steps of the proposed framework.

The current decision is temporarily changed to ABSTAIN until the delayed confirmatory measurement becomes available where a query is issued. The abstained decision is resolved after a fixed delay using the delayed confirmatory glucose measurement. Logistic regression (LR) [25] is used in the main experiment due to its stability and transparency. To assess the robustness of the study, it is replaced with eXtreme Gradient Boosting (XGB or XGBoost) [26]. In this way, it is possible to assess whether the performance gain comes from the EVSI policy rather than from a specific predictor. The robustness is evaluated under multiple missing scenarios. The scenarios evaluated include MCAR, MAR, and MNAR. MCAR10, MCAR30, and MCAR50 are used to represent missing rates of 10%, 30%, and 50%, respectively. MAR-TA refers to the time of day and activity. MNAR-G250 refers to the case when glucose levels exceed 250 mg/dL.

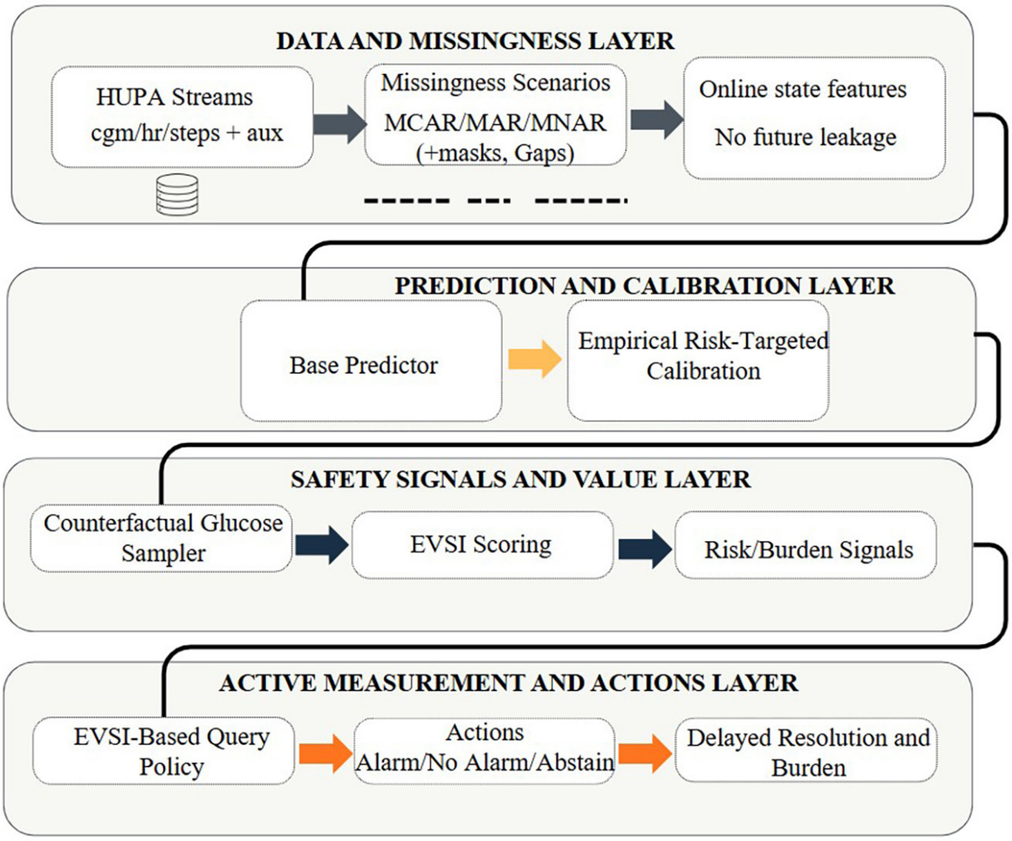

Fig. 1. Layers and steps of the framework

3.2 Mathematical formulation

These six mathematical components represent the decision-making process of the system. As such, equations (1)–(3) provide definitions of each component that will be used as input into the decision layer. These definitions are provided in a general form, consistent with traditional supervised alarm prediction literature. The primary methodological contributions occur in (4), which represents the risk-targeted calibration of the model parameters, and (5)–(6), which describe how the framework uses the EVSI to determine whether a delayed query is necessary based on whether the incomplete evidence is sufficient for a safe alarm decision.

The hyperglycemia alarm is defined as in (1). It is decision-oriented and therefore directly linked to clinical decision-making within a given timeframe. Each of these decisions is based only on information available up to time *t*.

$$y_t = \begin{cases} 1, \text{ if } g_{t+i}^{\text{true}} \geq T \ \forall \ i = 0, \ldots, m-1 \\ 0, \text{ else} \end{cases} \tag{1}$$

The space of action is defined in (2). The system can produce an alarm, not produce an alarm, or abstain. Abstain is used temporarily when there is delayed confirmation, which occurs under both budgetary and delay constraints.

$$a_t \in \{\text{ALARM, NOALARM, ABSTAIN}\} \tag{2}$$

In contrast to some other developments that improve model-agnostic clinical decision-making tools [27], our predictive model does not rely on the type of algorithm being used. The predicted probability (3) is an estimate of the chance of a critical event. Some researchers treat missing data or gaps as “noise,” but we treat them as indicators of possibly useful information. Both the Calibration Layer and the Query Policy are model-agnostic; they do not depend upon the specific predictive model.

$$P_t = P(y_t = 1 \mid \varphi_t) \tag{3}$$

As shown in equations (1), (2), and (3), the standard event, action, and predictive probabilities are used as inputs into the EVSI decision layer. While these components represent the main probabilistic inputs required by the EVSI decision layer, they do not represent the main contributions. The main contribution occurs later in the framework, where it determines if the existing incomplete evidence is sufficient for making a safe alarm decision given the constraints related to missing information, delay, and operational burden.

Threshold calibration τ (4) is the first empirical risk-targeted step to reduce missed events. It is based on empirical quantiles of predicted probabilities for calibration examples that have been positively labeled. CAL is the Calibration set. It denotes the calibration split (subject-disjoint from train/test), used only to select τ via quantiles over positive-labeled examples.

$$\tau = \text{Quantile}_{(r_0)}\left\{p_t : y_t = 1, t \in \text{CAL}\right\} \tag{4}$$

The EVSI component of the proposed framework represents the core decision mechanism. In classical value-of-information analysis, EVSI is used to determine whether collecting additional information is expected to improve a downstream decision. In this work, EVSI is adapted to a real-time clinical alarm context involving missing glucose values, explicit delay, query burden, abstention burden, and both false-negative and false-positive consequences. Therefore, the query decision is not based solely on predictive uncertainty about glucose levels. Instead, it is based on the expected value of delayed information for improving alarm safety.

(5) To estimate the EVSI [28], [29], we use a Monte Carlo simulation [30] to evaluate the utility of collecting additional data. Here, we apply EVSI in real-time, in a subject-specific manner, to make the decision to issue a query for the missing value of interest. We include overheads (delays and operating costs) to provide a more practical application. We develop a relatively simple and efficient model to approximate EVSI in real time. We use Bayesian Ridge regression for the mean glucose, conditioned on the context and locally varying noise modeled by KNN. Samples that are generated using this approach maintain local heterogeneity.

$$\text{EVSI}(t) \approx C_{0(t)} - \left(\left(\frac{1}{K}\right)\sum_{k=1}^{K} C_1(g_t^k, t) + \text{Overhead}\right) \tag{5}$$

Where:

$C_{0(t)}$ – is the expected future burden without knowledge of the missing glucose value.

$C_1(g_t^k, t)$ – represents the cost associated with replacing the missing glucose value with sampled plausible values.

g_t^k + Overhead – includes all additional burdens imposed by the query, abstention, and delay.

The mean total burden in (6) includes the query burden, the abstention burden, the waiting time burden, the burden of unresolved cases, and the penalty of normalized errors. Burden is the clinical-operational cost of managing missing data (measurement/verification, delays, and FN/FP consequences). The weights are relative.

$$\begin{aligned}\text{Burden} = {} & c_Q\text{*QueryRate} + c_A\text{*AbstainRate} + c_T\text{*(MeanTTR*ResolvedRate)} \\ & + c_U\text{*UnresolvedRate} + (c_{FN}\text{*FN} + c_{FP}\text{*FP})/n\end{aligned} \tag{6}$$

Where:
c_Q – The operational burden related to the query request is set to 1.0.
c_A – The operational burden related to the temporary abstain step is set to 0.25.
c_T – The waiting time burden per minute (time-to-resolution) is set to 0.01/minute.
c_U – The penalty for unresolved cases after the specified delay is set to 0.75.
c_{FN} – The clinical cost of missed critical events because of false negatives set to 5.0.
c_{FP} – The clinical cost of unnecessary alarms due to false positives set to 1.0.

Formulas (1)–(3) follow standard episode-based event definitions and probabilistic prediction. Formula (4) uses empirical risk-targeted calibration on a held-out CAL. Formulas (5)–(6) build on classical value-of-information and cost-sensitive decision-making; our novelty lies in adapting these principles to streaming missing-data management under explicit budget and delay constraints. In Equation (6), the total number of evaluated decisions normalizes FN and FP, whereas strict FNR is computed relative to true positive-event cases.

3.3 Dataset

The dataset used is the HUPA-UCM Diabetes Dataset, which is a longitudinal multimodal dataset for Type 1 diabetes [31]. The main variables included are time, glucose, heart rate, steps, calories, carb input, bolus volume delivered, and basal rate. For evaluation, ground-truth glucose values are stored before masking. The sampling interval is five minutes. For evaluation, ground-truth glucose is preserved before masking so that delayed resolution can be assessed against known values.

3.4 Testing

In order to create simulated missing glucose data, a separate MCAR (p) mask is employed. A variety of possibilities are evaluated for the mask ($p = 0.10$, $p = 0.30$, and $p = 0.50$). Additionally, the mask is used as a MAR TA mask based on both time of day and activity. For MNAR_G250, missing is defined as a function of glucose level, with values above 250 mg/dL being more likely to be missing. For the main scenario-wise comparison, Table 1 reports results for budget = 0.10 and delay = 30 minutes. Additional tables report separate robustness, ablation, and sensitivity runs using the same nominal budget and delay settings and should be interpreted within their respective experimental settings.

4 RESULTS

The overall goal of this study is to improve decision-making under missing data by prioritizing the reduction of missed critical events (stricter False Negative Rate—FNR) and maintaining minimal operational burden in clinical settings.

The primary metrics are strict FNR (resolved) and mean total burden, which represents both the total operational burden and total error cost associated with querying. All results were obtained on the HUPA dataset, with splits at the individual subject level, across multiple folds.

Results show that targeted delayed confirmation can reduce missed critical events. It is compared to baselines. Additional experiments provide a means for testing the robustness of the results, comparing the results with a more efficient querying baseline, and evaluating the importance of each component through an ablation study.

4.1 Experimental results

A view of how the policy behaves over a sample subject's path with MCAR30 and under a low budget can be found in Figures 2–4. Figure 2 shows the real and masked glucose trajectories. Figure 3 shows the predicted event probability together with the decision threshold. Figure 4 shows EVSI (t), the query threshold, and the issued query points over time. It illustrates the query threshold and EVSI values over time, showing how missing, predictive risk, and delay-related cost interact in the decision process.

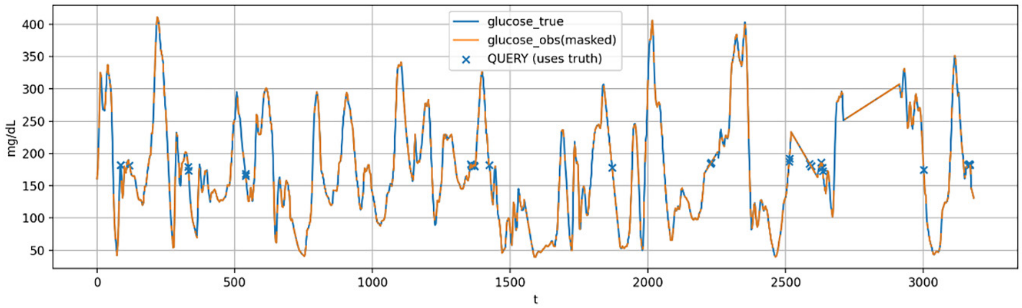

Fig. 2. Glucose real versus masked

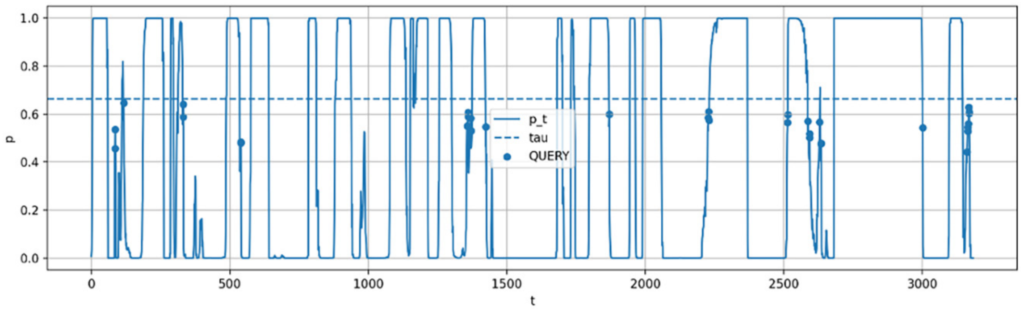

Fig. 3. Predicted probability and threshold

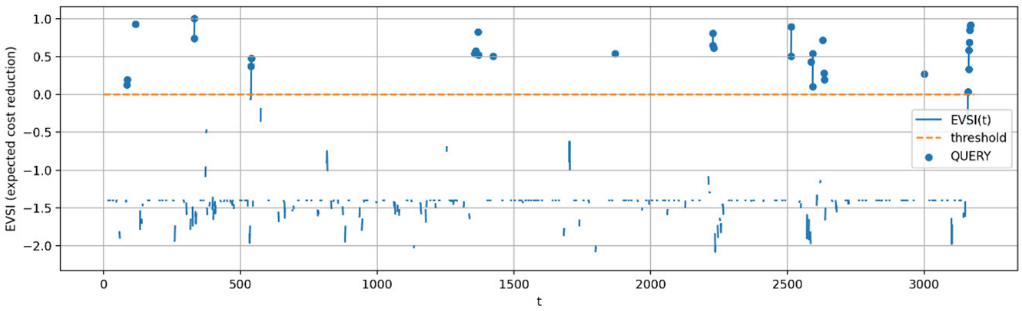

Fig. 4. EVSI(t) and query threshold

Figure 5 shows the representative MCAR30 policy as a Risk-Burden Frontier. Policies that are closest to the bottom left corner of the frontier represent the best Risk-Burden Trade-Off because they have the lowest (Strict FNR) and total burden (Total Burden). EVSI_POS offers the most feasible operating point, while EVSI_BIND has less risk; however, it has a higher total burden. The x-axis represents mean total burden. Lower values are better. The y-axis is labeled Strict FNR. Our aim is to find values of FNR that represent the fewest number of missed Critical Events. Lower strict FNR values are preferred because they indicate fewer missed critical events. Since EVSI_POS is positioned below OFFLINE_CRC, it is clear that EVSI_POS has fewer missed Critical Events than OFFLINE_CRC and is only slightly increasing the burden. On the other hand, ALWAYS_QUERY is located at the extreme right side of the frontier. It indicates that indiscriminate querying is operationally ineffective.

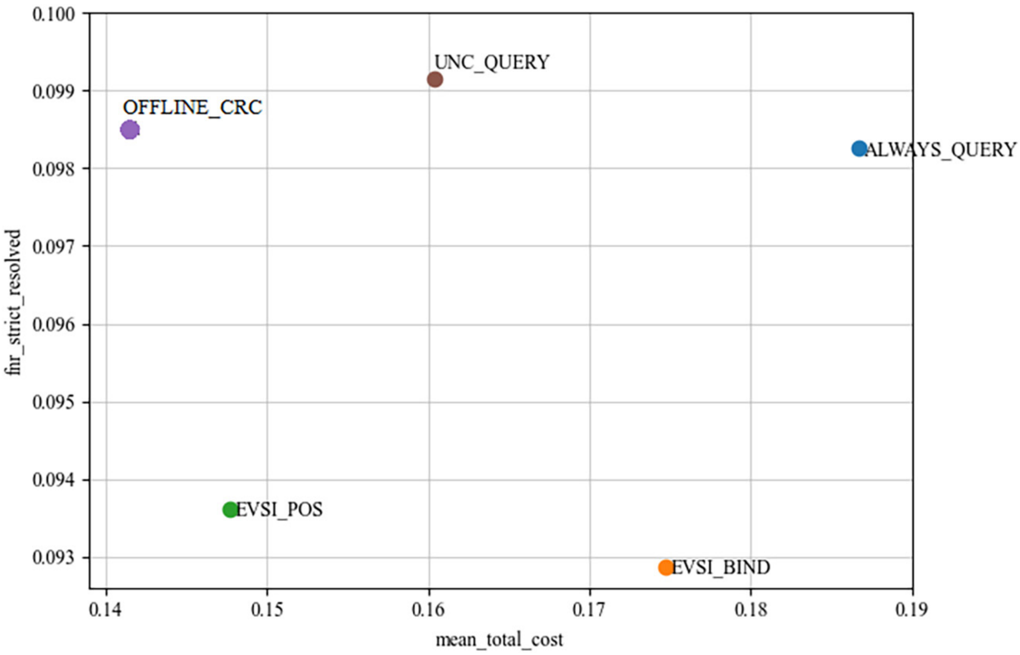

Fig. 5. Risk burden main plot (frontier), MCAR = 30, bm = 0.10, delay = 30 min

4.2 Main scenario

Table 1 reports the representative setting with budget = 0.10 and delay = 30 minutes across all five scenarios and policies. Several patterns are clear. EVSI_POS produces lower or comparable strict FNR than ALWAYS_QUERY while using substantially fewer queries in most scenarios. The most significant query count variations are shown in MCAR30 and MCAR50.

EVSI_BIND is a more conservative EVSI-based policy variant. It can further reduce strict FNR, but this reduction is achieved at the cost of higher query frequency and greater operational burden. UNC_QUERY does not match the benefits of EVSI_POS; therefore, being proximal to the decision boundary does not represent a sufficient surrogate measure for the utility of delayed measurement. As one would expect, the amount of benefit decreases in the negative outcome MNAR_G250 scenario because the most clinically relevant missing values will have the highest probability of being missing. Table 1 shows that EVSI_POS consistently reduces strict FNR relative to

OFFLINE_CRC in the MCAR and MAR scenarios, while requiring only a modest query rate and a small increase in mean burden. EVSI_BIND can further reduce strict FNR in some MCAR settings, but its most consistent effect is higher query frequency and higher operational burden.

Figure 6 provides the comparisons of the strict FNR across the five major missing data scenarios. The largest gains of EVSI-based querying are observed in MCAR30 and MCAR50, whereas the improvement becomes smaller in the more adverse MNAR_G250 setting. Within the MCAR scenarios, larger improvements were observed as the missing-data rate increased. Query rate is computed over all evaluated decision times. Under the budgeted setting used in Table 1, all query-capable policies are constrained by the query budget b = 0.10. ALWAYS_QUERY denotes querying every eligible missing candidate allowed by the budget, not querying every missing glucose value without constraint.

Table 1. Main scenario-wise comparison

Scenarios	Policy	FNR Strict	Burden	Query Rate
MCAR10	ALWAYS_QUERY	.111	.173	.0099
MCAR10	EVSI_BIND	.109	.169	.0089
MCAR10	EVSI_POS	.110	.160	.0029
MCAR10	OFFLINE_CRC	.111	.157	.0000
MCAR10	UNC_QUERY	.111	.164	.0038
MCAR30	ALWAYS_QUERY	.098	.187	.0297
MCAR30	EVSI_BIND	.093	.175	.0267
MCAR30	EVSI_POS	.094	.148	.0086
MCAR30	OFFLINE_CRC	.099	.142	.0000
MCAR30	UNC_QUERY	.099	.160	.0120
MCAR50	ALWAYS_QUERY	.082	.231	.0500
MCAR50	EVSI_BIND	.072	.208	.0460
MCAR50	EVSI_POS	.074	.165	.0153
MCAR50	OFFLINE_CRC	.082	.155	.0000
MCAR50	UNC_QUERY	.083	.185	.0190
MAR_TA	ALWAYS_QUERY	.084	.162	.0126
MAR_TA	EVSI_BIND	.082	.160	.0126
MAR_TA	EVSI_POS	.082	.147	.0041
MAR_TA	OFFLINE_CRC	.084	.143	.0000
MAR_TA	UNC_QUERY	.084	.154	.0067
MNAR_G250	ALWAYS_QUERY	.114	.156	.0083
MNAR_G250	EVSI_BIND	.114	.154	.0073
MNAR_G250	EVSI_POS	.114	.145	.0014
MNAR_G250	OFFLINE_CRC	.114	.143	.0000
MNAR_G250	UNC_QUERY	.114	.151	.0046

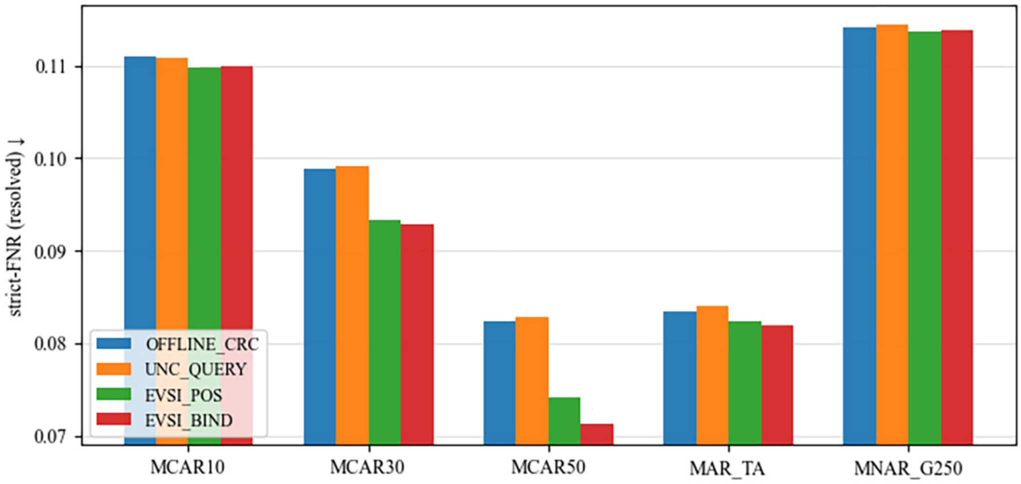

Fig. 6. Comparison scenario wise of strict FNR

4.3 Robustness across base predictors

The results in Table 2 (reports a separate robustness analysis should not be interpreted as a direct numerical replication of Table 1) suggest that EVSI-based querying shows a similar pattern when XGBoost replaces LR. This suggests that the decision strategy, rather than the predictive backbone, drives the overall performance. It reports a separate robustness analysis designed to test whether the EVSI-based policy gains remain visible when the base predictor is changed from LR to XGBoost. Figure 7 further supports this hypothesis. The strict-FNR gains of EVSI_POS compared with OFFLINE_CRC remain visible when the base predictor is changed from LR to XGB. Therefore, these gains appear to result from the EVSI_POS decision policy rather than from an individual model. Moreover, the sign and size of the gain for Missing at Random—Time/Activity (MAR_TA) are as would be expected for the models used.

Table 2. Robustness across LR and XGBoost

Scenario	Base Predictor	Policy	FNR strict	Mean Burden	Query Rate
MAR_TA	LR	EVSI_BIND	.1122	.1997	.0125
MAR_TA	LR	EVSI_POS	.1122	.1915	.0066
MAR_TA	LR	OFFLINE_CRC	.1140	.1850	.0000
MAR_TA	XGB	EVSI_BIND	.1133	.1997	.0119
MAR_TA	XGB	EVSI_POS	.1126	.1906	.0062
MAR_TA	XGB	OFFLINE_CRC	.1181	.1905	.0000
MCAR30	LR	EVSI_BIND	.1111	.1833	.0247
MCAR30	LR	EVSI_POS	.1123	.1661	.0113
MCAR30	LR	OFFLINE_CRC	.1186	.1588	.0000
MCAR30	XGB	EVSI_BIND	.0997	.1759	.0277
MCAR30	XGB	EVSI_POS	.0999	.1551	.0131
MCAR30	XGB	OFFLINE_CRC	.1173	.1613	.0000

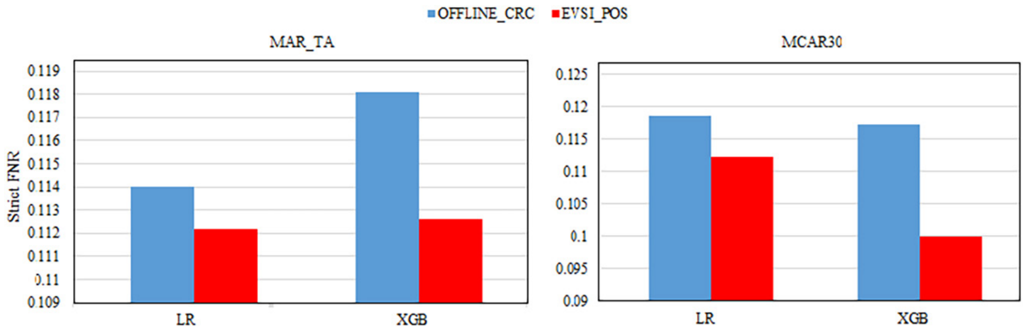

Fig. 7. Robustness comparison of strict FNR across LR and XGBoost

4.4 Ablation

Table 3 illustrates how the full EVSI pipeline and the unconditional alternative behaved similarly across the two scenarios. This suggests that most of the empirical gains come from the overall policy logic rather than from a highly specialized internal sampling mechanism. Table 3 reports a separate ablation run and should be used to compare FULL and UNCOND variants within the same ablation setting.

Table 3. Ablation of the counterfactual sampler

Scenario	Variant	Policy	FNR Strict	Mean Total Burden	Query Rate
MAR_TA	FULL	EVSI_BIND	.0636	.1332	.0127
MAR_TA	FULL	OFFLINE_CRC	.0650	.1176	.0000
MAR_TA	UNCOND	EVSI_BIND	.0632	.1309	.0115
MCAR30	FULL	EVSI_BIND	.0840	.1788	.0284
MCAR30	FULL	OFFLINE_CRC	.0921	.1512	.0000
MCAR30	UNCOND	EVSI_BIND	.0838	.1764	.0273

4.5 Sensitivity analysis and runtime overhead

We also carried out a detailed sensitivity analysis for the stability of the heuristic penalty weights. For the representative case we set the parameters budget = 0.10 and delay = 30 min. We varied the operational burden weights in Equation (6) by –20%, –10%, +10%, and +20% relative to their baseline values, while keeping the penalty weights for clinical false-negatives and false-positives constant. Our aim here was to check if an EVSI-based decision-making layer would be heavily dependent upon one manually set penalty level or whether its qualitative ordering would remain stable across different penalty levels. As presented in Table 4, the risk–burden orderings are preserved across all perturbations as well for both options: EVSI_POS remains the lower-burden EVSI option, while EVSI_BIND generally behaves as the more conservative option, requiring more queries and higher burden while achieving lower or comparable strict FNR depending on the scenario.

Table 4. Penalty-weight sensitivity analysis under ±10% and ±20% perturbations

Test Case	Description	Strict FNR Range	Query-Rate Range	Δ Burden vs OFFLINE_CRC Range
MCAR30	EVSI_POS	.093–.095	.008–.009	.004–.007
MCAR30	EVSI_BIND	.092–.095	.024–.027	.023–.039
MCAR50	EVSI_POS	.074–.075	.012–.021	.009–.012
MCAR50	EVSI_BIND	.071–.073	.041–.044	.032–.059

The Monte Carlo EVSI calculation introduced a small online computational overhead. With K = 64 Monte Carlo samples, the mean runtime was 2.25 ms per missing-glucose candidate decision, with a median of 2.10 ms and a 95th percentile of 3.46 ms. Results are shown in Table 5. These results indicate that the online EVSI decision step adds only a small per-decision overhead in the evaluated CPU environment.

Table 5. Runtime overhead of the Monte Carlo EVSI calculation

Monte Carlo Samples	Mean ms/Decision	Median ms/Decision	95th Percentile
16	2.09	2.04	2.56
32	2.09	2.03	2.59
64	2.25	2.1	3.46

4.6 Statistical comparison with QBC (Wilcoxon)

Query-by-Committee (QBC) was included as an uncertainty-based active querying baseline. Disagreement among multiple predictive models is used as a proxy for uncertainty. The cases with higher disagreement are prioritized for querying.

To assess whether the observed policy differences are consistent across folds, we performed a paired Wilcoxon signed-rank test. It compares EVSI_POS with the QBC baseline using the paired Wilcoxon signed-rank test [32], [33]. For each fold, we compute the metric difference (Δ = EVSI_POS – QBC) for strict-FNR (resolved) and mean total burden.

A negative Δ indicates that EVSI_POS is better. In the per-scenario analysis (n_pairs = 3), a consistent advantage of EVSI_POS is suggested (all Δ negative; r_rb = –1). Although the tests are underpowered (p = 0.25), they do not reach significance. When we pool both scenarios (pooled, n_pairs = 6), EVSI_POS consistently shows better statistics than QBC in both strict FNR and burden (p = 0.03125 for both; r_rb = –1), demonstrating consistent improvement under the same budget and delay constraints. Table 6 reports the Wilcoxon comparison of EVSI_POS against QBC.

Table 6. Wilcoxon EVSI_POS vs. QBC

Scope	Scenario	Metric	n_pairs	median Δ	*p*-Value	r_rb
per-scenario	MCAR30	strict-FNR (resolved)	3	–0.003875	.25	–1
per-scenario	MCAR30	mean total burden	3	–0.003394	.25	–1
per-scenario	MNAR_G250	strict-FNR (resolved)	3	–0.001057	.25	–1
per-scenario	MNAR_G250	mean total burden	3	–0.003759	.25	–1
pooled	ALL	strict-FNR (resolved)	6	–0.002151	.03125	–1
pooled	ALL	mean total burden	6	–0.003576	.03125	–1

5 CONCLUSIONS

The proposed framework was evaluated through a broad set of experiments. From the results obtained, EVSI_POS provides a better balance between risk and burden in most scenarios by reducing the FNR at a moderate cost. EVSI_BIND is more cautious. When a high-value query is selected, EVSI_BIND temporarily abstains until the delayed confirmatory information becomes available. Therefore, EVSI_BIND is more cautious and can reduce risk in several settings, but this is achieved at higher query frequency and burden. Within the MCAR scenarios, larger improvements were observed as the missing data rate increased. The overall result shows that EVSI-based Active Acquisition improves the risk-burden trade-off compared to the Offline and Heuristic Baseline methods. These results demonstrate the benefits of EVSI-based active acquisition across the evaluated scenarios. The benefits of EVSI-based Active Acquisition remain when compared to Baseline methods under both predictive models used in this study (LR to XGB) and under the realistic measurement delay/budget constraint conditions. Hence, we can conclude that the advantages of EVSI-based Active Acquisition derive from its decision-focused strategy, rather than from the specific model.

Compared to MCAR/MAR, the MNAR-G250 represents essentially an adversarial case. This is because the probability of missing is directly related to the value of the signal itself (e.g., glucose is missing more often when it is high). This causes the available information to become biased (selection bias), making it more difficult to calibrate and assess risk based on observed data alone. Second, the advantage of active acquisition can decrease when the delay is large. This is because query resolution arrives too late in a regime where critical events can develop quickly. Compared to MCAR/MAR, some configurations do not indicate that an EVSI-based policy failed on a theoretical basis. Rather, it indicates that MNAR limits how much information can be recovered from data without first identifying the missing data mechanism.

6 FUTURE WORK

In the near term, we plan to extend the framework in several directions. First, the risk-targeted calibration layer will be developed and integrated with a fully formulated CRC/LTT-based risk control mechanism. This will ensure the creation of a tighter link between empirical measures of downstream performance and sample risk targeted calibration. Second, the framework will incorporate more expressive counterfactual value models. Third, it will be extended from short-horizon glucose alarming to a larger class of medical time-series decision tasks, including other physiological monitoring settings, multimodal patient streams, and a variety of clinically structured missing patterns. Future versions of the framework will include more complex delayed confirmations. A more practical approach for deployment would likely involve separating the full EVSI computation from the wearable sensor. Wearable devices typically have limited processing capabilities. They are designed primarily to sense physiological signals and transmit these data to a secondary processing system. Examples include a user's smartphone, a home gateway, a hospital edge server, or even a cloud-based monitoring service.

This type of architecture is most appropriate when resources are constrained in terms of power consumption. Each wearable device can remain focused on its primary functions of sensing and transmitting. The EVSI-based decision process, which requires greater computational capability, is handled by the more powerful secondary processing system.

While delayed confirmation was simulated using fixed time intervals in the current prototype, measurement availability in real clinical settings may vary due to sensor contact loss, patient movement, network instability during data transfer, or workflow-related delays. The EVSI framework can be generalized to account for this variability by taking an expectation over a waiting-time distribution. Instead of assuming a single deterministic delay, the decision algorithm would evaluate the expected value of waiting across possible arrival times of the confirmatory measurement.

A more extensive validation process for the framework will be performed using larger healthcare datasets.

7 ACKNOWLEDGMENTS

This publication was made possible with the financial support of AKKSHI. Its content is the responsibility of the author, and the opinion expressed therein is not necessarily the opinion of AKKSHI.

8 REFERENCES

[1] W. Du, W. Côté, and Y. Liu, "SAITS: Self-attention-based imputation for time series," *Expert Systems with Applications*, vol. 219, p. 119619, 2023. https://doi.org/10.1016/j.eswa.2023.119619

[2] L. P. Le, X.-H. Nguyen Thi, T. Nguyen, M. A. Riegler, P. Halvorsen, and B. T. Nguyen, "Missing data imputation for noisy time-series data and applications in healthcare," *arXiv preprint arXiv:2412.11164*, 2024. https://doi.org/10.48550/arXiv.2412.11164

[3] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, and Y. Li, "BRITS: Bidirectional recurrent imputation for time series," *Advances in Neural Information Processing Systems*, vol. 31, 2018. https://doi.org/10.48550/arXiv.1805.10572

[4] Y. Tashiro, J. Song, Y. Song, and S. Ermon, "CSDI: Conditional score-based diffusion models for probabilistic time series imputation," *Advances in Neural Information Processing Systems*, vol. 34, 2021. https://doi.org/10.48550/arXiv.2107.03502

[5] L. Qian *et al.*, "CSAI: Conditional self-attention imputation for healthcare time-series," *IEEE Journal of Biomedical and Health Informatics*, 2025. https://doi.org/10.1109/JBHI.2025.3648181

[6] M. Liu *et al.*, "Handling missing values in healthcare data: A systematic review of deep learning-based imputation techniques," *Artificial Intelligence in Medicine*, vol. 142, p. 102587, 2023. https://doi.org/10.1016/j.artmed.2023.102587

[7] W. Du *et al.*, "TSI-bench: Benchmarking time series imputation," *arXiv preprint arXiv:2406.12747*, 2024. https://doi.org/10.48550/arXiv.2406.12747

[8] L. Qian *et al.*, "Beyond random missingness: Clinically rethinking for healthcare time series imputation," *arXiv preprint arXiv:2405.17508*, 2024. https://doi.org/10.48550/arXiv.2405.17508

[9] Z. Wang, L. Yang, L. Sun, Q. Wen, and Y. Wang, "Task-oriented time series imputation evaluation via generalized representers," *Advances in Neural Information Processing Systems*, vol. 37, pp. 137403–137431, 2024. https://doi.org/10.52202/079017-4365

[10] J. A. Raj, L. Qian, and Z. Ibrahim, "Modular deep learning for multivariate time-series: Decoupling imputation and downstream tasks," *arXiv preprint arXiv:2411.03941*, 2024. https://doi.org/10.48550/arXiv.2411.03941

[11] F. Z. Poleto, J. M. Singer, and C. D. Paulino, "Missing data mechanisms and their implications on the analysis of categorical data," *Statistics and Computing*, vol. 21, no. 1, pp. 31–43, 2011. https://doi.org/10.1007/s11222-009-9143-x

[12] A. M. Alharthi *et al.*, "Improving penalized logistic regression model with missing values in high-dimensional data," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 2, pp. 40–54, 2022. https://doi.org/10.3991/ijoe.v18i02.25047

[13] A. N. Angelopoulos, S. Bates, E. J. Candès, M. I. Jordan, and L. Lei, "Learn then test: Calibrating predictive algorithms to achieve risk control," *Annals of Applied Statistics*, vol. 19, no. 2, pp. 1641–1662, 2025. https://doi.org/10.1214/24-AOAS1998

[14] A. N. Angelopoulos, S. Bates, A. Fisch, L. Lei, and T. Schuster, "Conformal risk control," in *International Conference on Learning Representations*, 2024. https://doi.org/10.48550/arXiv.2208.02814

[15] Y. Qin, M. van der Schaar, and C. Lee, "Risk-averse active sensing for timely outcome prediction under cost pressure," *Advances in Neural Information Processing Systems*, vol. 36, pp. 6397–6411, 2023. https://doi.org/10.52202/075280-0279

[16] C. Tran and F. Fioretto, "Data minimization at inference time," *Advances in Neural Information Processing Systems*, vol. 36, 2023. https://doi.org/10.48550/arXiv.2305.17593

[17] L. M. G. Steuten *et al.*, "A systematic and critical review of the evolving methods and applications of value of information in academia and practice," *PharmacoEconomics*, vol. 31, no. 1, pp. 25–48, 2013. https://doi.org/10.1007/s40273-012-0008-3

[18] A. Heath, N. Kunst, and C. Jackson, Eds., *Value of Information for Healthcare Decision-Making*. Chapman and Hall/CRC, 2024. https://doi.org/10.1201/9781003156109

[19] B. G. Mohammed and D. S. Hasan, "Smart healthcare monitoring system using IoT," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17, no. 1, pp. 141–152, 2023. https://doi.org/10.3991/ijim.v17i01.34675

[20] L. Ocares-Cunyarachi and L. Andrade-Arenas, "Design of a mobile app to monitor and control in real time type 2 diabetes mellitus in Peru," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17, no. 10, pp. 176–192, 2023. https://doi.org/10.3991/ijim.v17i10.38207

[21] A. Y. Al-Zoubi, A. Tahat, R. A. Wahsheh, M. Taha, L. Al-Tarawneh, and O. Hasan, "A bachelor degree program in IoT engineering: Accreditation constraints and market demand," *International Journal of Engineering Pedagogy (iJEP)*, vol. 12, no. 4, pp. 17–34, 2022. https://doi.org/10.3991/ijep.v12i4.31429

[22] E. Zanaj, L. Balliu, G. Basha, and E. Gjata, "Studying the behavior of a modified deep learning model for disease detection through x-ray chest images," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 5, 2024. https://doi.org/10.14569/IJACSA.2024.0150585

[23] L. Balliu *et al.*, "Enhancing heart disease prediction accuracy by comparing classification models employing varied feature selection techniques," *Serbian Journal of Electrical Engineering*, vol. 21, no. 3, pp. 375–390, 2024. https://doi.org/10.2298/SJEE2403375B

[24] A. Xhemali, E. Zanaj, G. Basha, and L. Balliu, "Enhancing energy efficiency prediction in assisted living through GA-FIS, PSO-FIS, and NSGA-II-FIS: A comparative evaluation," in *Automation, Robotics & Communications for Industry 4.0/5.0, ARCI 2025*, 2025, p. 62.

[25] F. E. Harrell, *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis* (2nd ed.). Cham: Springer, 2015. https://doi.org/10.1007/978-3-319-19425-7

[26] S. Hakkal *et al.*, "XGBoost to enhance learner performance prediction," *Computers and Education: Artificial Intelligence*, vol. 7, p. 100254, 2024. https://doi.org/10.1016/j.caeai.2024.100254

[27] K. Kassem *et al.*, "An innovative artificial intelligence-based method to compress complex models into explainable, model-agnostic and reduced decision support systems with application to healthcare (NEAR)," *Artificial Intelligence in Medicine*, vol. 151, p. 102841, 2024. https://doi.org/10.1016/j.artmed.2024.102841
[28] M. Sadatsafavi *et al.*, "Expected value of sample information calculations for risk prediction model validation," *Medical Decision Making*, vol. 45, no. 3, pp. 232–244, 2025. https://doi.org/10.1177/0272989X251314010
[29] A. Heath, I. Manolopoulou, and G. Baio, "Efficient Monte Carlo estimation of the expected value of sample information using moment matching," *Medical Decision Making*, vol. 38, no. 2, pp. 163–173, 2018. https://doi.org/10.1177/0272989X17738515
[30] A. E. Ades, G. Lu, and K. Claxton, "Expected value of sample information calculations in medical decision modeling," *Medical Decision Making*, vol. 24, no. 2, pp. 207–227, 2004. https://doi.org/10.1177/0272989X04263162
[31] J. I. Hidalgo, J. Alvarado, M. Botella, A. Aramendi, J. M. Velasco, and O. Garnica, "HUPA-UCM Diabetes Dataset [Data set]," *Mendeley Data*, vol. V1, 2024. https://doi.org/10.1016/j.dib.2024.110559
[32] K. Okoye and S. Hosseini, "Wilcoxon statistics in R: Signed-rank test and rank-sum test," in *R Programming: Statistical Data Analysis in Research*, Springer Nature Singapore, 2024, pp. 279–303. https://doi.org/10.1007/978-981-97-3385-9_13
[33] R. Gilad-Bachrach, A. Navot, and N. Tishby, "Query by committee made real," *Advances in Neural Information Processing Systems*, vol. 18, pp. 443–450, 2005. https://doi.org/10.5555/2976248.2976304

9 AUTHORS

Elma Zanaj received an M.S. degree in Telecommunications Engineering from Polytechnic University of Tirana in 2003 and a Ph.D. degree in Electronics, Informatics, and Telecommunications Engineering from Marche Polytechnic University, Ancona, Italy, in 2008. She is currently a Full Professor at the Polytechnic University of Tirana, Albania, with research focusing on WSNs applications, IoT, and the use of AI in healthcare (E-mail: ezanaj@fti.edu.al).

Lorena Balliu is a full-time Assistant Lecturer at the Department of Computer Science Fundamentals, Polytechnic University of Tirana. She is currently pursuing a Ph.D. at the Department of Electronics and Telecommunications. Her research interests include Artificial Intelligence, Machine Learning algorithms, and their applications in telemedicine and data analysis, supported by both academic and private sector experience (E-mail: lorena.balliu@fti.edu.al).

Anita Xhemali obtained her M.Sc. degree in Computer Science from the Polytechnic University of Tirana in 2005. She is currently a Ph.D. candidate at the Faculty of Information Technology. She is currently a full-time Lecturer at the Faculty of Electrical Engineering. Her research explores the integration of fuzzy logic systems with intelligent IoT architecture and advanced healthcare, developing efficient computational approaches (E-mail: anita.xhemali@fti.edu.al).

Gledis Basha obtained his B.Sc. (2012) and M.Sc. (2014) degrees from the Faculty of Information Technology, Polytechnic University of Tirana (UPT). He is currently a full-time Assistant Lecturer and a Ph.D. candidate at the Faculty of Information Technology. His research focuses on information technology systems and healthcare technologies, particularly their practical applications in engineering and industry (E-mail: gbasha@fti.edu.al).