

Fuzzy Fusion of Biometric Signals for Early Estimation of Vaso-Occlusive Crisis Risk: An Experimental Study

Blaise Muhala Luhepa^{1*}, Jean Pepe Buanga Mapetu^{2*}, Gisèle Makengo Kafuti^{3**}, Dieuleveut Nianga Kayakaya^{4*}, Nathanael Kasoro Mulenda^{5*}, Rostin Mabela Makengo^{6****}, Herman Matondo Mananga^{7****}

*Department of Mathematics, Statistics and Computer Science, University of Kinshasa, Kinshasa, DR Congo

**Department of Biology, University of Kinshasa, Kinshasa, DR Congo

***Centre for Research in Mathematics Education (CREM), DR Congo

blaise.muhala@unikin.ac.cd¹, jpbuanga@gmail.com², giselemkengo@gmail.com³, niangadieuleveut@gmail.com⁴,
nathanael.kasoro@unikin.ac.cd⁵, rostin.mabela@unikin.ac.cd⁶, herman.matondo@unikin.ac.cd⁷

Article Info

Article history:

Received 2026-08-10

Revised 2026-09-21

Accepted 2026-10-02

Keyword:

*Fuzzy logic,
Remote monitoring,
Sickle cell disease,
Vaso-occlusive crisis,
Wearable devices.*

ABSTRACT

Early detection of vaso-occlusive crises remains challenging because crisis onset is multifactorial and physiological responses vary across patients. This study evaluates a fuzzy-fusion method that combines six patient-centered biometric deviations with measurement-quality and contextual information to produce a continuous pre-crisis ranking score. The analysis included 500 patients and 321,988 15-minute windows after exclusion of active-crisis windows. A window was labeled positive when the next recorded vaso-occlusive crisis began within 30 minutes after the window end. Patients were separated between training and testing (375/125), and decision thresholds were selected on the training set. The test set contained 216 positive windows among 80,842 windows (0.27%). Fuzzy fusion achieved a PR AUC of 0.642, F1-score of 0.642, precision of 0.581, and recall of 0.718, compared with PR AUC 0.343 and F1-score 0.496 for hard thresholds. Logistic regression achieved the highest discrimination (PR AUC 0.787; F1-score 0.652), whereas fuzzy fusion generated fewer false-positive windows at its training-derived operating point. The fuzzy score was not well calibrated as a probability and should be interpreted as an ordinal risk index. The results support methodological feasibility and auditability, not clinical validation.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Sickle cell disease remains one of the most burdensome inherited disorders, particularly in regions where access to specialized care is limited. Vaso-occlusive crises are among its most frequent and disabling manifestations. They cause acute pain, repeated medical consultations, and numerous hospitalizations [1,2]. Management often begins only after the crisis is already established. Earlier recognition of a physiological drift that may precede an episode could therefore support monitoring without replacing clinical judgment.

Wearable devices enable longitudinal monitoring of parameters such as oxygen saturation, heart rate, heart rate

variability, skin temperature, activity, and indices derived from photoplethysmography. In sickle cell disease, Shah et al. demonstrated the feasibility of non-invasive monitoring of peripheral vasoconstriction biomarkers using a wrist-worn device [2]. More general studies on connected sensors nevertheless emphasize that predictive value depends on measurement quality, calibration, context of use, and interindividual variability [3,4,5,6].

Machine learning has already been applied in several studies on sickle cell disease, particularly for diagnosis, classification, and monitoring. However, the review by Machado et al. identifies recurrent limitations, including small sample sizes, heterogeneous datasets, risk of

overfitting, and insufficient explainability [7]. These challenges become particularly important in remote monitoring, where time series are incomplete, noisy, and highly dependent on patient behavior.

Fuzzy logic provides a relevant framework in this context. Rather than imposing a sharp boundary between normal and abnormal states, it represents gradual levels of deviation. A slight decrease in SpO₂ may therefore contribute weakly to the score and contribute more strongly as it becomes more pronounced or occurs together with other changes. This flexibility is suitable for imperfect data and decisions informed by expert knowledge [8]. Although fuzzy logic does not resolve clinical uncertainty by itself, it makes the computational mechanism more explicit than a fully opaque model [9].

Because vaso-occlusive crises are multifactorial, a single physiological channel is unlikely to be sufficient in all situations. Multimodal-fusion studies highlight the value of combining complementary sources while also showing that the aggregation method strongly affects performance and result interpretability [10,11,12,13]. The challenge is therefore not to accumulate variables, but to preserve their physiological meaning and distinguish a coherent drift from a variation explained by activity, stress, or poor signal quality.

To the best of our knowledge, the gap is not the absence of fuzzy inference or multimodal physiological prediction in healthcare. Rather, few studies in sickle cell disease have examined an auditable short-horizon pipeline that jointly combines patient-centered biometric deviations, explicit measurement-quality information, contextual modifiers, and patient-separated evaluation for windows immediately preceding a recorded vaso-occlusive crisis. Here, the term "risk" denotes a ranking score produced by the system and must not be interpreted as a calibrated clinical probability.

We address the following question: on a remote-monitoring dataset derived from patients with sickle cell disease, what trade-off does fuzzy fusion provide, compared with hard-threshold counting and logistic regression, for distinguishing windows occurring within the 30 minutes preceding a recorded vaso-occlusive crisis? Our contribution comprises four elements: patient-centered orientation and normalization of six physiological deviations; gradual fusion through explicit membership functions and rules; separate integration of signal quality and simplified context; and patient-level train-test separation with metrics and uncertainty estimates suited to rare outcomes. The objective is methodological feasibility and transparent comparison, not clinical validation.

A. State of the Art

The literature relevant to this study can be divided into five areas: biomarkers associated with vaso-occlusive crises, remote monitoring using wearable devices, multimodal fusion, representation of uncertainty, and transparency in decision-support systems. These areas have progressed separately, while their integration into a continuous pre-crisis score remains poorly documented in sickle cell disease.

1) Crisis Biomarkers and Monitorable Signals in Sickle Cell Disease:

Vaso-occlusive crises cannot be reduced to a single biomarker. The review by Khurana et al. describes a set of inflammatory, hemolytic, endothelial, and vascular markers associated with acute episodes [14]. Machado et al. further show that artificial intelligence has already been applied to several tasks related to sickle cell disease, although using highly heterogeneous data and protocols [7]. Finally, Dampier highlights the value of continuous digital tools for moving beyond retrospective documentation of pain alone [15]. These studies support the investigation of a drift state combining several signals rather than the search for a universal threshold.

The study by Shah et al. provides an important indication of feasibility. The authors used wrist-worn photoplethysmography to longitudinally monitor markers of vasoconstriction in people with sickle cell disease [2]. This result does not demonstrate that a crisis can be accurately predicted 30 minutes in advance, but it confirms that relevant peripheral information can be collected outside the laboratory. Its fusion with oxygenation, heart rate, respiration, temperature, and heart rate variability remains an open question.

2) Wearables and Remote Monitoring: Strong Promise, Still Incomplete Integration:

Wearable devices now play an important role in continuous monitoring. Ayena et al. emphasize the value of multimodal data for studying pain, as well as the importance of security, signal quality, and cautious interpretation [3]. Bignami et al. note that multiparametric devices remain sensitive to calibration, sensor drift, and differences between users [4]. Köhler et al. distinguish three complementary uses of smartwatches: monitoring, nudging behavior, and prediction [5]. Performance reported in other conditions also shows that sensor reliability depends on context and cannot be assumed to remain constant [6].

The value of these devices also depends on the architecture surrounding the measurement process. Taherdoost describes integration of sensors into the Internet of Things as a condition for continuous monitoring, with requirements related to transmission, security, and quality [16]. Mukherjee et al. highlight the gradual shift toward more flexible biosensors and personalized approaches [17]. In practice, raw data become useful only after quality control, contextualization, and synthesis.

3) Multimodal Fusion: From Demonstrated Potential to Challenges in Clinical Implementation:

Multimodal fusion is specifically intended to combine complementary sources. Xu et al. describe the benefits and challenges of such integration in medical diagnosis [11]. Krones et al. review methods combining imaging, text, time series, and tabular data [12], whereas Siam et al. discuss recent multimodal architectures in healthcare [18]. Graph-based approaches further show that fusion may concern relationships among variables rather than simple

concatenation [19]. Recent reviews nevertheless converge on one point: data heterogeneity and clinical evaluation remain major obstacles [13].

For time series from wearable sensors, the challenges are even more concrete: asynchronous measurements, missing data, motion artifacts, and subject dependence. In epilepsy, Hasiya et al. show that feature fusion can improve prediction, although results vary with the pre-event duration and data quality [10]. This observation is methodologically transferable: performance depends as much on signal preparation as on the fusion algorithm [10,11].

4) Fuzzy Logic and Uncertainty: An Approach Suited to Progressive Physiological Drift:

Fuzzy logic was designed to represent categories with gradual boundaries [8]. This property reflects the nature of many physiological changes: an increase in heart rate or a decrease in saturation does not have the same meaning depending on its magnitude, duration, and the other signals present. Recent studies have used fuzzy systems to represent diagnostic uncertainty or construct explicit decision rules [20,21]. These findings support the use of a gradual score without validating the thresholds selected for sickle cell disease.

Uncertainty nevertheless extends beyond fuzzy logic alone. Campagner et al. note that it is rarely addressed explicitly in medical AI studies [9]. Aliyeva and Mehdiyev also show that the evaluation of explanation methods should account for the partial reliability of information [22]. In our framework, signal quality is therefore treated as information distinct from the risk level. This distinction will be essential in future deployment: an uncertain measurement must not be equated with low risk.

5) Explainability: From Post Hoc Explanations to Intrinsic Interpretability:

Explainability has become a central issue in medical decision-support systems. Recent reviews show that most solutions still rely on post hoc explanations applied to complex models [23,24]. Liu et al. emphasize that acceptance by healthcare professionals depends both on understanding the result and on integrating the system into care practices [25]. A readable structure is therefore useful, but insufficient: interpretability must also be evaluated with the intended users.

A fuzzy approach provides structural transparency: membership functions, activated rules, and channel contributions can be inspected. This advantage is more limited than demonstrated clinical interpretability. In the present study, no evaluation involving physicians or patients was conducted. We therefore refer to an auditable mechanism or intrinsic interpretability rather than validated clinical explainability [8,23,25].

TABLE I
CRITICAL SYNTHESIS OF RECENT LITERATURE

Domain	Key findings from the literature	Limitation for the present study
Vaso-occlusive crisis biomarkers	The crisis is multifactorial and involves several physiological mechanisms.	Few wearable signals have been validated for short-term anticipation.
Wearable devices and remote monitoring	Non-invasive longitudinal monitoring is technically feasible.	Signal quality, calibration, and individual variability remain critical.
Multimodal fusion	Combining complementary sources may improve prediction.	Fusion methods and evaluation protocols are highly heterogeneous.
Fuzzy logic	Gradual transitions and explicit rules are suitable for imprecise data.	Selected parameters must be justified and tested on independent data.
Explainability and decision support	Understanding the mechanism facilitates system auditing.	Structural interpretability does not replace evaluation with users.

6) Critical Synthesis and Positioning of Our Contribution:

The literature therefore provides several useful building blocks: measurable peripheral signals, devices capable of monitoring them over time, fusion methods, and principles for gradual representation. It also identifies the limitations that should guide evaluation: noise, missing data, low prevalence, differences between patients, and the difficulty of translating a score into a clinical decision.

Our methodological novelty lies in this problem-specific integration rather than in proposing a new fuzzy inference operator. Existing fuzzy medical decision systems demonstrate the usefulness of gradual rules [20,21], while multimodal-health studies emphasize the value and difficulty of combining heterogeneous sources [10-13,18,19]. The present work combines these principles for short-horizon vaso-occlusive-crisis estimation using individualized deviations, an explicit quality gate, contextual modulation, and an auditable rule structure. The contribution concerns feasibility and the observed performance/false-positive trade-off on this dataset; it does not constitute biomarker validation, proof of robustness across devices or centers, or demonstrated clinical interpretability.

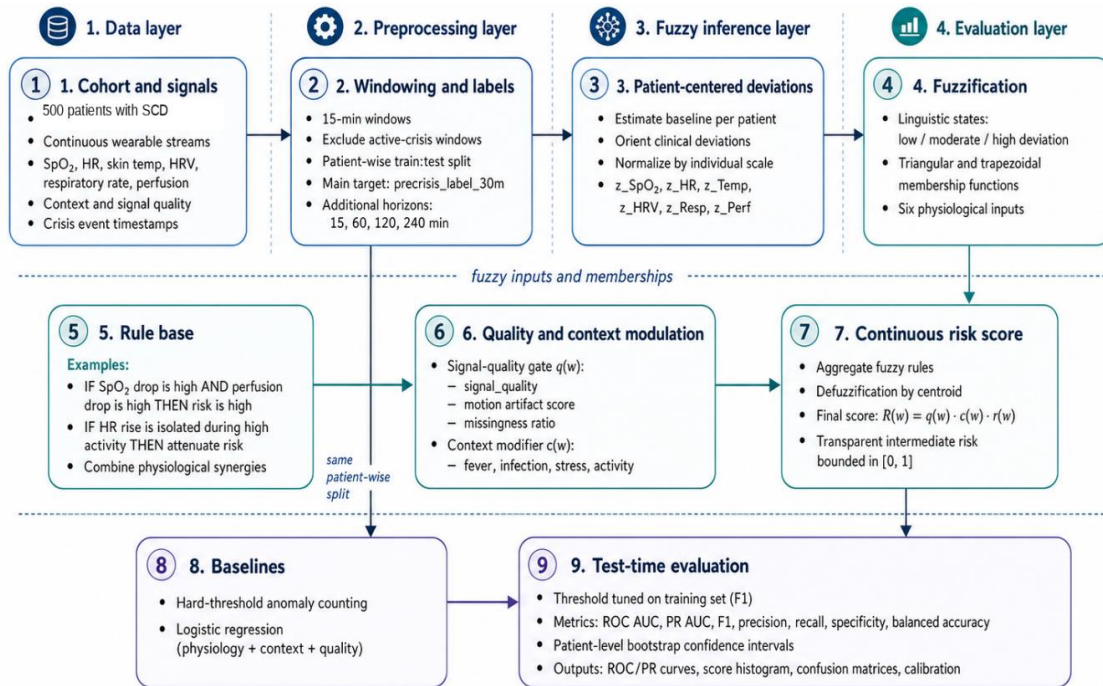


Figure 1. Architecture of the experimental pipeline for estimating a pre-crisis vaso-occlusive score.

II. METHOD

A. Study Design and Experimental Objective

This study is a methodological experiment conducted on an analytical dataset derived from data collected from patients with sickle cell disease at the HN Center in Mont-Ngafula, Kinshasa, Democratic Republic of the Congo. It examines whether a fuzzy system can transform several physiological deviations into a continuous and auditable score and compares this score with two reference methods. The analytical unit is a time window rather than a patient or a crisis episode; consequently, repeated windows from the same patient are treated as temporally related observations.

B. Data Structure

The dataset includes 500 profiles of people with sickle cell disease. Each profile contains individual baseline values, windowed physiological observations, contextual variables, quality indicators, and recorded crisis events. The analytical XLS dataset is organized into three linked tables:

- a patients table describing profile characteristics and individual baseline values;
- a crisis_events table containing the recorded crisis onset times;
- a windowed_observations table containing the windows used for the analysis.

Observations are organized into 15-minute analytical windows. Each window w associated with patient i contains six physiological summaries, contextual variables, and quality indicators. The primary target equals 1 when the next

recorded vaso-occlusive crisis begins within 30 minutes after the end of the window and 0 otherwise. Windows already labeled as occurring during an active crisis are removed before this target is evaluated. The number of windows is therefore much larger than the number of patients, and successive windows from one patient must not be interpreted as independent clinical events.

TABLE II
GROUPS OF VARIABLES USED IN THE STUDY

Group	Main variables	Role
Raw biometrics	SpO ₂ , heart rate, skin temperature, HRV, respiratory rate, perfusion	Instantaneous physiological description
Profile-centered deviations	spo2_delta_baseline, heart_rate_delta_baseline, skin_temp_delta_baseline, hrv_delta_baseline, resp_rate_delta_baseline, perfusion_delta_baseline	Main inputs to the fuzzy system
Context	activity, sleep, hydration, stress, fever, infection, medication intake	Score modulation
Signal quality	signal_quality, motion_artifact_score, missingness_ratio, imputed_flag	Reliability control
Labels	precrisis_label_30m, during_crisis_label, time_to_next_crisis_min	Training and evaluation

The present analysis uses the window-level records supplied in the analytical table; it does not reconstruct raw sensor waveforms or create additional sliding windows from those waveforms. Native device sampling frequency, device-specific firmware settings, and raw waveform processing are therefore outside the scope of the analytical file used in this study. Reproducibility in the present paper concerns the transformation of the supplied window-level variables into

model inputs and predictions. This limitation is important because sensor-level reliability cannot be inferred solely from model performance at the aggregated-window level.

The six biometric inputs were selected to represent complementary physiological dimensions plausibly relevant to short-term deterioration rather than to assert six clinically validated crisis biomarkers. SpO2 represents oxygenation; heart rate and HRV capture complementary cardiovascular/autonomic changes; skin temperature may reflect systemic or peripheral thermal change; respiratory rate describes ventilatory response; and the perfusion variable represents peripheral vascular information. The strongest direct wearable precedent in sickle cell disease concerns photoplethysmography-based vasoconstriction monitoring [2], while broader biomarker evidence supports a multifactorial view of vaso-occlusive crises [14]. Accordingly, the selected signals are treated as hypothesis-driven inputs whose independent clinical value still requires prospective validation.

Context and measurement quality are modeled separately from physiology. Fever, infection, and stress may increase the plausibility that simultaneous physiological deviations are clinically meaningful, whereas high activity can explain an isolated heart-rate increase and therefore attenuates that interpretation. Hydration, sleep, and medication intake are retained in the dataset as contextual descriptors but are not all used in the simplified modifier of the current prototype. Signal-quality, motion-artifact, missingness, and imputation indicators are retained to distinguish unreliable measurement from low physiological risk [3-6,9].

C. Data Preprocessing

1) Exclusion of Windows During an Active Crisis:

Because the objective is anticipation, all windows labeled as active crisis are excluded from the main analysis. The model must distinguish a pre-crisis state from a non-crisis state rather than recognize an episode that is already established.

2) Construction of Deviations From Baseline:

For each patient i , physiological variable k , and window w , let:

$$x_{i,k}(w) \quad (1)$$

$x_{i,k}(w)$, the value observed in the window;

$$b_{i,k} \quad (2)$$

$b_{i,k}$, the individual baseline value for this variable.

The raw deviation from baseline is defined as:

$$\Delta_{i,k}(w) = x_{i,k}(w) - b_{i,k} \quad (3)$$

Because the variables do not all have the same clinical direction, a polarity coefficient η_k is introduced:

- a) -1 when a decrease in the signal increases the risk score, particularly for SpO2, HRV, and perfusion;
- b) +1 when an increase in the signal raises the score, particularly for heart rate, skin temperature, and respiratory rate.

The oriented deviation is then:

$$d_{i,k}(w) = \eta_k \Delta_{i,k}(w) \quad (4)$$

To compare variables expressed in different units, this deviation is normalized:

$$z_{i,k}(w) = \frac{d_{i,k}(w)}{\sigma_{i,k} + \epsilon} \quad (5)$$

where $\sigma_{i,k}$ represents an individual variability scale and $\epsilon > 0$ is a numerical stabilization term.

This transformation describes a deviation relative to the patient's individual profile rather than a universal clinical abnormality. The patient-specific baseline and variability values are used only for that patient, and all windows from a held-out patient remain outside the training partition. In a prospective implementation, baseline values and variability scales must be estimated exclusively from observations available before the prediction time; using future observations to define these quantities would constitute temporal data leakage.

3) Definition of the Pre-Crisis Label:

For a horizon h , the binary label associated with window w is defined as:

$$y_i^{(h)}(w) = 1 \text{ if } 0 < t^{\text{crise}_i} - t(w) \leq h; 0 \text{ otherwise.}$$

where t^{crise_i} denotes the onset time of the next recorded crisis after the end of window w , and $t(w)$ is the window end time. If more than one crisis onset is recorded after a window, the label is determined by the nearest upcoming onset; the window is counted once rather than duplicated across events. Windows belonging to an active crisis are excluded beforehand, so no already-established crisis window is relabeled as pre-crisis. Closely spaced events can nevertheless generate temporally correlated eligible windows, which motivates grouped evaluation at the patient level.

In this study, the main analysis uses the horizon $h = 30$ minutes:

$$y_i^{(30)}(w) \quad (6)$$

This choice represents a short-term anticipation scenario. It is not a clinically validated horizon. A future study should compare several horizons and measure the actual delay between the first alert and the onset of each crisis.

D. Proposed Fuzzy-Fusion System

1) Linguistic Variables:

The fuzzy system receives the following six oriented and normalized deviations:

$$z_i(w) = (z_{i,\text{SpO}_2}(w), z_{i,\text{HR}}(w), z_{i,\text{Temp}}(w), z_{i,\text{HRV}}(w), z_{i,\text{Resp}}(w), z_{i,\text{Perf}}(w)) \quad (7)$$

Each input is described using three linguistic categories representing the intensity of the physiological deviation:

- Low deviation;
- Moderate deviation;
- High deviation.

These categories are represented by triangular or trapezoidal membership functions. A generic triangular function is written as:

$$\mu_{\Delta}(z; a, b, c) = \begin{cases} 0, & z \leq a, \\ \frac{z-a}{b-a}, & a < z \leq b, \\ \frac{c-z}{c-b}, & b < z < c, \\ 0, & z \geq c. \end{cases} \quad (8)$$

A generic trapezoidal function is written as:

$$\mu_{\text{trap}}(z; a, b, c, d) = \begin{cases} 0, & z \leq a, \\ \frac{z-a}{b-a}, & a < z \leq b, \\ 1, & b < z \leq c, \\ \frac{d-z}{d-c}, & c < z < d, \\ 0, & z \geq d. \end{cases} \quad (9)$$

Figure 2 presents the functions used in the study for the six inputs. The breakpoints are prototype design parameters and should not be interpreted as clinically validated thresholds.

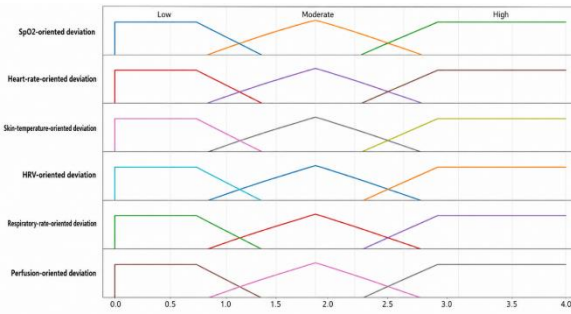


Figure 2. Membership functions for the six physiological deviations used in the study.

The overlap between categories prevents an artificial discontinuity: the same observation may partially belong to two adjacent levels.

TABLE III
ORIENTATION OF THE SIX INPUT VARIABLES

Variable	Risk-related direction	Input used
SpO2	decrease	$z_i, \text{SpO2}$
Heart rate	increase	z_i, HR
Skin temperature	increase	z_i, Temp
HRV	decrease	$z_i, \text{HRV after orientation}$
Respiratory rate	increase	z_i, Resp
Peripheral perfusion	decrease	z_i, Perf

2) Fuzzy Rule Base:

The inference system uses IF-THEN rules. The rules below illustrate the main configurations included in the prototype:

- R1: IF the decrease in SpO2 is high AND the decrease in perfusion is high, THEN the risk is high.
- R2: IF heart rate and temperature increase moderately AND HRV decreases moderately, THEN the risk is moderate to high.
- R3: IF heart rate increases markedly BUT activity is high and the other deviations remain low, THEN the risk is low to moderate.

- R4: IF several moderate deviations occur simultaneously, THEN the risk may become high even without an isolated extreme deviation.

For a rule r composed of m_r antecedents, the activation level is defined as:

$$\alpha_r(w) = \min_{j=1, \dots, m_r} \mu_{A_{r,j}}(z_j(w)) \quad (10)$$

where $A_{r,j}$ denotes the fuzzy set associated with the j th antecedent of rule r .

The rule consequents are aggregated using the maximum operator:

$$\mu_{\text{risk}}(u) = \max_r \min(\alpha_r(w), \mu_{B_r}(u)) \quad (11)$$

where B_r is the output fuzzy set associated with rule r .

3) Defuzzification:

The continuous fuzzy score is obtained using the centroid method:

$$\hat{r}_i(w) = \frac{\int_0^1 u \mu_{\text{risk}}(u) du}{\int_0^1 \mu_{\text{risk}}(u) du} \quad (12)$$

This operation provides a continuous value between 0 and 1. It facilitates window ranking and inspection of the decision mechanism, but it does not automatically convert the score into a calibrated probability [8].

E. Integration of Signal Quality and Context

1) Signal-Quality Gate:

Measurements from wearable devices may be affected by motion, signal loss, or imputation. We therefore introduce a quality factor $q_i(w)$:

$$q_i(w) = \max(0, \min(1, s_i(w) [1 - m_i(w)] [1 - a_i(w)])) \quad (13)$$

where:

- $s_i(w)$ is the overall signal quality;
- $m_i(w)$ is the proportion of missing data;
- $a_i(w)$ is the motion-artifact score.

In the prototype, $q_i(w)$ decreases the contribution of a window as overall signal quality worsens and as missingness or motion-artifact burden increases. This attenuation is a computational reliability gate, not evidence that the patient is at low risk. In a clinical implementation, sufficiently poor quality should trigger an "unreliable estimate" state or system abstention rather than a reassuring score. The present experiment does not include a dedicated channel-masking or noise-injection study; consequently, robustness to complete sensor loss or progressively degraded signals is not established here.

2) Contextual Modifier:

Context is represented by the following modifier:

$$c_i(w) = 1 + \lambda_1 F_i(w) + \lambda_2 I_i(w) + \lambda_3 S_i(w) - \lambda_4 A_i(w) \quad (14)$$

where:

- $F_i(w)$ indicates the presence of fever;
- $I_i(w)$ indicates an infection;

- $S_i(w)$ represents stress;
- $A_i(w)$ represents high physical activity.

The first three factors increase the score when the context strengthens the plausibility of the physiological drift, whereas activity attenuates the interpretation of isolated tachycardia. The coefficients λ_1 to λ_4 are prototype parameters and should be examined through sensitivity analysis.

3) Final Score:

The final prototype score is defined as:

$$R_i(w) = \min(1, \max(0, q_i(w)c_i(w)\hat{r}_i(w))) \quad (15)$$

The bounding function keeps the score within $[0,1]$. $R_i(w)$ is used as a continuous ranking index and is not presented as an individual probability of crisis.

This construction makes three distinct elements visible: deviation intensity, measurement quality, and context. It supports more direct inspection of the calculation, but its robustness to noise and missing data still requires experimental evaluation [5,6].

F. Reference Methods

The fuzzy system is compared with two reference methods.

1) Hard-Threshold Method:

For each variable k , a threshold τ_k is defined. The number of deviations exceeding their threshold is:

$$H_i(w) = \sum_{k=1}^6 \mathbf{1}[z_{i,k}(w) \geq \tau_k] \quad (16)$$

The hard-threshold score is then normalized:

$$p_i^{\text{hard}}(w) = \frac{H_i(w)}{6} \quad (17)$$

This method represents a simple count of abnormalities and does not model gradual transitions between deviation levels.

2) Logistic Regression:

The second reference method is logistic regression:

$$p_i^{\text{log}}(w) = \sigma(\beta_0 + \sum_{k=1}^6 \beta_k z_{i,k}(w) + \sum_{\ell=1}^L \gamma_\ell c_{i,\ell}(w)) \quad (18)$$

The vector $c_i(w)$ contains contextual and quality variables. Logistic regression and fuzzy fusion therefore use more information than the hard-threshold method. The hard-threshold method is intentionally retained as a simple rule-based benchmark, not as a matched-input ablation. Consequently, differences relative to hard thresholds cannot be attributed to fuzzy logic alone; a strict ablation would require identical input information and preprocessing with only the fusion operator changed.

TABLE IV
CHARACTERISTICS OF THE COMPARED METHODS

Method	Inputs	Nature	Mechanism transparency
Hard thresholds	Physiological deviations	Binary rules	High, but simplified
Logistic regression	Deviations, context, and quality	Parametric statistical model	Moderate
Proposed fuzzy fusion	Deviations, context, and quality	Gradual rules	Structurally high

G. Experimental Protocol

1) Train-Test Split:

To prevent the same patient from contributing to both training and testing, the split is performed at the patient level. Let:

$$\mathcal{P} = \{1, \dots, N\} \quad (19)$$

the set of N profiles. The partition is defined as:

$$\mathcal{P} = \mathcal{P}_{\text{train}} \cup \mathcal{P}_{\text{test}}, \quad \mathcal{P}_{\text{train}} \cap \mathcal{P}_{\text{test}} = \emptyset \quad (20)$$

All windows belonging to the same profile are assigned to the same subset. The resulting partition contains 375 training patients (241,146 windows) and 125 test patients (80,842 windows). Patient identifiers are mutually exclusive between the two partitions, so no window from a test patient is used to fit the logistic-regression coefficients, select model decision thresholds, or define training performance.

2) Decision-Threshold Selection:

The score $R_i(w)$ is converted into a binary prediction according to the following rule:

$$\hat{y}_i(w) = \mathbf{1}[R_i(w) \geq \tau_{\text{opt}}] \quad (21)$$

For each method, the decision threshold τ_{opt} is selected on the training set by maximizing the F1-score and is then frozen before test-set evaluation. No test label is used for threshold selection. The confusion matrices and threshold-dependent metrics therefore describe prespecified operating points, whereas ROC AUC and PR AUC evaluate ranking performance across all possible thresholds.

3) Uncertainty Estimation:

Confidence intervals are calculated using patient-level bootstrap resampling to preserve within-patient dependence among successive windows. The reported intervals therefore reflect uncertainty under changes in the composition of patients in the held-out set rather than an assumption that all 80,842 test windows are independent.

This procedure limits underestimation of uncertainty caused by resampling successive windows. Nevertheless, the present study relies on a single train-test split; repeated grouped validation remains necessary [26,27,28].

H. Evaluation Metrics

Given the rarity of positive windows, several complementary metrics are reported.

1) Classification Metrics:

From true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), we calculate: precision:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (22)$$

recall, or sensitivity:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (23)$$

specificity:

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (24)$$

the F1-score:

$$F_1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (25)$$

balanced accuracy:

$$\text{Balanced accuracy} = \frac{\text{Recall} + \text{Specificity}}{2} \quad (26)$$

2) ROC AUC and PR AUC:

Because the positive class is very rare, the precision-recall curve is particularly informative. We therefore report:

- ROC AUC, which measures overall discrimination;
- PR AUC, which places greater emphasis on detection of the rare class.

These metrics are calculated at the window level. For an alerting system, they should be supplemented with event-level measures, including detected crises, anticipation time, and the number of false alerts per patient-day [26,27,28,29].

TABLE V
METRICS REPORTED IN THE STUDY

Metric	Formula or definition	Relevance
Precision	TP/(TP+FP)	Proportion of positive predictions that are correct
Recall (sensitivity)	TP/(TP+FN)	Ability to detect pre-crisis windows
Specificity	TN/(TN+FP)	Stability on negative windows
F1-score	Harmonic mean of precision and recall	Trade-off at the selected threshold

A. Main Quantitative Results

TABLE VI
TEST-SET PERFORMANCE FOR THE 30-MINUTE PRE-CRISIS HORIZON

Method	Threshold	ROC AUC	PR AUC	F1	Precision	Recall / sensitivity	Specificity	Balanced accuracy
Logistic regression	0.950	0.999 [0.999, 1.000]	0.787 [0.763, 0.821]	0.652 [0.635, 0.668]	0.492 [0.475, 0.514]	0.968 [0.936, 0.989]	0.997 [0.997, 0.998]	0.982 [0.967, 0.993]
Fuzzy fusion	0.675	0.999 [0.999, 0.999]	0.642 [0.614, 0.689]	0.642 [0.605, 0.670]	0.581 [0.543, 0.620]	0.718 [0.665, 0.779]	0.999 [0.998, 0.999]	0.858 [0.832, 0.889]
Hard thresholds	0.850	0.994 [0.991, 0.996]	0.343 [0.317, 0.363]	0.496 [0.472, 0.525]	0.398 [0.375, 0.421]	0.657 [0.592, 0.732]	0.997 [0.997, 0.998]	0.827 [0.795, 0.865]

On this dataset, the hard-threshold method showed the lowest performance, with a PR AUC of 0.343 and an F1-score of 0.496. Fuzzy fusion achieved a PR AUC of 0.642 and an F1-score of 0.642, with a specificity of 0.999 at the selected threshold. These results show that gradual combination of deviations classifies pre-crisis windows better than simple binary counting. However, the entire improvement cannot be attributed to fuzzy logic alone because the methods do not use exactly the same information.

Logistic regression showed the highest raw discrimination, with a PR AUC of 0.787, an F1-score of 0.652, and a recall of 0.968. At the selected threshold, it detected more positive windows but also generated 216 false positives, compared with 112 for fuzzy fusion. This result describes a sensitivity-precision trade-off in the experiment. It is not sufficient to conclude that alert fatigue would be reduced, which should be evaluated at the level of alert episodes and monitoring duration.

The 95% confidence intervals in Table VI were obtained by resampling patients, not individual windows. This is important because the 216 positive test windows are sparse

Balanced accuracy	Mean of recall and specificity	Comparison under class imbalance
ROC AUC	Area under the ROC curve	Overall discrimination
PR AUC	Area under the precision-recall curve	Performance on the rare class

III. RESULTS AND DISCUSSION

The evaluation includes all 500 profiles. We examine discrimination, results at the selected threshold, score distribution, and calibration. Because the positive class is very rare, PR AUC and the confusion matrices are given greater interpretive weight than overall accuracy [26,29].

After excluding windows during active crises, 321,988 windows remain: 241,146 for training and 80,842 for testing. The split includes 375 training profiles and 125 test profiles. The test set contains 216 positive and 80,626 negative windows, giving a positive prevalence of $216/80,842 = 0.00267$ (approximately 0.27%). For a non-informative ranking, the expected PR AUC is approximately equal to this prevalence (0.0027), which is the baseline shown in Figure 5. The results therefore concern a highly imbalanced window-level task rather than 80,842 independent clinical events [26,27,29].

relative to the 125 test patients, and patient-specific PR AUC or F1 can be undefined for patients with no positive window and highly unstable for patients with very few positive windows. We therefore avoid reporting a potentially misleading unweighted average of per-patient F1 scores. Patient-level bootstrap intervals provide a heterogeneity-sensitive uncertainty estimate for the aggregate metrics, but they do not replace a dedicated patient-stratified analysis of event-level sensitivity and false-alert burden. Such an analysis requires sufficient crisis events per patient and is a priority for external validation.

The confidence intervals support the stability of the reported aggregate values within this held-out cohort, but no paired hypothesis test was performed to claim statistical superiority between fuzzy fusion and logistic regression. The comparison should therefore be interpreted in terms of observed discrimination and operating trade-offs rather than a formal test of method dominance.

B. Distribution of the Fuzzy Score

Figure 3 presents the distribution of the fuzzy score on the test set. Non-pre-crisis windows are concentrated mainly in the lower part of the scale, whereas windows labeled as pre-crisis are located mostly between 0.55 and 0.85. This separation is consistent with the fuzzy rules and parameters selected for the analysis.

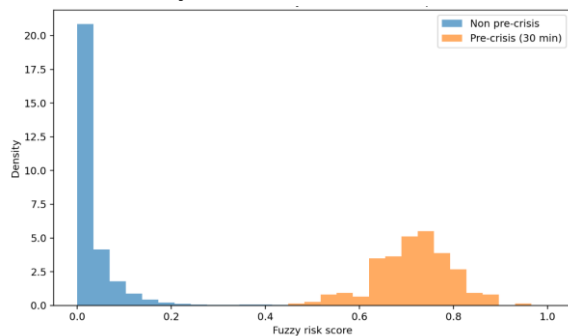


Figure 3. Distribution of the fuzzy score on the test set.

The score therefore provides a continuous ordering of the dataset windows, whereas the hard-threshold method imposes a discontinuity. This property may be useful for prioritization, but it should not be interpreted as a direct measure of clinical proximity to an actual crisis. The observed distribution depends on the characteristics of the dataset and the selected membership functions.

C. Analysis of ROC and PR Curves

The ROC curves in Figure 4 show very high values for all three methods: 0.994 for hard thresholds and 0.999 for both fuzzy fusion and logistic regression. In a problem in which 99.73% of windows are negative, ROC AUC may conceal substantial differences in precision for the rare class [29].

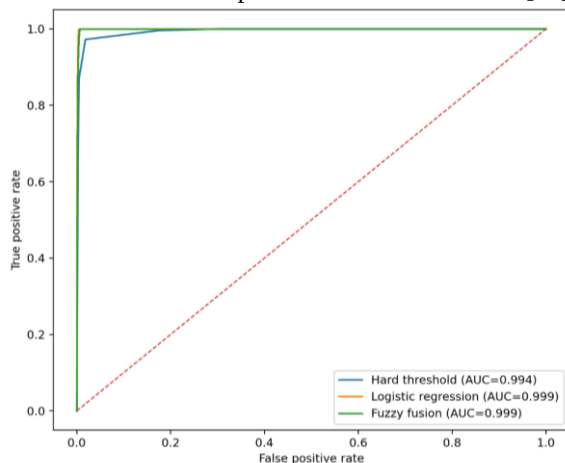


Figure 4. ROC curves for the three compared methods.

Figure 5, based on precision-recall curves, distinguishes the methods more clearly. Logistic regression performs best, fuzzy fusion occupies an intermediate position, and the hard-threshold method performs worst. The horizontal prevalence reference is approximately 0.0027; thus the reported PR AUC values of 0.343, 0.642, and 0.787 are well above a non-

informative ranking for this test set. Because PR AUC is prevalence-sensitive, these values should be compared only with explicit attention to the 0.27% positive-window prevalence [28,29].

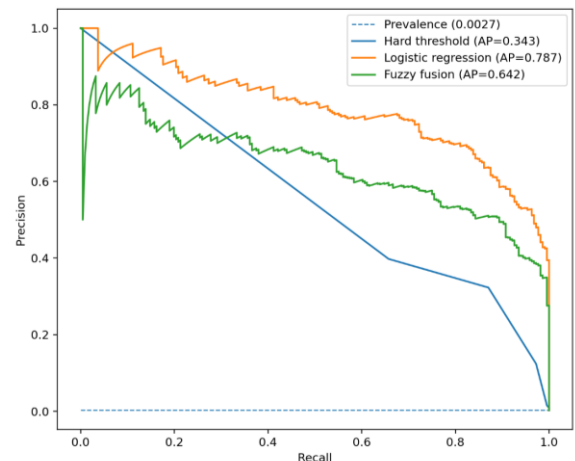


Figure 5. Precision-recall curves for the three compared methods.

D. Confusion Matrices and Operating Trade-Off

The confusion matrices in Figure 6 specify the selected operating point. The hard-threshold method produced 142 true positives, 74 false negatives, and 215 false positives. Fuzzy fusion obtained 155 true positives, 61 false negatives, and 112 false positives. Logistic regression reached 209 true positives and 7 false negatives, with 216 false positives.

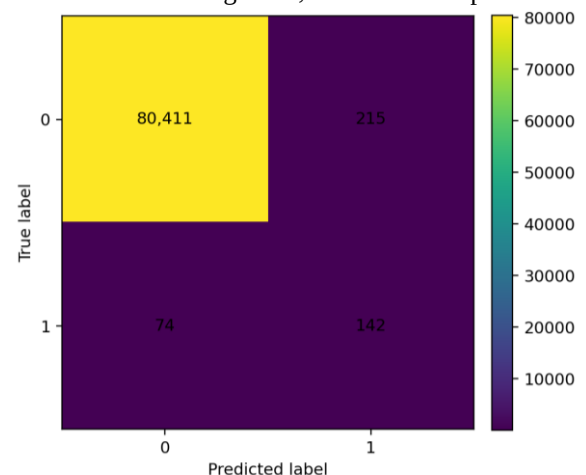


Figure 6a. Confusion matrix for the hard-threshold method.

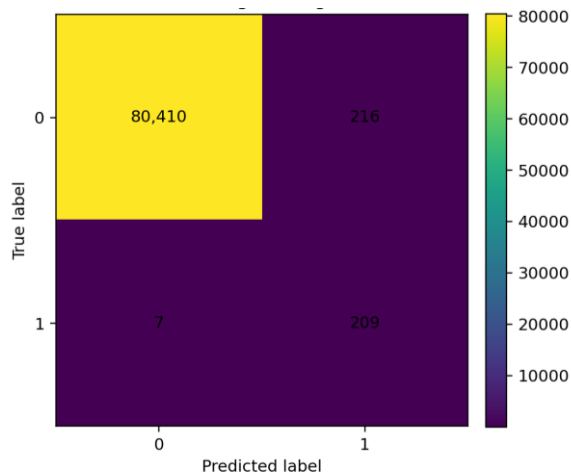


Figure 6b. Confusion matrix for logistic regression.

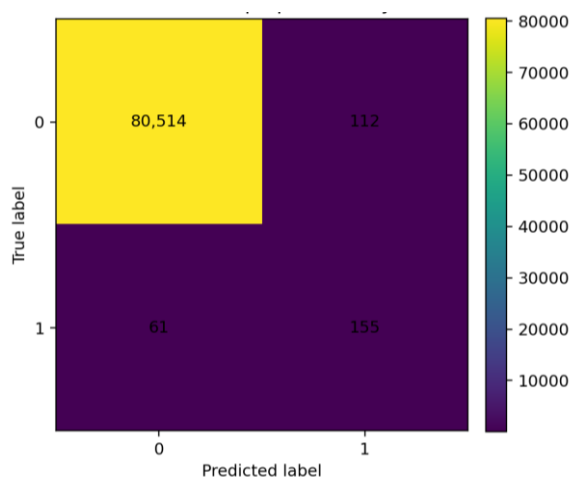


Figure 6c. Confusion matrix for the proposed fuzzy-fusion method.

Compared with hard thresholds, fuzzy fusion increases the number of correctly classified pre-crisis windows (155 versus 142) while reducing false positives (112 versus 215). Compared with logistic regression, fuzzy fusion reduces false positives (112 versus 216) and improves precision (0.581 versus 0.492), but it also misses substantially more positive windows (61 false negatives versus 7) and has lower recall (0.718 versus 0.968). These asymmetric errors have different potential clinical consequences. The present window-level experiment cannot determine whether the reduction in false positives compensates for the increase in missed pre-crisis windows; that question requires event-level alert analysis, anticipation-time measurement, and clinical utility assessment.

E. Score Calibration and Interpretation of the Generated Risk

Figure 7 presents calibration of the fuzzy score. The observed frequency of positive windows remains considerably lower than the mean score values in the higher intervals. The score therefore overestimates risk when interpreted as a probability.

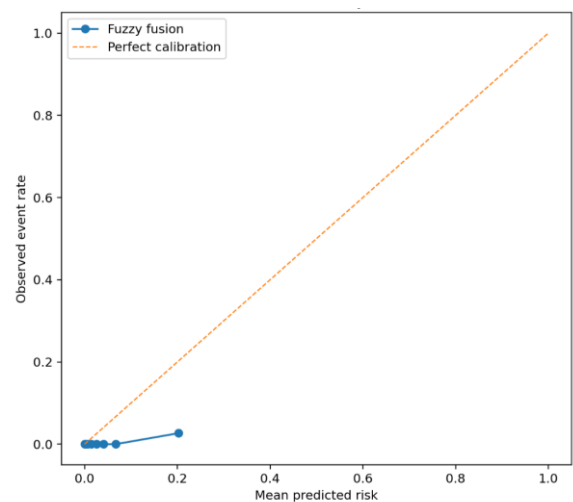


Figure 7. Calibration curve for the fuzzy score.

This observation confirms that $R_i(w)$ should currently be considered an ordinal index. Good discrimination does not guarantee good calibration [29]. Before the score is communicated as a probability, recalibration on independent data would be required, together with indicators such as the Brier score, calibration slope, and uncertainty around the estimates [28,29].

F. Physiological Interpretation of Channel Contributions

Figure 8 shows the mean contributions of the channels to the fuzzy score. In pre-crisis windows, SpO₂, peripheral perfusion, and skin temperature contribute more than heart rate, HRV, and respiratory rate. This hierarchy reflects the characteristics observed in the data and the rule base; it does not constitute a clinical discovery regarding the actual importance of the biomarkers.

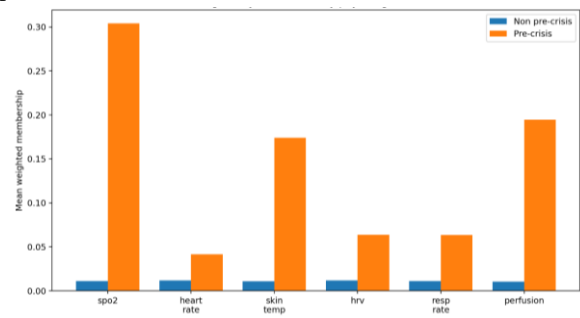


Figure 8. Mean contributions of the variables to the fuzzy score.

The value of this decomposition is primarily methodological. It identifies which channels activate the fuzzy score and checks whether the output follows the intended rule structure. Figure 8 is not an ablation analysis: it does not measure the change in predictive performance after removing each signal or contextual variable. Therefore, the present experiment cannot determine whether the gain over the simple benchmark is genuinely distributed across modalities or dominated by a small subset of predictors. A matched-input ablation and a sensitivity analysis of membership-function breakpoints, rule weights, contextual

coefficients, and signal-quality parameters are required to answer that question.

G. Failure-Case Analysis

At the selected operating point, fuzzy fusion produces 112 false-positive and 61 false-negative windows. The available aggregate results do not permit causal adjudication of every error, so the following mechanisms should be read as model-based hypotheses rather than observed clinical causes. False positives may occur when fever, infection, stress, transient vascular change, or residual motion-related disturbance jointly activate several rules without a crisis beginning in the next 30 minutes. False negatives may occur when a crisis is preceded by weak or atypical changes in the six monitored channels, when relevant physiology is absent from the selected variables, or when the quality gate attenuates a genuine but unreliable signal.

These failure modes directly inform future evaluation. Each predicted alert should be linked to raw sensor traces, quality flags, contextual annotations, and the corresponding clinical record so that errors can be adjudicated at the crisis-event level. Prospective studies should additionally report false alerts per patient-day, the proportion of crises detected, time from first alert to crisis onset, and performance under deliberate channel removal or noise injection. This would distinguish shortcomings of the fuzzy rules from failures caused by sensing, missing context, or the chosen 30-minute horizon.

H. General Discussion

Three findings emerge from the experiment. First, patient-centered oriented deviations provide an individualized representation of change relative to profile baselines. Second, fuzzy fusion performs better than the simple hard-threshold benchmark on PR AUC and F1 while producing fewer false-positive windows at the selected operating point. Third, logistic regression retains the best overall discrimination and substantially higher recall. The contribution of fuzzy fusion therefore lies in an auditable, continuous rule-based trade-off rather than in universal predictive superiority. Because the hard-threshold benchmark does not use exactly the same context and quality information, the comparison does not isolate the causal effect of fuzzy logic itself.

This positioning is consistent with remote-monitoring requirements, where signal quality, understandable alert logic, and integration into clinical workflow are important alongside discrimination [30-33]. Nevertheless, the current evidence remains limited by a single patient-level split, temporal dependence among neighboring windows, one fixed 30-minute horizon, absence of a matched-input ablation, absence of dedicated sensor-degradation experiments, and lack of external or prospective validation. System interpretability is inferred from its explicit structure rather than measured with physicians or patients, and the generated score is not calibrated as a clinical probability.

IV. CONCLUSIONS

This study evaluated a fuzzy-fusion approach for early estimation of vaso-occlusive crisis risk using multiple biometric signals, measurement quality, and contextual variables. The proposed system transforms individualized physiological deviations into a continuous and auditable score that ranks windows occurring within the 30 minutes preceding the onset of a vaso-occlusive crisis.

Results obtained from data for 500 patients show that fuzzy fusion performs better than the simple hard-threshold benchmark on PR AUC and F1-score and produces fewer false-positive windows at its training-derived operating point. Logistic regression nevertheless retains the highest overall discrimination and recall. Because the compared methods do not use exactly identical information, these results do not demonstrate that fuzzy logic alone causes the observed improvement. Fuzzy fusion should instead be viewed as an auditable approach offering a specific trade-off among ranking performance, score continuity, precision, recall, and transparency of the decision mechanism.

These results should be interpreted cautiously. The study uses a fixed 30-minute pre-crisis horizon, a single patient-level train-test split, window-level outcomes with temporal dependence, and prototype fuzzy parameters. The score is an ordinal ranking index rather than a calibrated clinical probability. The findings therefore provide methodological evidence that the proposed pipeline is feasible on this dataset, but they do not establish clinical usefulness, transportability across centers or devices, or prospective crisis prediction performance.

Future work should include prospective, external, and multicenter validation using new longitudinal data from a more diverse population of patients with sickle cell disease. It should compare several anticipation horizons; use repeated grouped or nested validation; quantify patient-level and event-level variability; perform matched-input ablations and sensitivity analyses; stress-test missing, noisy, or unavailable sensors; and report event sensitivity, anticipation time, and false alerts per patient-day. Calibration or explicit abstention should be implemented before a score is communicated clinically, and the usefulness and interpretability of the system should be evaluated directly with healthcare professionals and patients.

REFERENCES

- [1] A. Kumar and S. Bhattacharya, "Sickle cell disease: a comparative perspective on global and national initiatives," *Front. Hematol.*, vol. 3, pp. 1457158, 2024, doi: 10.3389/frhem.2024.1457158.
- [2] P. Shah, S. Sy, M. Chen, M. C. K. Khoo, T. D. Coates, and S. Veluswamy, "Wearable Device Photoplethysmography As a Viable Tool to Longitudinally Monitor Vasoconstriction Biomarkers for Predicting Vaso-Occlusive Crisis in Sickle Cell Disease: Feasibility and Validation Study," *JMIR Hum. Factors*, vol. 13, pp. e75465, 2026, doi: 10.2196/75465.
- [3] J. C. Ayena, A. Bouayed, M. Ben Arous, Y. Ouakrim, K. Loulou, D. Ameyed, I. Savard, L. El Kamel, and N. Mezghani, "Predicting chronic pain using portable devices: a scoping review of sensor capabilities, data security, and standards compliance," *Front. Digit. Health*, vol. 7, pp. 1581285, 2025, doi: 10.3389/fdgh.2025.1581285.

- [4] E. G. Bignami, A. Fornaciari, S. Fedele, et al., "Wearable Devices in Healthcare Beyond the One-Size-Fits All Paradigm," *Sensors*, vol. 25, no. 20, pp. 6472, 2025, doi: 10.3390/s25206472.
- [5] C. Köhler, A. Bartschke, D. Fürstenau, T. Schaaf, and E. Salgado-Baez, "The Value of Smartwatches in the Health Care Sector for Monitoring, Nudging, and Predicting: Viewpoint on 25 Years of Research," *J. Med. Internet Res.*, vol. 26, pp. e58936, 2024, doi: 10.2196/58936.
- [6] N. Barrera, M. Solorzano, Y. Jimenez, et al., "Accuracy of Smartwatches in the Detection of Atrial Fibrillation: A Systematic Review and Diagnostic Meta-Analysis," *JACC Adv.*, vol. 4, no. 11, pp. 102133, 2025, doi: 10.1016/j.jacadv.2025.102133.
- [7] T. F. Machado, F. C. B. Neto, M. S. Gonçalves, C. G. Barbosa, and M. E. Barreto, "Exploring machine learning algorithms in sickle cell disease patient data: A systematic review," *PLoS One*, vol. 19, no. 11, pp. e0313315, 2024, doi: 10.1371/journal.pone.0313315.
- [8] R. Saatchi, "Fuzzy Logic Concepts, Developments and Implementation," *Information*, vol. 15, no. 10, pp. 656, 2024, doi: 10.3390/info15100656.
- [9] A. Campagner, E. M. Biganzoli, C. Balsano, C. Cereda, and F. Cabitza, "Modeling unknowns: A vision for uncertainty-aware machine learning in healthcare," *Int. J. Med. Inform.*, vol. 203, pp. 106014, 2025, doi: 10.1016/j.ijmedinf.2025.106014.
- [10] T. Hasija, M. Kuschel, M. Jackson, S. Dailey, H. Menne, C. Reinsberger, S. Vieluf, and T. Loddenkemper, "Improving portable-based seizure prediction by feature fusion using an explainable growing network," *Artif. Intell. Med.*, vol. 169, art. 103228, 2025, doi: 10.1016/j.artmed.2025.103228.
- [11] X. Xu, J. Li, Z. Zhu, et al., "A Comprehensive Review on Synergy of Multi-Modal Data and AI Technologies in Medical Diagnosis," *Bioengineering*, vol. 11, no. 3, pp. 219, 2024, doi: 10.3390/bioengineering11030219.
- [12] F. Kroner, U. Marikkar, G. Parsons, A. Szmul, and A. Mahdi, "Review of multimodal machine learning approaches in healthcare," *Inf. Fusion*, vol. 114, pp. 102690, 2025, doi: 10.1016/j.inffus.2024.102690.
- [13] F. Demrozi, M. Farmanbar, and K. Engan, "Multimodal AI (MMAI) for next-generation healthcare: data domains, algorithms, challenges, and future perspectives," *Curr. Opin. Biomed. Eng.*, pp. 100632, 2025, doi: 10.1016/j.cobme.2025.100632.
- [14] K. Khurana, S. Mahajan, S. Acharya, et al., "Clinical Biomarkers of Acute Vaso-Occlusive Sickle Cell Crisis," *Cureus*, vol. 16, no. 3, pp. e56389, 2024, doi: 10.7759/cureus.56389.
- [15] C. Dampier, "What is the future of digital tools to help manage pain in sickle cell disease patients?," *Expert Rev. Hematol.*, vol. 17, no. 9, pp. 539–541, 2024, doi: 10.1080/17474086.2024.2388188.
- [16] H. Taherdoost, "Wearable Healthcare and Continuous Vital Sign Monitoring with IoT Integration," *Comput. Mater. Contin.*, 2024, doi: 10.32604/cmc.2024.054378.
- [17] M. D. Mukherjee, P. Gupta, V. Kumari, et al., "Wearable biosensors in modern healthcare: Emerging trends and practical applications," *Talanta Open*, vol. 11, pp. 100486, 2025, doi: 10.1016/j.talo.2025.100486.
- [18] M. K. Siam, et al., "Multimodal Models in Healthcare: Methods, Challenges, and Future Directions for Enhanced Clinical Decision Support," *Information*, vol. 16, no. 11, pp. 971, 2025, doi: 10.3390/info16110971.
- [19] M. Vaida and Z. Huang, "Multimodal graph neural networks in healthcare: a review of fusion strategies across biomedical domains," *Front. Artif. Intell.*, vol. 8, pp. 1716706, 2026, doi: 10.3389/frai.2025.1716706.
- [20] S. Leelavathy, R. Jayaprakash, and I. M. Mary, "Harnessing Fuzzy Logic Framework to Quantify Diagnostic Uncertainty in Medical Decision Support," *Fuzzy Inf. Eng.*, vol. 17, no. 4, pp. 425–446, 2025, doi: 10.26599/FIE.2025.9270070.
- [21] A. El-Ibrahimi, O. Daanouni, Z. Alouani, et al., "Fuzzy based system for coronary artery disease prediction using subtractive clustering and risk factors data," *Intell.-Based Med.*, vol. 11, pp. 100208, 2025, doi: 10.1016/j.ibmed.2025.100208.
- [22] K. Aliyeva and N. Mehdiyev, "Uncertainty-aware multi-criteria decision analysis for evaluation of explainable artificial intelligence methods: A use case from the healthcare domain," *Inf. Sci.*, vol. 657, pp. 119987, 2024, doi: 10.1016/j.ins.2023.119987.
- [23] H. Hoghooghi Esfahani, S. Toyonaga, and K. Oyibo, "The application of explainable artificial intelligence in the prediction, diagnoses, treatment, and management of chronic diseases: A systematic review," *Digit. Health*, vol. 11, pp. 20552076251355669, 2025, doi: 10.1177/20552076251355669.
- [24] Q. Abbas, W. Jeong, and S. W. Lee, "Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges," *Healthcare*, vol. 13, no. 17, pp. 2154, 2025, doi: 10.3390/healthcare13172154.
- [25] Y. Liu, C. Liu, J. Zheng, C. Xu, and D. Wang, "Improving Explainability and Integrability of Medical AI to Promote Health Care Professional Acceptance and Use: Mixed Systematic Review," *J. Med. Internet Res.*, vol. 27, pp. e73374, 2025, doi: 10.2196/73374.
- [26] G. S. Collins, P. Dhiman, J. Ma, et al., "Evaluation of clinical prediction models (part 1): from development to external validation," *BMJ*, vol. 384, pp. e074819, 2024, doi: 10.1136/bmj-2023-074819.
- [27] R. D. Riley, L. Archer, K. I. E. Snell, et al., "Evaluation of clinical prediction models (part 2): how to undertake an external validation study," *BMJ*, vol. 384, pp. e074820, 2024, doi: 10.1136/bmj-2023-074820.
- [28] R. D. Riley, G. S. Collins, L. Kirton, et al., "Uncertainty of risk estimates from clinical prediction models: rationale, challenges, and approaches," *BMJ*, vol. 388, pp. e080749, 2025, doi: 10.1136/bmj-2024-080749.
- [29] B. Van Calster, G. S. Collins, A. J. Vickers, et al., "Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: overview and guidance," *Lancet Digit. Health*, vol. 7, no. 12, pp. 100916, 2025, doi: 10.1016/j.landig.2025.100916.
- [30] M. R. Scheid, B. Friedmann, M. Oppenheim, J. S. Hirsch, T. P. Zanos, et al., "Development and validation of a clinical portable deep learning based continuous in-hospital deterioration prediction model," *Nat. Commun.*, vol. 16, pp. 9513, 2025, doi: 10.1038/s41467-025-65219-8.
- [31] L. P.-Y. Ting, H.-P. Chen, A.-S. Liu, C.-Y. Yeh, P.-L. Chen, and K.-T. Chuang, "Early Detection of Patient Deterioration from Real-Time Wearable Monitoring System," *Proc. IJCAI* 2025, pp. 9871–9879, 2025, doi: 10.24963/ijcai.2025/1097.
- [32] P. Nong, et al., "Current Use And Evaluation Of Artificial Intelligence And Predictive Models In US Hospitals," *Health Aff.*, vol. 44, no. 1, pp. 90–98, 2025, doi: 10.1377/hlthaff.2024.00842.
- [33] Y. Daineko and A. Seitnur, "Artificial Intelligence for Predictive Health Analytics: Challenges, Methods, and Immersive System Integration," *Procedia Comput. Sci.*, vol. 265, pp. 636–641, 2025, doi: 10.1016/j.procs.2025.07.232.