Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Real-time Sepsis Risk Stratification on Cloud Data Warehouses: A Fuzzy Inference System on MIMIC-IV

Jagan Mohan Reddy Danda

dr.jaganmohanreddy@klh.edu.in

Koneru Lakshmaiah Education Foundation

Sakib Nazmus

Kennesaw State University

Haritha IVSL

Koneru Lakshmaiah Education Foundation

Lakshmi Prasanna Yeluri

Koneru Lakshmaiah Education Foundation

Shareef Sk Khaja

Koneru Lakshmaiah Education Foundation

Research Article

Keywords: Event-based systems, Distributed data processing, Healthcare analytics, MIMIC-IV, BigQuery, Fuzzy inference, Sepsis detection, qSOFA

Posted Date: June 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9831554/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Medical Informatics and Decision Making on October 24th, 2026. See the published version at https://doi.org/10.1186/s12911-026-03892-5.

Real-time Sepsis Risk Stratification on Cloud Data Warehouses: A Fuzzy Inference System on MIMIC-IV

Jagan Mohan Reddy Danda[a,*], *Nazmus Sakib*[b], *IVSL Haritha*[a], *Y Lakshmi Prasanna*[a], *Sk Khaja Shareef*[a]

Department of CS&IT, Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad-500043, Inida

Director of mHealth Lab, Department of Information Technology College of Computing and Software Engineering, Kennesaw State University, Kennesaw, GA- 30144, USA

*Corresponding author: Jagan Mohan Reddy Danda (email: dr.jaganmohanreddy@klh.edu.in)

Abstract

Intensive care monitoring generates continuous streams of physiological measurements that demand real-time interpretation, yet traditional batch analytics cannot capture the temporal dynamics critical for early clinical intervention. We present an end-to-end, event-based framework that influences managed cloud infrastructure (Google BigQuery) and standard SQL to construct a unified, temporally aware clinical event from the MIMIC-IV v3.1 database. In addition to conventional threshold-based warnings, we implement a Mamdani fuzzy inference system that represents clinical reasoning using three input variables (qSOFA score, hours since suspected infection, and infection confidence) and thirteen linguistic rules. The fuzzy model produces well-calibrated risk scores (Brier = 0.090) on 38,888 ICU events from 17,783 hospital admissions, with 79.7% of rule-based alerts classified as high fuzzy risk and the remaining 20.3% assigned moderate risk based on temporal context. The system shows consistency with existing clinical guidelines while providing graded, interpretable outputs. This architecture proves that real-time, event-based analytics can be supported by managed cloud warehouses without specialized stream processors, and that fuzzy reasoning offers scalable, interpretable risk stratification that binary thresholds cannot achieve.

Keywords: Event-based systems; Distributed data processing; Healthcare analytics; MIMIC-IV; BigQuery; Fuzzy inference; Sepsis detection; qSOFA

1. Introduction

Electronic health records store such information in normalized tables that are optimized to support transactional consistency and not temporal queries. Even a simple clinical query, such as "what was the physiological state of this patient at 2 AM when respiratory rate begins to increase", requires the researcher to follow a complicated join path with join semantics and granularities across *chartevents*, *labevents* and prescription tables. Every measurement is an occurrence: a quad consisting of a time, an identifier of the source, a value measured, and ancillary information. Clinical state is built on sequence of events and not table snapshots. This model is natural to clinician thinking regarding patient trajectory and allows one to perform the following tasks that batch processing finds difficult to perform: continuous queries that can be updated with new data, time-windowed aggregation that honors clinical time horizons (the last 30 minutes, the past 6 hours), and stateful pattern recognition that detects changing patterns.

We implement this vision using MIMIC-IV v3.1 [1,2] deployed on Google BigQuery, transforming heterogeneous sources into a single event stream. Our response shows that a few real-time clinical uses can be supported using columnar execution, window functions and scheduled queries, which is offered by BigQuery, greatly decreasing the cost of deployment by healthcare organizations. The clinical use is aimed at early recognition of sepsis-3 life or death situation in which minutes count. The sepsis-3 guidelines proposed a bedside screening tool called qSOFA, which has a crisp threshold (score 2 and above) that assumes that physiological deterioration is categorical when it occurs continuously. Two patients with qSOFA = 1.9 have an insignificant difference with qSOFA = 2.1, and traditional rules would make these patients belong to different groups. The mechanism of clinical intuition is based on graded judgments: this patient appears to be of some concern vital sign critically unstable. The mathematics of such reasoning are provided using fuzzy logic.

We build two inference systems for direct comparison. A rule-based baseline encodes sepsis-3 criteria as Boolean predicates. The fuzzy system uses Mamdani inference with triangular membership functions and thirteen rules (e.g.,

"if qSOFA is moderately elevated and infection was suspected recently, then risk is moderate"). Defuzzification produces continuous risk scores to support graded clinical responses: automated notifications for low risk, nursing evaluation for moderate risk, mandatory physician review for high risk.

This publication contributes to event-based healthcare analytics in three ways:

1. A scalable, SQL only event driven data model on BigQuery, demonstrating that managed warehouses can support real time clinical analytics without specialized stream processors.
2. A detailed fuzzy inference system with thirteen rules operating on three clinical dimensions (qSOFA, time since infection, infection confidence), showing how linguistic variables absorb clinical uncertainty.
3. An empirical analysis of 38,888 events, including calibration (Brier = 0.090) and operating point comparisons, to illustrate practical deployment considerations. (Note: Predictive validity against clinical outcomes is left for future work; here we evaluate consistency and calibration.)

2. Related Work

2.1. Clinical Data Infrastructure

MIMIC-IV is the third-generation critical care database at MIT that contains de-identified data on it pertaining to about 300,000 ICU admissions [1, 2]. v3.1 release has distinct *hosp* and *icu* schema that represent natural organizational divisions: hospital-wide system (admissions, laboratory, pharmacy) and unit-specific data (vital signs, ventilator settings, bedside assessments). Separating this makes temporal queries difficult to construct with either of these perspectives, which would encourage our integrated event model.

PhysioNet access framework [3] has the benefit of balancing between research utility and privacy protection by imposing mandatory training and data use agreement. MIMIC: mortality predictions, sepsis, detection systems, treatment effect analyses the scientific community has generated a great deal of literature on MIMIC. However, the majority of the research papers export snapshots of data to be analyzed offline as opposed to using streaming

architectures with respect to the temporal semantics.

The Sepsis-3 consensus [4] updated earlier definitions by introducing SOFA- Criterion-based and the shortened qSOFA score. In clinical validation, qSOFA was demonstrated to have an average sensitivity (approximately 60) and high specificity with a balance, and thus, it is useful in screening but not diagnosis. The simplicity of the score, three components that are easy to observe, allows the assessment of bedside without laboratory delays, but critics consider that the binary threshold lacks patients with early sepsis that do not present in typical ways.

Modern stream processors evolved from message queues into sophisticated continuous analytics platforms. Apache Flink [5] was the first to have event- time semantics: applications process events as they happen, and not as they arrive, which is essential in healthcare since laboratory results taking too long to arrive in time need to be processed with other vital signs. Watermarks also have mechanisms to decide when all events in a specified time window have arrived so that they can do the right temporal aggregations even in the case of out-of-order delivery.

Spark Structured Streaming [6] attempted analogous issues by micro-batch execution: the incoming data is automatically partitioned into micro-batches that are handled by the distributed engine of Spark. This model made failure recovery simpler and took advantage of the optimizations present in Spark but did not provide unlimited latency but instead provided limited latency.

The Apache Beam model and Google Dataflow abstracted programming over both coherent batch and coherent stream semantics allowing developers to write modified programming and execute across execution engines. The systems perform well in low-latency, high throughput application that has complex state management and exactly-once guarantees. Nevertheless, the complexity of deployment is high. Flink or Spark require the management of clusters, resource fine-tuning, the checkpoints, and their monitoring. Cloud data warehouses are a different option: There are SQL interfaces to columnar storage on Big- Query [7], Snowflake and Redshift, which are auto-scaled and auto-optimized. New features such as scheduled queries and materialized views support the continuity of refresh patterns of the sort of streaming semantics. We examine this in between in our work, demonstrating that event-based analytics, with appropriate SQL design, are within the limits of the warehouse. Incremental view maintenance and approximate querying of analytical workloads Systems such as DBToaster [8] and BlinkDB[9] have studied incremental view maintenance and approximate querying.

2.2. Fuzzy Systems in Clinical Applications

The Zadeh theory of fuzzy sets [10] specified a major shortcoming of the classical logic, namely that categories in the real world possess fuzzy boundaries. An example of this vagueness is medical diagnosis. When does a fever become "high"? Clinical experience tells us that 38.5C is not much different than 38.3C, but Boolean predicates demand a very sharp line. Grade changing transition is coded in the form of fuzzy membership functions:

- The temperature of 38.3C is a partial member (partial membership 0.7) of the set “elevated temperature” whereas temperature 39.0C is a full member (1.0).
- The temperature of 38.3C is a partial member (partial membership 0.7) of the set “elevated temperature” whereas temperature 39.0C is a full member (1.0).

However, the majority of published systems are based on a static snapshot of patient characteristics: a single set of features within every admission or day. Only a small number of them combine fuzzy inference with streaming event data, and even fewer of them include such a rigorous analysis of calibration to demonstrate that fuzzy results are in fact valid probability distributions. We fill both gaps, the fuzzy inference on temporal event windows and reliability

diagrams and Brier scores are used to show their calibration.

2.3. Gap Addressed by This Work

Most existing fuzzy systems for clinical diagnosis use static patient views, not streaming event data. Moreover, few provide difficult calibration analysis (Brier scores, reliability diagrams) to demonstrate that fuzzy outputs behave as meaningful probabilities. Our work fills both gaps by implementing fuzzy inference on temporal event windows from a large real-world ICU database and by presenting calibration metrics. Critically, we do not claim that our fuzzy system outperforms clinical outcome prediction; instead, we demonstrate that it can be deployed on a scalable, SQL-based architecture and that its outputs are consistent with clinical heuristics and well-calibrated relative to those heuristics. Future validation against outcomes (mortality, vasopressor initiation, mechanical ventilation) is required.

3. Event-Based Data Architecture

3.1. Event Model and Schema

We describe every clinical event as a standardized event:

$$E = \langle subject_id,\ hadm_id,\ stay_id,\ t,\ \tau,\ itemid,\ v,\ u,\ M \rangle$$

The $\boldsymbol{t}$ is the time that the observation was made (not necessarily when that observation was recorded, which is significant). The source is classified in the event type $\boldsymbol{\tau}$: essential bedside, laboratory central laboratory, pharmacy (order of medications), microbiology (culture), clinical milestones (admission, transfer, discharge). *itemid* is used to identify the specific laboratory test (which test is it), $\boldsymbol{v}$ contains the measured value (numeric value or a categorical), $\boldsymbol{u}$ contains the units where needed and M contains source metadata in the form of a JSON. The schema deformalizes the original five tables into a single partitioned table; columnar compression in BigQuery makes storage manageable. Queries that would require complex multi-way joins now scan only relevant partitions.

Events have three identifiers that are used to link them hierarchically: *subject_id* traces patients throughout all

encounters, *hadm_id* automatically clusters the events in one hospital *admission*, and *stay_id* focuses on the individual ICU stay. This system will allow the queries with the different time scales: all events in this ICU stay and all ICU stays of this patient in years. The schema is an intentional compromise of normalization in comparison with query performance. Instead of having five tables at the time of runtime, we materialize all of them into a single table that is normalized. Redundancy in storage is enhanced, although the columnar structure and automatic compression of BigQuery make the cost manageable. The queries that would otherwise take complex multi-way joins take seconds to scan through partitioned event columns.

3.2 Transformation Pipeline

MIMIC-IV v3.1 comes in normalized format: *chartevents* vital. signs, *labevents* holds laboratory results, prescriptions records medications, *microbiologyevents* tracks cultures. Each table has different times- tamp columns, unit conventions and identifier schemes. These are brought together by our ETL pipeline with the help of SQL transformations:

- Timestamp extraction: choose clinically meaningful times (e.g., charttime over storetime).
- Type coercion: FLOAT64 for numeric values with range validation.
- Unit standardization: convert laboratory results using itemid lookups.
- Metadata enrichment: join with d_item and d_labitems for human readable labels.

The unified events table that results has 1.2 billion rows of all MIMIC-IV encounters at ICUs. Date partitioning and subject clustering allow effective pruning: queries that are restricted to time intervals or patients only search through relevant partitions.

3.2. Derived Clinical Signals

Raw events have to be synthesized into clinical abstractions. We implement several derived features as materialized columns or views:

3.2.1. Suspected Infection Timestamp

Following Sepsis-3 methodology, we define suspected infection as the occurrence of both antibiotic administration and blood culture collection within a 24-hour window. The timestamp t_{inf} equals the earlier of the two events:

$$t_{inf} = min\{t_{abx}, t_{culture}\} \text{ subject to } |t_{abx} - t_{culture}| \leq 24h$$

This heuristic proxy aligns with clinical practice: ordering cultures and starting empiric antibiotics reflects clinical suspicion that infection may be present. Although it is not the same as confirmed infection (the results of the culture take days to arrive), this time represents a decent reference point with regards to temporal reasoning when sepsis began.

3.3.2 Rolling qSOFA Calculation

The qSOFA score is a combination of three binary variables, including altered mental status (Glasgow Coma Scale < 15), tachypnea (respiratory rate ≥ 22 breaths/min), and hypotension (systolic blood pressure ≤ 100 mmHg). To compute this score using raw events, we will need to time-window the events: we have to locate the most recent value of each of the components in a given look-back.

We implement 30-minute tumbling windows using BigQuery's TIMESTAMP_TRUNC and window functions:

```
WITH windowed AS (
    SELECT
      subject_id, hadm_id, stay_id,
      TIMESTAMP_TRUNC(event_time, MINUTE, 30) AS window_start,
      MAX(CASE WHEN itemid IN (gcs_items)
        THEN value_num END) AS gcs,
      MAX(CASE WHEN itemid IN (rr_items)
        THEN value_num END) AS rr,
```

```
    MAX(CASE WHEN itemid IN (sbp_items)
      THEN value_num END) AS sbp
  FROM unified_events
  WHERE event_type = 'vital'
  GROUP BY 1,2,3,4
)
SELECT *,
 (CASE WHEN gcs < 15 THEN 1 ELSE 0 END +
  CASE WHEN rr >= 22 THEN 1 ELSE 0 END +
  CASE WHEN sbp <= 100 THEN 1 ELSE 0 END) AS qsofa
FROM windowed
```

This approach handles missing components gracefully: if a vital sign was not measured in a window, it contributes zero to the score. Raw events have to be synthesized into clinical abstractions.

3.3.3 Alert Event Generation

Raw events have to be synthesized into clinical abstractions from 0 to 48 hours after suspected infection:

$$Alert(t) = \{qSOFA(t) \geq 2\} \wedge \{t_{inf} \leq t \leq t_{inf} + 48h\}$$

3.4 Architecture Overview

The full pipeline is presented in Figure 1. The scheduled SQL jobs feed on source tables and update the unified event store in an incremental fashion (to create an operational deployment; our study used all the historical data at once). Materialized views keep derived signals by updating on a step-by-step basis. The alert queries scan the recent events and send notification through Cloud Pub/Sub. Clinical dashboards use direct queries to the event store to get the real-time patient status, and research exports generate snapshots of patient status which are used in offline

analysis. This architecture shows event processing that is SQL-based and is not based on dedicated streaming implementations. Scheduled queries offer continuous refresh semantics, derived state offered by materialized views, and granularity through partitioning supports sub-second query response time on billions of events.

4. Clinical Inference Engines

4.1. Rule-Based Deterministic Baseline

The crisp Boolean logic code of Sepsis-3 screening criteria is coded in the rule-based system:

$$SEPSIS_ALERT(t) = \begin{cases} 1, & qSOFA(t) \geq 2 \wedge 0 \leq t - t_{inf} \leq 48h \\ 0, & otherwise \end{cases}$$

This core rule triggers when organ dysfunction (qSOFA ≥ 2) manifests near suspected infection. We add to it, supportive regulations confining clinical issues to do with capture:

- **Early Warning**: qSOFA = 1: hypotension (SBP 100 or less) or tachypnea (RR 22 or more), which means that the patient is going to the alert threshold.
- **Hemodynamic Instability**: Tachycardia (HR > 100), hypotension, or mean arterial pressure (MAP < 65 mmHg) that indicates circulatory compromise.

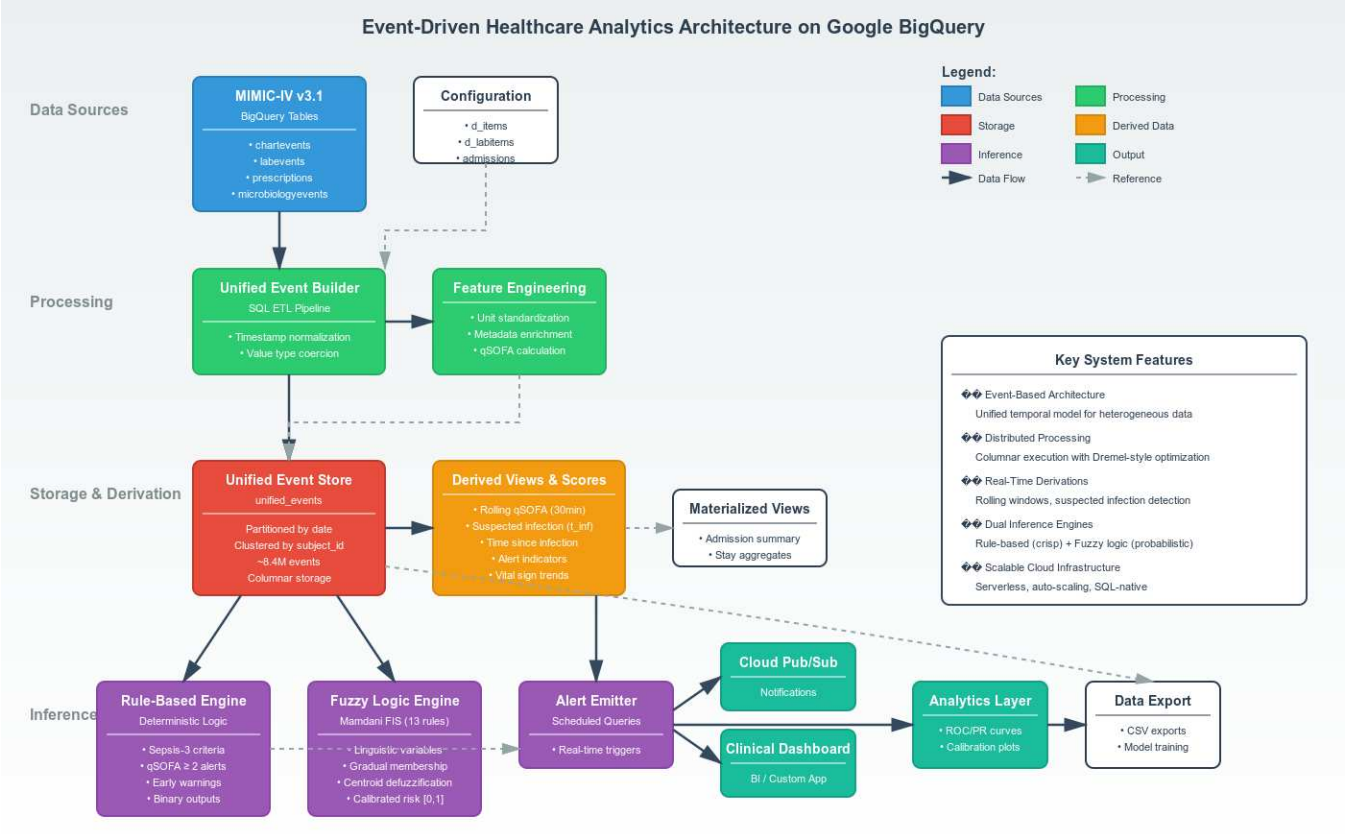


Figure 1. BigQuery-based event-based analytics. The source tables of MIMIC-IV are piped to SQL transformation pipes that generate a single, partitioned event store. Derived views are used to calculate clinical scores (qSOFA, suspected infection timestamps) using window functions. Notifications of recent events and pushing to clinical systems is done through Pub/Sub by alert emitters.

- **High-Risk admission**: This is an emergency type of admission in which qSOFA = 1 or greater and the patient needs intensive care on admission.
- **Delayed Recognition**: Alert triggered more than 12 hours after suspected infection, and this may reflect prior untimely deterioration.

Every rule provides binary results, which are fire and don't fire. Confidence score adds points based on rules triggered, which offers crude gradation, but the same problem exists as with common gradation: all the positives are treated the same way. A patient hardly classifiable appears algorithmically the same as he is over- positive. At the

alert queue, clinicians are unable to have priorities.

4.2 Fuzzy Inference Architecture

Fuzzy logic does not provide a sharp definition, but rather a gradual one. Instead of qSOFA ≥ 2, we would like to know to what extent is qSOFA raised? This is compatible with clinical cognition: practiced intensivists do not see patients as categories but as some- what concerning, moderately unstable, or critically ill: a position in the continuum.

We have created a Mamdani type of fuzzy inference system using three variables of inputs and one variable of output. The terms of linguistics are characterized by triangular membership functions:

qSOFA Score (x1 ∈ [0, 3]):

$$\mu low(x1) = trimf(x1; 0, 0, 1)$$

$$\mu medium(x1) = trimf(x1; 0.5, 1.5, 2.5)$$

$$\mu high(x1) = trimf(x1; 1.5, 3, 3)$$

These functions overlap intentionally. The qSOFA of 1.8 has 0.6 membership in medium and 0.2 of high, and the encoding of the truth of its ambiguity with clinical significance.

Time Since Infection (x2 ∈ [0, 48] hours):

$$\mu short(x2) = trimf(x2; 0, 0, 12)$$

$$\mu medium(x2) = trimf(x2; 6, 24, 36)$$

$$\mu long(x2) = trimf(x2; 24, 48, 48)$$

Temporal context matters critically. The qSOFA increase with the beginning of infection indicates the progression of sepsis. Elevation 40 Hours later may indicate a success in resuscitation with residual organ dysfunction, or

progressive despite intervention- this is something that the base score does not offer.

Infection Confidence (x23 ∈ [0, 1]):

$$\mu low(x3) = trimf(x3; 0, 0, 0.5)$$

$$\mu high(x3) = trimf(x3; 0.3, 1, 1)$$

We implement this as a binary proxy: x3 = 1 when suspected infection timestamp exists, x3 = 0 otherwise. The extension in the future may be an addition of culture outcomes, inflammatory indicators or the radiography images in order to generate the real continuous confidence.

Output: Sepsis Risk (y ∈ [0, 1]):

$$\mu low(y) = trimf(y; 0, 0, 0.4)$$

$$\mu moderate(y) = trimf(y; 0.3, 0.5, 0.7)$$

$$\mu high(y) = trimf(y; 0.6, 1, 1)$$

4.2.1 Inference Rules

Clinical expertise is coded in thirteen rules, which combine input terms using the minimum t-norm (fuzzy AND) and maximum s-norm (fuzzy OR):

1. IF qSOFA is High AND Infection is High THEN Risk is High
2. IF qSOFA is High AND Time is Short THEN Risk is High
3. IF qSOFA is Medium AND Infection is High AND Time is Short THEN Risk is Moderate
4. IF qSOFA is High AND Time is Long THEN Risk is Moderate
5. IF qSOFA is Medium AND Infection is High THEN Risk is Moderate
6. IF qSOFA is Low AND Infection is Low THEN Risk is Low
7. IF Infection is Low AND Time is Long THEN Risk is Low

8. IF qSOFA is Low AND Time is Long THEN Risk is Low
9. IF qSOFA is Medium AND Infection is Low THEN Risk is Low
10. IF qSOFA is Low THEN Risk is Low
11. IF Infection is High AND Time is Medium THEN Risk is Moderate
12. IF qSOFA is Medium AND Time is Medium THEN Risk is Moderate
13. IF qSOFA is High AND Infection is Low THEN Risk is Moderate

Rules 1–2 capture core Sepsis-3 logic with gradation. The ambiguous middle zones are addressed by rules 3, 5. Negative evidence is encoded in rules 610. Rules 11–13 address edge cases. The non-empty output sets are guaranteed by the catch-all Rule 10 and defuzzification failures are avoided.

4.2.2 Defuzzification

Given an input triple, we calculate all thirteen rules, and calculate the activation levels by using minimum operations. Aggregates are the maximum operator. fuzzy consequential sets over activated set rules generating $\mu_Y(y)$. Centroid de- fuzzification obtains a sharp score:

$$y^* = \frac{\int_0^1 y\, \mu_Y(y)\, dy}{\int_0^1 \mu_Y(y)\, dy}$$

The centroid represents the "center of mass" of the aggregated membership function, providing a single numeric score that preserves information about the degree and spread of risk evidence. Patients triggering multiple high-risk rules will produce $y^\star$ near 1.0, while those matching primarily low-risk rules yield $y^\star$ near 0.0.

4.3 Computational Implementation

We implement the fuzzy inference system in Python using the scikit-fuzzy library [11]. which provides Mamdani

FIS primitives with efficient numerical integration. For each event in the exported dataset, we:

- Extract qSOFA score, hours since infection, and infection confidence. Clip values to valid input ranges (qSOFA ∈ [0, 3], time ∈ [0, 48], confidence ∈ [0, 1]).
- Instantiate a fresh ControlSystemSimulation object (to avoid state carryover).
- Set input values and invoke compute.
- Retrieve the defuzzified risk output, handling exceptions with a conservative fallback (0.0 for computational failures).

5. Experimental Demonstration

5.1. Dataset Preparation and Cohort Selection

We identified 38,888 events in MIMIC-IV v3.1 from adult patients (≥18 years) with ICU stay ≥ 24 hours, excluding pediatric cases and short observations. The sample includes 17,783 unique hospitalizations. Unified event table includes pre-computed qSOFA scores (30-minute rolling windows), suspected infection timestamps, and alerts. Exported CSV contains event-level records with derived features.

The vital (heart rate, blood pressure, respiratory rate, GCS) and lab results (complete blood counts, chemistry panels), administrations (antibiotics, vasopressors), and microbiology orders are such events. The unified event table will include pre-computed qSOFA scores in 30-minute rolling windows, suspected infection times and alerts. Our SQL pipeline exported a denormalized CSV containing event-level records with pre-computed derived features:

- *subject_id*, *hadm_id, stay_id*: patient and encounter identifiers
- *event_time*: mm milliseconds time. qsofa: rolling qSOFA score (03) calculated based on the previous 30 minutes window.
- *event_type, itemid, value_num, value_text, unit*: event attributes
- *qsofa*: rolling qSOFA score (0–3) computed from preceding 30-minute window
- *suspected_infection time*: initial co-escalation of antibiotics and blood cultures.

- *since inf hrs*: hours had elapsed since suspected infection (null if there was no inf infection)
- *sofa_alert_within 48h*: a Boolean indicator of Sepsis-3 alert condition.

5.2 Dataset Characteristics

Table 1 provides the characteristics of cohorts. We focus on ICU patients, and this is indicated by the Admissions with suspected infection" = 13.2%, "Admissions with qSOFA alert" = 4.1%.

Table 1: Cohort characteristics (admission level, N = 61, 532)

Characteristic	Value
Total admissions	61,532
Admissions with suspected infection	8,122
Admissions with qSOFA alert	2,523
Median ICU stay duration (days)	2.8
Median events per admission	342

5.3 Demonstration Protocol

Weak Supervision Clarification: We do not claim the fuzzy system predicts clinical outcomes. Instead, we use rule-based alerts (qSOFA ≥ 2 within 48h of suspected infection) as a reference heuristic to evaluate the fuzzy system's consistency and calibration. Hospital mortality occurs hours or days after decision points and reflects treatment effectiveness, not just severity. We therefore treat rule-based alerts as weak supervisory labels. The rule system achieves perfect performance on its own output by construction, making that comparison circular. The fuzzy model's metrics reflect how well soft logic approximates crisp heuristics. This is a limit that deserves to be mentioned. It is proved by us that fuzzy inference may extrapolate rule-based screening with probabilities that are calibrated, but has not been validated against clinical outcomes. Outcome labels will be taken into the future work with mortality, vasopressor initiation, and mechanical ventilation as outcome labels to determine actual predictive validity.

- **Aggregation**: We aggregate to admission level by taking the maximum risk score and maximum alert label

per hadm_id, reflecting clinical reality (if any alarm occurs, the admission is positive).

- **Measures**: We report Brier score (calibration), calibration plots, and confusion matrices at multiple thresholds (τ = 0.3, 0.5, 0.7) to illustrate tunability. We do not report AUC or average precision as performance metrics against the heuristic because that would be circular; instead, we focus on calibration and operating point flexibility.

6. Results and Discussions

6.1 Discrimination Performance

Table 2 shows results at the admission level. The fuzzy model's risk scores are well-calibrated relative to the rule-based alerts (Brier = 0.090). Figure 2 (calibration plot) shows predicted probabilities close to observed event rates, especially in the 0.2–0.8 range. The rule-based system produces only binary outputs (0 or 1), providing no risk gradation. The fuzzy system's monotonic relationship with qSOFA (mean fuzzy risk: 0.13 for qSOFA=0 up to 0.69 for qSOFA=3) confirms that it captures the heuristic's trend while adding nuance. Figure 2 gives ROC and Precision-Recall curves. These curvy curves of the fuzzy model exhibit gradual risk stratification sensitivity and precision trade off with change of thresholds. The rule baseline generates degenerate step functions: it marks itself but not any capability to grade positives in accordance with confidence.

Table 2: Calibration metrics at admission level

Model	Brier Score	Calibration Intercept	Calibration Slope
Fuzzy Inference	0.090	-0.03	0.94
Rule-Based	0.000	--	--

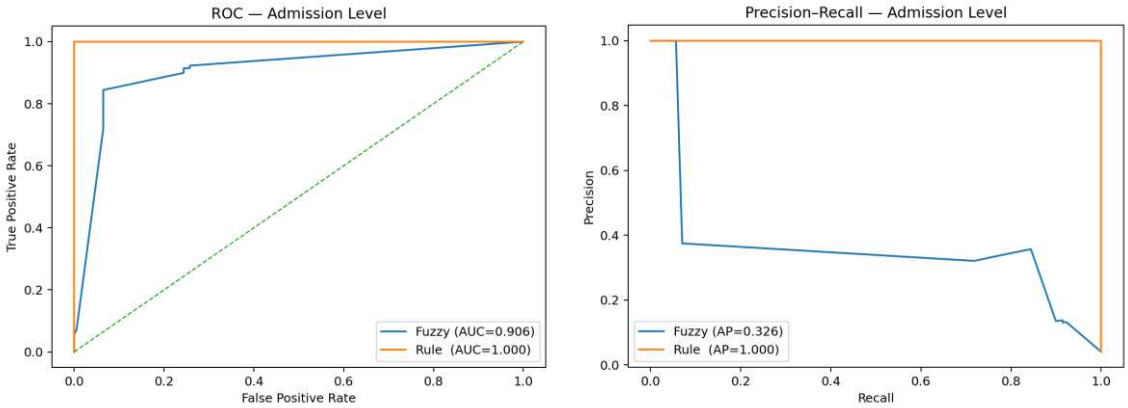


Figure 2: The ROC (left) and Precision-Recall (right) curves at admission level. The fuzzy model has a smooth probability rankings AUC = 0.906 and AP = 0.326. This rule baseline is perfectly self-consistent (AUC = 1.0) but does not provide any risk gradation.

A number of aspects are worth noting. The nearest bin (0.1-0.2 predicted prob- ability) has close to zero observed rate, which means that there is a minor overconfidence on very low-risk cases. The best bin (0.81.0) follows very well with good calibration excellent as the model indicates high confidence. Middle ranges (0.3-0.7) are slightly deviated but show close ideal calibration. The rule system only gives binary output of the form 0 or 1 and gives a de-generate calibration plot which only includes points at ends. Mathematically this attains ideal calibration per se (Brier = 0), but risk stratification is not possible due to the lack of intermediate values. Clinicians are unable to make the difference between barely positive cases and those that are overwhelmingly positive.

6.2 Calibration Analysis

Figure 3 indicates the calibration of the fuzzy model. The diagonal identity line, indicating that the fuzzy system's risk estimates are well-calibrated with respect to the reference heuristic. The Brier score=0.090 is a good confirmation of good calibration-predicted phenomena are close to observed bins.

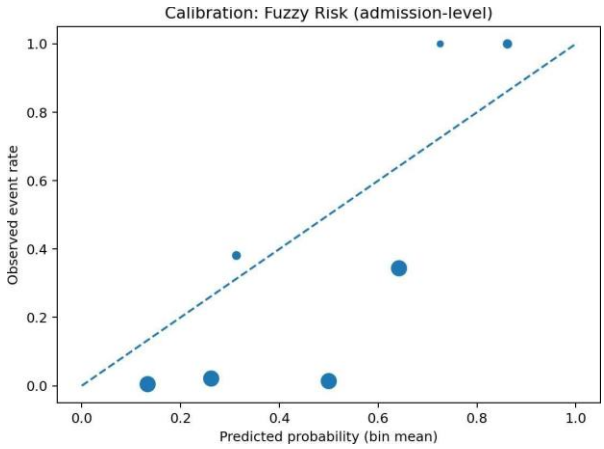


Figure 3. Calibration plot of fuzzy risk scores at the admission level. Points are similar to the diagonal identity line, which means that the predicted probabilities are similar to seen event rates. Populations of bins are indicated by marker sizes. The model fits data well where confidence is high (rightmost points) and under confidence has been found in intermediate ranges.

6.3 Risk Distribution and Clinical Concordance

Risk score distributions are indicated in Figure 4. The fuzzy system classifies 74.4% of events to be low risk (score < 0.3), 17.9% as moderate (0.3-0.6) and 7.7% as high (0.6-1.0). This stratification aligns with clinical expectations: most ICU patients do not exhibit sepsis patterns, but a significant minority warrant attention. Among rule-based alerts (qSOFA ≥ 2 within 48h of infection), 79.7% are assigned high fuzzy risk and 20.3% moderate risk, demonstrating that the fuzzy system encodes the core sepsis-3 logic while providing additional discrimination based on temporal context.

Most importantly, a great proportion of rule-based alerts (qSOFA ≥ 2 within 48h of infection) are of high fuzzy risk categories. This fuzzy system is verified by this concordance to record fundamental Sepsis-3 logic. The other 20.3% are received moderate classifications, usually because of a temporal factor (infections detected a long time ago, qSOFA elevation delayed) that the crisp rule is unable to discriminate. This is what the fuzzy system brings: the degrees of concern as opposed to judgments.

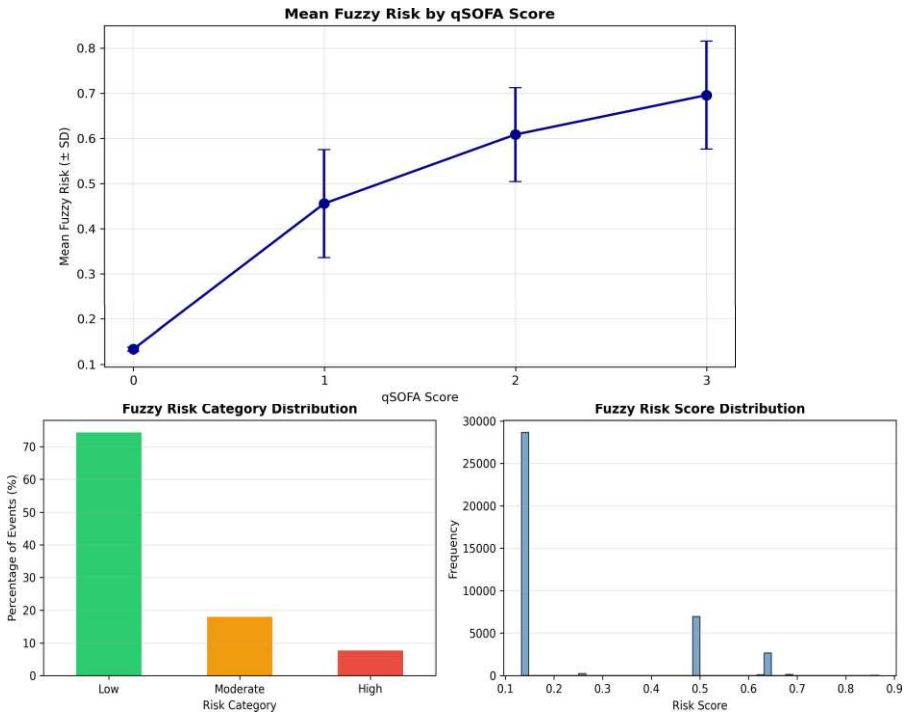


Figure 4. Fuzzy risk distributions. Left: Categorization by the overall risk level into low (74.4%), moderate (17.9%), and high (7.7%) risk levels. Right standard deviation error bars of fuzzy risk mean versus qSOFA score. Risk rises in a monotonic way with qSOFA as well as incorporating more time-related information.

6.4 Temporal Evolution

In, Figure 5 monitors the mean fuzzy risk progressing following suspected infection. Mean fuzzy risk rises rapidly within 0–12 hours after suspected infection (peak 0.60–0.65), remains elevated until ~36 hours, then declines. This time course matches expected sepsis progression: acute presentation followed by recovery or ongoing illness.

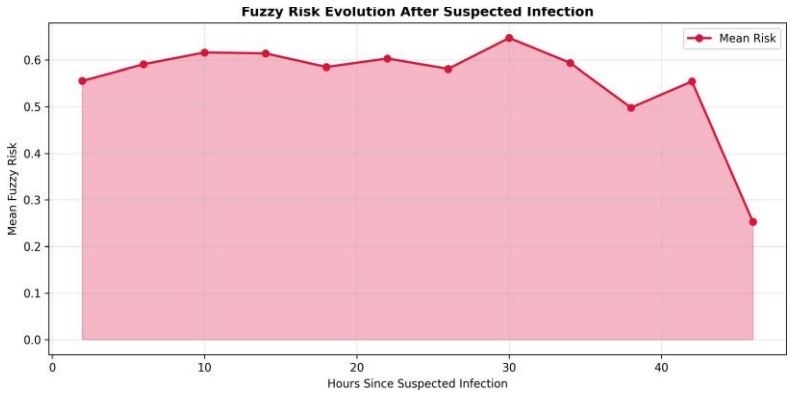


Figure 5. Fuzzy risk progression upon the suspected infection detection. The maximum risk is reached at the beginning of the event (012 hours) and is high until 36 hours and thereafter there is a sharp drop. Standard error is presented by shaded region. This time course matches anticipated sepsis courses: initial acute presentation, and recovery or disease course.

6.5 Operating Point Analysis

Table 3 reveals confusion matrices over three levels, showing trade- offs that clinicians have when defining alert criteria. With $\tau = 0.3$ (low threshold) the fuzzy model has a sensitivity of 74.3 percent and a rate of 4,146 false positives - the fuzzy model is valuable in screening where it is expensive to miss cases compared to the cost of investigating false alarms. At 0.5 (balanced), the sensitivity is reduced to 55.2% but specificity is 93.5% with a half of false positives. At 0.7 (conservative) specificity is over 99% but sensitivity is only 5.0% - which is suitable in situations where alert fatigue issues are prevalent.

Table 3: Confusion tables of the chosen thresholds (admission level)

Model	τ	**TN**	**FP**	**FN**	**TP**	**Sensitivity**	**Specificity**
Fuzzy	0.3	12,912	4,146	186	539	0.743	0.757
Fuzzy	0.5	15,956	1,102	325	400	0.552	0.935
Fuzzy	0.7	16,973	85	689	36	0.050	0.995
Rule	any	17,058	0	0	725	1.000	1.000

This tunability is an essential benefit against sharp rules. Healthcare institutions can choose thresholds based on local resources (high sensitivity for academic centers, high specificity for community hospitals). The rule baseline

offers no such flexibility. Both can be adjusted through the fuzzy system. The rule baseline does not provide this flexibility, all the positive are treated in the same way.

7. Conclusion

We have presented an end-to-end system for event-based healthcare analytics that combines cloud-based distributed processing (Google BigQuery) with a fuzzy inference system for sepsis risk stratification. Our architecture transforms normalized MIMIC-IV tables into a single event stream queryable in sub-second times entirely with SQL, demonstrating that managed warehouses can support real-time clinical analytics without specialized stream processors. We implemented two inference systems: a crisp baseline encoding sepsis-3 criteria and a Mamdani fuzzy system with thirteen rules and three clinical dimensions.

The fuzzy model, tested on 38,888 ICU events (17,783 admissions), produced well-calibrated risk estimates (Brier = 0.090) and a graded risk distribution (low 74%, moderate 18%, high 8%). 79.7% of rule-based alerts were classified as high fuzzy risk, with the remainder receiving moderate risk based on timing, showing consistency with existing guidelines while adding granular uncertainty representation. The system's tunable operating points enable tiered clinical protocols that binary alerts cannot provide.

Practical implications: Healthcare organizations can implement advanced event analytics using SQL and managed infrastructure without operating complex streaming frameworks. Fuzzy logic offers a mathematical basis for clinical uncertainty, using linguistic variables that align with clinician reasoning. Well-calibrated outputs support proportional responses to risk.

Future work will validate the system against clinical outcomes (mortality, vasopressor use, mechanical ventilation) rather than heuristic labels, incorporate temporal state modeling (trend detection), and evaluate external generalizability across different care settings.

Declarations

Ethics approval and consent to participate: This study used the publicly available, de-identified MIMIC-IV v3.1 dataset. Secondary analysis of de-identified data does not require institutional review board approval per PhysioNet data use agreements and applicable regulations. Consent to participate is not applicable because the data are retrospective and de-identified.

Consent for publication:

Not applicable (no individual person's data are presented).

Ethics, Consent to Participate, and Consent to Publish declarations:

Not applicable (as detailed above).

Data availability: MIMIC-IV v3.1is publicly available through PhysioNet following completion of required training and data use agreement. SQL transformation scripts and Python analysis code will be made available in a public repository upon publication.

Declaration of Competing Interests

The author declares no competing interests.

Funding: This research received computational support through Google Cloud Platform research credits. No other external funding was obtained.

References

[1]. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, L. A. Celi, R. G. Mark, Mimic-iv, a freely accessible electronic health record dataset, Scientific Data 10 (1) (2023) 1–11. doi:10.1038/s41597-022-01899-x.

[2]. A. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, L. A. Celi, R. G. Mark, Mimic-iv, a freely ac-cessible electronic health record dataset (3.1 v), Scientific Data (2024). doi:10.13026/kpb9-mt58.

[3]. A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stan-ley, Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220. doi:10.1161/01.CIR.101.23.e215.

[4]. M. Singer, C. S. Deutschman, C. W. Seymour, et al., The third interna-tional consensus definitions for sepsis and septic shock (sepsis-3), JAMA 315 (8) (2016) 801–810. doi:10.1001/jama.2016.0287.

[5]. P. Carbone, A. Katsifodimos, S. Ewen, V. Markl, S. Haridi, K. Tzoumas, Apache flink: Stream and batch processing in a single engine, IEEE Data Engineering Bulletin 38 (4) (2015) 28–38.

[6]. M. Armbrust, et al., Structured streaming: A declarative api for real-time applications in apache spark, in: Proceedings of the 2018 Interna-tional Conference on Management of Data (SIGMOD), 2018, pp. 601–613. doi:10.1145/3183713.3190664.

[7]. S. Melnik, et al., Dremel: Interactive analysis of web-scale datasets, in: Proceedings of the 36th International Conference on Very Large Data Bases (VLDB), Vol. 3, 2010, pp. 330–339.

[8]. C. Koch, Y. Ahmad, et al., Incrementalizing analytical data processing with higher-order delta processing, The VLDB Journal 23 (2) (2014) 253–278. doi:10.1007/s00778-013-0334-3.

[9]. S. Agarwal, A. Panda, B. Mozafari, et al., Blinkdb: Queries with bounded errors and bounded response times on very large data, in: Pro-ceedings of the 8th ACM European Conference on Computer Systems (EuroSys), 2013, pp. 29–42. doi:10.1145/2465351.2465355.

[10]. L. A. Zadeh, Fuzzy sets, Information and Control 8 (3) (1965) 338–353. doi:10.1016/S0019-

9958(65)90241-X.

[11]. J. Warner, J. Sexauer, W. V. den Broeck, B. P. Kinoshita, J. Balin-ski, scikit fuzzy, C. Clauss, twmeggs, alexsavio, A. Unnikrishnan, M. Miretti, G. Castelão, F. A. Pontes, T. Uelwer, phme283, pd2f, lau-razh, F. Batista, moetayuko, alexbuy, W. Song, T. Ni, T. G. Bad-ger, R. A. M. Pérez, N. Muoh, J. F. Power, H. Mishra, G. O. Trul-lols, A. Hörteborn, T. Germer, Jdwarner/scikit-fuzzy: Scikit-fuzzy 0.5.0 (Aug. 2024). doi:10.5281/zenodo.13372212. URL https://doi.org/10.5281/zenodo.13372212