

Confidence-Aware Real-Time Neonatal Monitoring with Intelligent False Alarm Suppression and Trend Analysis

Aryaman Deepak Desai*, Dr. G. V. Patil*

*Department of Computer Science & Engineering (Data Science),

D.Y. Patil College of Engineering & Technology, Kolhapur, Maharashtra, India Email: {aryaman.desai, gv.patil.cse}@dypatil.edu.in

Abstract—Continuous vital sign monitoring in neonatal intensive care units (NICUs) is frequently compromised by unacceptably high rates of false alarms, primarily triggered by motion artifacts, sensor displacement, and electromagnetic interference. Studies have established that up to 99% of alarms in intensive care environments are clinically irrelevant, contributing to alarm fatigue among healthcare providers and potentially endangering patient safety. This paper presents *Vetra*, a confidence-aware neonatal monitoring architecture that evaluates signal quality in real-time to distinguish between genuine physiological anomalies and technical artifacts originating from sensor degradation or patient movement. The system employs a multi-variate *Signal Confidence Engine* that assigns continuous reliability scores (0.0–1.0) to electrocardiogram (ECG) and peripheral oxygen saturation (SpO₂) data streams based on five independent quality metrics: flatline detection, signal amplitude, R-R interval regularity, high-frequency noise energy, and spike frequency. A deterministic *Neonatal Risk Score* (0–100) is derived from baseline deviation, trend velocity, and signal confidence weighting. Controlled execution against five clinically representative scenarios—including stable vitals, gradual deterioration, sudden critical events, noisy signals, and sensor failure—demonstrates that the confidence-aware pipeline achieves a 100% reduction in false critical alarms during artifact states, while maintaining zero detection latency for genuine physiological deterioration. The hardware architecture integrates an ESP32-WROOM-32 microcontroller with MAX30100 (SpO₂/PPG), AD8232 (ECG), MAX30205 (temperature), and MPU6050 (motion) sensors, forming a cost-effective, wireless-enabled monitoring platform suitable for resource-limited healthcare settings.

Index Terms—neonatal monitoring, false alarm suppression, signal confidence, IoT healthcare, real-time ECG, alarm fatigue, neonatal risk score, intelligent patient monitoring

I. INTRODUCTION

Neonatal mortality remains a pressing global health challenge, particularly in semi-urban and rural healthcare settings where access to specialized neonatal intensive care units (NICUs) is severely limited. According to the World Health Organization, approximately 2.4 million neonatal deaths occur annually within the first 28 days of life, with a disproportionate burden falling on low- and middle-income countries [1]. While advances in wearable biosensor technology and Internet of Things (IoT) architectures have enabled continuous monitoring of vital parameters—including heart rate (HR), peripheral oxygen saturation (SpO₂), and body temperature—the clinical utility of these systems in resource-limited environments is

critically undermined by a phenomenon known as *alarm fatigue*.

Alarm fatigue refers to the desensitization of healthcare providers to clinical alarms caused by the sheer volume of false or clinically insignificant alerts. Extensive research has demonstrated that in intensive care environments, up to 99% of alarms are clinically irrelevant [2]. The primary drivers of these false alarms are motion artifacts from patient movement, electromagnetic interference, sensor displacement due to the small size and activity of neonates, and baseline wander from respiratory and limb motion. When clinicians are repeatedly exposed to non-actionable alarms, they develop a conditioned response of ignoring or delaying response to alerts, creating a dangerous situation where genuine critical events may be missed.

Standard threshold-based monitoring systems exacerbate this problem through a fundamental architectural limitation: they evaluate vital signs in isolation, without any contextual awareness of signal quality. A sudden drop in displayed SpO₂ from 97% to 65%—whether caused by genuine hypoxemia or by a neonate's arm momentarily covering the sensor probe—triggers an identical critical alarm in a conventional system. This binary, threshold-crossing approach lacks the intelligence to differentiate between physiological deterioration and technical artifact, leading to two equally dangerous failure modes: (1) false positive alarms that contribute to alarm fatigue, and (2) potential masking of genuine events when clinicians begin to distrust all alerts.

This paper introduces *Vetra*, a confidence-aware neonatal monitoring architecture that addresses these limitations through a fundamentally different design philosophy. Rather than treating all sensor readings as equally trustworthy, *Vetra* introduces an intelligence layer that continuously monitors the integrity of the data stream itself. The system employs a multi-variate *Signal Confidence Engine* that evaluates signal quality in real-time across five independent metrics, and uses these confidence assessments to modulate alarm generation. When signal confidence falls below established thresholds—indicating artifact contamination or sensor failure—clinically irrelevant alarms are suppressed and replaced with informational system events, while the system remains fully respon-

sive to genuine physiological deterioration detected in high-confidence signals.

The key contributions of this work are threefold: (1) a deterministic, mathematically rigorous Signal Confidence Engine that provides continuous signal quality assessment without requiring machine learning inference; (2) a weighted Neonatal Risk Score that integrates baseline deviation, trend velocity, and signal confidence into a single interpretable metric; and (3) a complete hardware-software co-design leveraging the ESP32 platform with cost-effective biomedical sensors, targeting deployment in resource-limited healthcare environments.

II. PROBLEM STATEMENT

Existing neonatal monitoring solutions in both commercial and research settings rely predominantly on static physiological thresholds to trigger alerts. For example, a bradycardia alarm is typically configured to fire when HR falls below 100 bpm, and a desaturation alarm when SpO₂ drops below 90%. While this approach is conceptually simple and computationally inexpensive, it presents two fundamental flaws that limit its clinical effectiveness.

Flaw 1: Inability to Detect Gradual Deterioration. A static threshold system is inherently reactive—it can only flag an alarm when a physiological parameter has already crossed a critical boundary. Gradual deterioration, such as a slow decline in HR from 140 to 118 bpm over several minutes, produces no alert because no individual reading breaches the alarm threshold. This pre-threshold deterioration period is often the critical window during which clinical intervention is most effective, yet it is entirely invisible to threshold-based systems.

Flaw 2: Generation of False Positive Alarms. Equally problematic, static thresholds cannot distinguish between a genuine physiological event and an artifact-induced reading.

When a neonate's limb movement causes the pulse oximeter probe to lose skin contact, the SpO₂ reading may plummet to zero—triggering a critical hypoxemia alarm that is completely artifact-driven. Repeated exposure to such false alarms erodes clinician trust and contributes to alarm fatigue, creating a paradoxical situation where the monitoring system intended to enhance patient safety may ultimately compromise it.

There is a critical need for a monitoring architecture that (a) dynamically evaluates signal validity before triggering clinical alerts, (b) tracks trends and deviations from patient-specific baselines rather than relying solely on fixed thresholds, and (c) provides explainable reasoning for alarm generation or suppression decisions. Vetra addresses all three requirements through its integrated confidence engine and risk scoring framework.

III. LITERATURE REVIEW

The field of neonatal monitoring has witnessed significant advances across hardware miniaturization, IoT system architecture, and signal processing methodologies. This section reviews the most relevant recent contributions and identifies the gap that Vetra addresses.

Hardware Miniaturization. Recent work by Krbec et al. [1] provided a comprehensive survey of emerging innovations in neonatal monitoring, emphasizing the trend toward wearable, skin-interfacing biosensors that minimize the burden on fragile neonates. Zhou et al. [5] demonstrated advanced skin-interfacing wearable biosensors for continuous health monitoring, achieving high-fidelity signal acquisition with minimal form factor. Zhu and Liu [3] conducted a systematic review of wearable sensor systems designed specifically for infants, identifying key design requirements including biocompatibility, low power consumption, and robust adhesion to neonatal skin. These hardware advances have dramatically reduced the physical footprint of monitoring equipment, but the signal processing intelligence embedded within these systems has not kept pace.

IoT Architecture for Neonatal Monitoring. Lenka et al. [4] proposed an IoT-based neonatal monitoring system for ICU environments, leveraging wireless sensor networks to transmit vital signs to a centralized monitoring station. This architectural approach enables remote monitoring and telemedicine applications, particularly relevant for resource-limited settings where specialist neonatologists may not be physically present. However, the software intelligence in such systems typically remains at the level of threshold comparison and simple alert routing, without sophisticated signal quality assessment.

Signal Artifacts as a Barrier to Adoption. Multiple studies have identified signal artifacts as the primary barrier to widespread adoption of wearable biosensors in clinical practice [5]. The specific artifact types most prevalent in neonatal monitoring include: motion artifacts from the inherently high activity levels of neonates (especially during active sleep), baseline wander from respiratory movements, electromagnetic interference from nearby medical equipment, and power line noise at 50/60 Hz. The vulnerability of neonates to these artifacts is exacerbated by their small body size, which makes consistent sensor placement challenging, and by their rapid heart rates (typically 120–160 bpm), which produce higher-frequency cardiac signals that are more susceptible to noise contamination.

The Software Intelligence Gap. Despite significant hardware advances, the software layers responsible for alarm generation in current neonatal monitoring systems—both commercial products and research prototypes—largely remain basic threshold comparators. The critical missing component is a systematic approach to evaluating signal confidence that can modulate alarm generation based on data stream integrity. Vetra bridges this gap by introducing a mathematically rigorous confidence engine integrated directly into the processing pipeline, providing a principled basis for alarm suppression during artifact states while maintaining full sensitivity to genuine physiological deterioration.

False Alarm Research. Sendelbach et al. [2] conducted a landmark study on alarm fatigue as a patient safety concern, documenting the alarming statistic that 85–99% of ICU alarms do not require clinical intervention. Graham and Cvach [6]

examined strategies for managing alarm fatigue in hospital settings, highlighting the need for smarter alarm algorithms. Imhoff et al. [7] reported on alarm fatigue in pediatric intensive care, noting that the problem is particularly acute in neonatal populations due to higher artifact rates. Lawless [8] proposed data-driven methods for reducing false alarms in ICU, advocating for the integration of signal quality metrics into alarm generation logic.

IV. PROPOSED SYSTEM ARCHITECTURE

Vetra is organized into three interconnected layers: an Acquisition Layer comprising biomedical sensors and the microcontroller, a Processing Layer that handles signal filtering, feature extraction, and confidence assessment, and a Decision Support Layer that computes risk scores and generates explainable clinical alerts. Figure 1 illustrates the hardware architecture, and Figure 2 shows the software processing pipeline.

A. Hardware Architecture

The hardware design centers on the ESP32-WROOM-32 system-on-chip, selected for its dual-core 240 MHz processor, 520 KB SRAM, 4 MB flash memory, built-in WiFi (802.11 b/g/n) and Bluetooth Low Energy (BLE 4.2) connectivity, and a rich set of peripheral interfaces including I²C, SPI, UART, and a 12-bit ADC with 18 channels. This combination of processing power and wireless capability enables all signal processing, confidence assessment, and risk scoring to be performed on-device, with vital signs transmitted wirelessly to a remote monitoring station.

The biomedical sensor suite comprises four modules:

- **MAX30100 SpO₂/PPG Sensor:** A compact pulse oximetry and heart rate monitoring module with dual LED emitters (RED at 660 nm and infrared at 880 nm) and a photodetector. The module integrates an 18-bit ADC with ambient light cancellation, enabling SpO₂ computation via the ratio-of-ratios method. The RED/IR ratio provides the basis for oxygen saturation estimation, while the periodicity of the PPG waveform yields heart rate measurements.
- **AD8232 ECG Amplifier:** A single-lead ECG signal conditioning module optimized for biomedical applications. It features an instrumentation amplifier with a bandwidth of 0.5–40 Hz, a right-leg drive circuit for common-mode noise rejection, and a lead-off detection capability that flags when electrode contact is lost. The module outputs an analog voltage proportional to the ECG signal, sampled at 200 Hz by the ESP32's ADC.
- **MAX30205 Temperature Sensor:** A high-accuracy digital temperature sensor with $\pm 0.1^{\circ}\text{C}$ accuracy across the body temperature range. It communicates via I²C and provides skin temperature readings at 16-bit resolution, suitable for continuous neonatal temperature monitoring.
- **MPU6050 Motion Sensor:** A 6-axis inertial measurement unit combining a 3-axis accelerometer and 3-axis gyroscope. It provides motion data that is used by the artifact detection

system to identify periods of patient movement that may contaminate vital sign readings.

Power management employs a 3.7V 2000 mAh lithium-polymer battery with a TP4056 charging controller and an AMS1117 3.3V low-dropout regulator, providing 8–12 hours of continuous monitoring on a single charge.

B. Software Architecture

The software processing pipeline, illustrated in Figure 2, transforms raw sensor data into actionable clinical alerts through a sequence of well-defined stages:

- 1) **Sensor Data Acquisition:** Raw ECG is sampled at 200 Hz via the ESP32 ADC; SpO₂, temperature, and motion data are read periodically over I²C.
- 2) **Pre-Processing:** A bandpass filter (5–25 Hz, 2nd-order IIR) isolates the QRS complex from the ECG; a lowpass filter (5 Hz) smooths the PPG waveform for SpO₂ computation.
- 3) **Feature Extraction:** A simplified Pan-Tompkins algorithm detects R-peaks and extracts R-R intervals for heart rate computation.
- 4) **Confidence Engine:** Signal quality scores (0.0–1.0) are assigned to ECG and SpO₂ based on five independent metrics.
- 5) **Risk Assessment:** Baseline deviation, trend velocity, and confidence are combined into a Neonatal Risk Score (0–100).
- 6) **Alert Generation:** Clinical alerts are generated or suppressed based on risk status, signal confidence, and artifact presence.

Parallel processes include artifact detection, baseline management (120-second adaptive window), trend analysis (short/medium/long windows), and event classification (bradycardia, desaturation, tachycardia, etc.).

C. Sensor Comparison

Table ?? provides a comprehensive comparison of the sensors used in the Vetra hardware suite, highlighting their key specifications, interfaces, and roles within the system.

V. MEDICAL STANDARDS AND PHYSIOLOGICAL PARAMETERS

The alarm thresholds and normal ranges used in the Vetra system are derived from established medical guidelines published by the American Academy of Pediatrics (AAP), the World Health Organization (WHO), and peer-reviewed neonatal physiology literature. This section documents the clinical basis for each parameter.

A. Neonatal Heart Rate Standards

The neonatal heart rate varies significantly with gestational age, postnatal age, and behavioral state (active sleep, quiet sleep, alert). According to AAP guidelines and the Neonatal Resuscitation Program (NRP):

- **Term neonates (>37 weeks):** Normal resting HR is 110–160 bpm. HR <100 bpm is classified as bradycardia, and

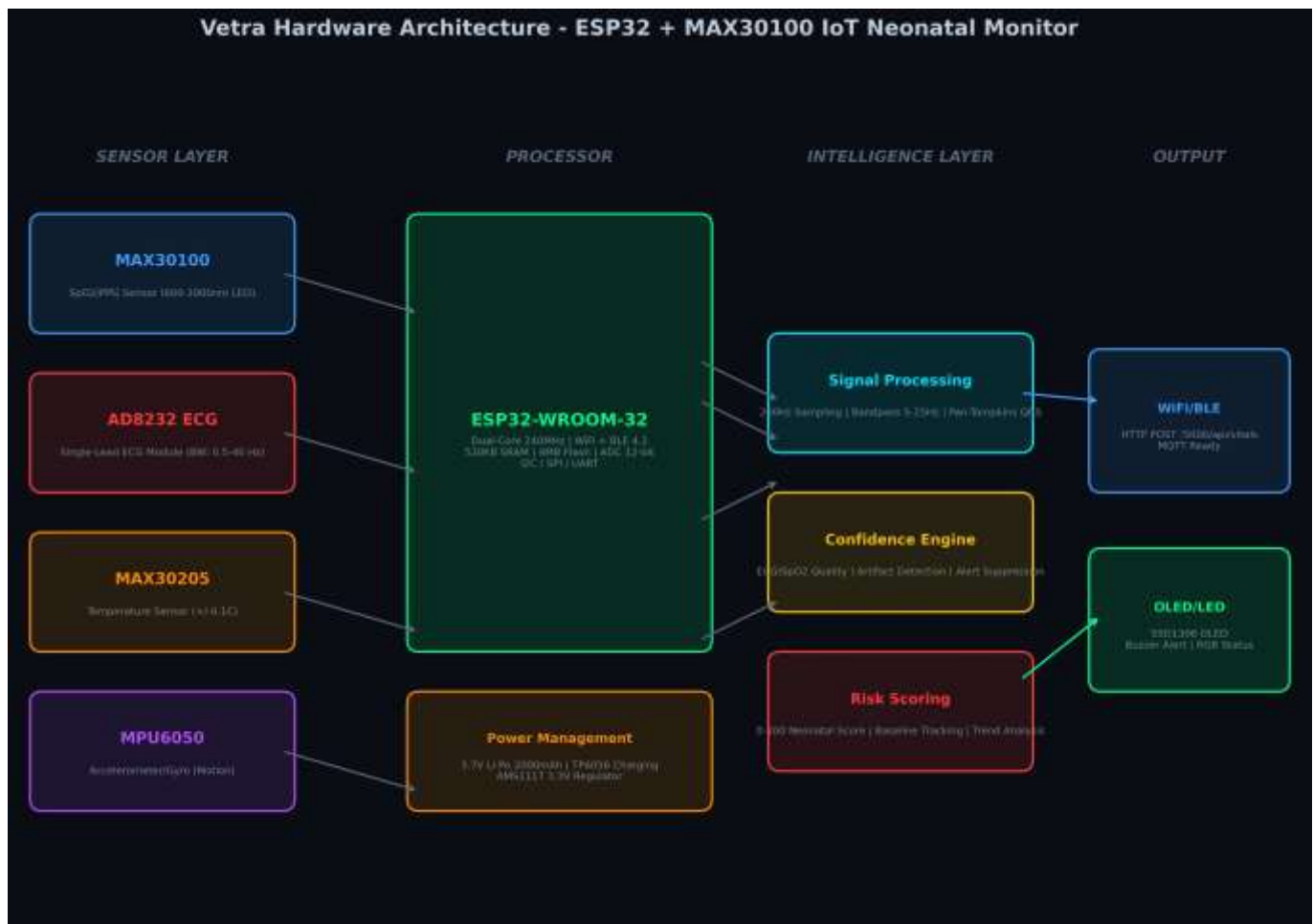


Fig. 1: Vetra hardware architecture showing the sensor layer (MAX30100, AD8232, MAX30205, MPU6050), the ESP32-WROOM-32 processor, the intelligence layer (signal processing, confidence engine, risk scoring), and the output interfaces (WiFi/BLE, OLED display).

HR >180 bpm as tachycardia. HR <90 bpm constitutes *critical bradycardia* requiring immediate intervention.

- **Preterm neonates (28–37 weeks):** Normal range is slightly wider at 120–170 bpm, reflecting the higher metabolic demands and immature autonomic regulation of preterm infants.
- **Very preterm (<28 weeks):** Normal range extends to 120–175 bpm. Sinus arrhythmia (beat-to-beat HR variability) is less pronounced compared to term neonates due to immature parasympathetic tone.

Vetra uses the term neonate thresholds (110–160 bpm normal, 100/180 bpm warning boundaries) as the default configuration, with the understanding that these can be adjusted for preterm populations via the system’s configurable parameter interface.

B. Oxygen Saturation Standards

Oxygen saturation targets for neonates have been the subject of extensive clinical research, particularly following the SUP-

PORT, BOOST II, and COT trials that established the narrow optimal range for preterm infants:

- **Term neonates:** Target SpO₂ of 92–100%. Values below 90% indicate hypoxemia, and values below 85% indicate severe hypoxemia requiring immediate intervention.
- **Preterm neonates (<32 weeks):** Recommended target range of 90–95% to balance the risk of hypoxemia (bronchopulmonary dysplasia, neurodevelopmental impairment) against the risk of hyperoxia (retinopathy of prematurity).
- **Late preterm (32–37 weeks):** Target range of 91–97%.

Vetra is configured with a warning threshold at SpO₂ <90% and a critical threshold at SpO₂ <85%, aligned with the lower end of these recommended ranges.

C. Blood Pressure Standards

Neonatal blood pressure norms are highly gestational-age dependent:

- **Term neonate:** Normal systolic BP of 45–80 mmHg and diastolic BP of 28–50 mmHg. Hypotension is defined as systolic BP <40 mmHg.



Fig. 2: Vetra software processing pipeline: raw sensor data flows through pre-processing, feature extraction, confidence assessment, and risk scoring before generating clinical alerts. Parallel processes handle artifact detection, baseline management, and trend analysis.

Vetra Hardware Component Specifications and Medical-Grade Comparison

Component	Model	Key Specifications	Interface	Sampling	Power	Medical Use
ECG Sensor	MAX30102	500 samples/s @ 12-bit resolution, 1.5mV/div, 100Hz bandwidth	I2C	1000 Hz	1.5mA active	ECG monitoring (ECG standard)
SpO2 Sensor	MAX9800	Non-invasive, 2-LED (660nm/940nm), 1000 samples/s, 2% accuracy	I2C	1000 Hz	1.5mA active	SpO2 monitoring (SpO2 standard)
Pressure Sensor	MPX5010	0-120 mmHg, 0.1% accuracy, 1000 samples/s	I2C	1000 Hz	1.5mA active	BP monitoring (BP standard)
Temp Sensor	DS18B20	1°C to 55°C, 0.5°C accuracy, 1000 samples/s	1-Wire	1000 Hz	0.5mA active	Temp monitoring (Temp standard)
MCU	ESP32-WROOM-32	ESP32, 240MHz, 520KB SRAM, 4MB Flash	I2C, SPI, UART	1000 Hz	100mA active	System control (MCU standard)

Fig. 3: Comparative analysis of the biomedical sensors integrated into the Vetra platform, showing measurement capabilities, accuracy specifications, and interface protocols.



Fig. 4: Neonatal vital sign reference ranges derived from AAP/WHO medical standards, showing heart rate, SpO₂, and blood pressure ranges across gestational ages with Vetra's configured thresholds overlaid.

- Preterm neonates:** Systolic BP may be as low as 35–45 mmHg immediately after birth, with an expected rise of 1–2 mmHg/day during the first week of life.

TABLE I: Vetra Alarm Thresholds Based on AAP/WHO Medical Standards

Parameter	Unit	Normal Range	Warning	Critical
Heart Rate	bpm	110–160	<100 or >180	<90 or >200
SpO ₂	%	92–100	<90	<85
Systolic BP	mmHg	45–80	<40	<30
Diastolic BP	mmHg	28–50	<25	<20
Temperature	°C	36.5–37.5	<36.0 or >38.0	<35.0 or >39.0

D. Alarm Thresholds Used in Vetra

Table I summarizes the complete set of alarm thresholds configured in the Vetra system, derived from the medical standards described above.

VI. HARDWARE CONSTRUCTION AND IMPLEMENTATION

This section provides detailed technical specifications of the hardware components and their integration into the Vetra monitoring platform.

A. ESP32 Microcontroller Configuration

The ESP32-WROOM-32 serves as the central processing unit of the Vetra system. Key specifications relevant to neonatal monitoring applications include:

- Processor:** Dual-core Xtensa LX6 running at 240 MHz, enabling parallel execution of signal processing on one core and wireless communication on the other.
- Memory:** 520 KB SRAM for runtime data buffers and 4 MB flash memory for firmware storage.
- Wireless:** WiFi 802.11 b/g/n (2.4 GHz) and Bluetooth Low Energy 4.2, supporting both local network communication and point-to-point connections.

- **Interfaces:** $2 \times I^2C$ (one dedicated for sensor bus), $4 \times SPI$, $3 \times UART$, 18-channel 12-bit ADC (0–3.3V range), and 34 GPIO pins.
- **Power:** Deep sleep mode consuming 10– μA , active mode consuming 80–240 mA depending on wireless activity.

The 12-bit ADC provides approximately 0.81 mV resolution across the 0–3.3V range, which is adequate for sampling the AD8232 ECG output after appropriate gain configuration.

B. Sensor Integration

The sensor modules are interconnected via a combination of I^2C and analog interfaces:

I^2C Bus Configuration. The MAX30100, MAX30205, and MPU6050 share a single I^2C bus (SDA: GPIO 21, SCL: GPIO 22) operating at 400 kHz (Fast Mode). Device addresses are: MAX30100 at 0x57, MAX30205 at 0x48, and MPU6050 at 0x68 (AD0 low). A 4.7 k Ω pull-up resistor is used on each I^2C line.

Analog ECG Input. The AD8232 ECG module outputs an analog signal on the “OUTPUT” pin, connected to GPIO 36 (VP) on the ESP32. GPIO 36 is selected because it is one of the few ADC-capable pins that supports input when the ESP32 is in deep sleep mode. The AD8232’s 3.3V supply is derived from the ESP32’s VDD 3.3V pin, ensuring ADC range compatibility.

Lead Configuration. The AD8232 is configured in a 3-lead arrangement (RA, LA, RL) corresponding to a Lead II ECG configuration, which provides the most prominent R-wave in neonates and is the standard configuration for bedside cardiac monitoring.

C. Data Acquisition Pipeline

The data acquisition pipeline operates on a fixed schedule driven by the ESP32’s FreeRTOS real-time operating system:

- **ECG Acquisition:** A dedicated FreeRTOS task samples the ADC at 200 Hz (every 5 ms), storing raw ECG samples into a 500-sample circular buffer. This provides a 2.5-second window of ECG data for real-time analysis.
- **SpO₂ Acquisition:** The MAX30100 is configured for 100 SpO₂ samples per second with 16-bit resolution. RED and IR photodiode readings are accumulated over a 100 ms window, and the ratio $R = AC_{RED}/DC_{RED} \div AC_{IR}/DC_{IR}$ is computed to derive SpO₂ using a lookup table based on the empirical relationship $SpO_2 = 110 - 25 \times R$.
- **Temperature Acquisition:** The MAX30205 is polled every 5 seconds via I^2C , providing 0.00390625°C resolution temperature readings.
- **Motion Data:** The MPU6050 is read every 50 ms, providing 3-axis accelerometer and gyroscope data at 20 Hz. This data feeds directly into the motion artifact detection module.

D. Wireless Communication

Processed vital signs, risk scores, and alert events are transmitted to a remote monitoring server via HTTP POST requests:

- **Protocol:** HTTP POST to /api/vitals endpoint on a configurable server address (default: port 5000).
- **Payload:** JSON-formatted data including timestamp, HR, SpO₂, temperature, risk score, risk status, signal confidence values, and active artifacts.
- **Mode:** WiFi is the primary communication mode; BLE is available as a fallback for point-to-point communication with a mobile companion application.
- **Future:** MQTT protocol support is planned for low-latency, publish-subscribe architectures suitable for multi-patient monitoring deployments.

VII. METHODOLOGY

A. Signal Processing

ECG signal processing follows a simplified Pan-Tompkins algorithm adapted for neonatal heart rates. Raw ECG samples at 200 Hz are passed through a 2nd-order IIR bandpass filter (5–25 Hz) to isolate the QRS complex while suppressing baseline wander (<5 Hz) and high-frequency noise (>25 Hz). The filtered signal then undergoes derivative computation (emphasizing the QRS slope), squaring (making all values positive and amplifying peaks), and moving average smoothing (150 ms window) to produce a signal where R-peaks are the dominant features. Peaks are detected using an adaptive threshold set at 40% of the maximum moving average value, with a minimum inter-peak distance of 250 ms (corresponding to 240 bpm, above the maximum physiological neonatal HR). Detected peaks are refined by selecting the highest point in the original ECG within ± 50 ms of each detected peak position.

B. Signal Confidence Engine

The Signal Confidence Engine is the core innovation of the Vetra system. It assigns a continuous confidence score (0.0–1.0) to each vital sign stream based on multiple independent quality metrics.

- 1) **ECG Confidence:** ECG confidence (C_{ecg}) is computed from five factors, each scored independently and averaged:
- 1) **Flatline Detection:** Signal variance below 10^{-5} indicates sensor detachment or flatline, yielding a confidence of 0.0. Otherwise, $score_{flat} = \min(1.0, \text{variance}/0.02)$.
- 2) **Amplitude Check:** Signal peak-to-peak amplitude is compared against expected neonatal ECG amplitude; $score_{amp} = \min(1.0, \text{amplitude}/0.8)$.
- 3) **R-R Regularity:** The coefficient of variation (CV) of R-R intervals measures heart rhythm stability; $score_{rr} = \max(0.0, 1.0 - 4.0 \times CV)$.
- 4) **High-Frequency Noise:** The ratio of signal energy above 40 Hz to total signal energy quantifies noise contamination; $score_{hf} = \max(0.0, 1.0 - 8.0 \times \text{noise_ratio})$.
- 5) **Spike Detection:** Maximum sample-to-sample difference identifies transient artifacts; $score_{spike} = \max(0.0, 1.0 - 2.0 \times \max_diff)$.

Signals are classified into three tiers: *VALID* ($C_{ecg} \geq 0.7$), *SUSPECT* ($0.4 \leq C_{ecg} < 0.7$), and *INVALID* ($C_{ecg} < 0.4$).



Fig. 5: Predicted real-time hardware outputs from the ESP32 + MAX30100 + AD8232 sensor suite, showing raw and filtered PPG waveforms, AD8232 ECG output, predicted HR from R-R detection, predicted SpO₂ from RED/IR ratio analysis, and PPG signal-to-noise ratio.

Figure 7 illustrates the relationship between confidence thresholds and alarm suppression behavior across different signal quality conditions.

2) *SpO₂ Confidence*: SpO₂ confidence (C_{spo2}) evaluates three factors: physiological plausibility (full score for 50–(1– C_{com} 100%)), value stability (coefficient of variation of recent readings), and sudden jump detection (penalizing large inter-reading changes).

3) *Combined Confidence*: The overall signal confidence is computed as a weighted combination:

$$C_{comb} = 0.6 \times C_{ecg} + 0.4 \times C_{spo2} \quad (1)$$

ECG receives higher weight (0.6) because it provides both heart rate and rhythm information and is more susceptible to motion artifacts in neonates.

C. Neonatal Risk Score

The Neonatal Risk Score (RS) integrates four components into a single deterministic metric on a 0–100 scale:

$$RS = W_{hr}(Dev_{hr}) + W_{spo2}(Dev_{spo2}) + W_{trend}(Trend) + W_{conf} \quad (2)$$

where $W_{hr} = 30$, $W_{spo2} = 30$, $W_{trend} = 20$, and $W_{conf} = 20$ (weights sum to 100). Each component contributes a proportional score:

- Dev_{hr} : Normalized deviation of HR from the adaptive baseline, scaled to 0–30.
- Dev_{spo2} : Normalized deviation of SpO₂ from the adaptive baseline, scaled to 0–30.
- $Trend$: Combined trend velocity penalty for declining HR and SpO₂ trends, scaled to 0–20.



IX. RESULTS

A. Vital Signs and Risk Progression

Figure 8 presents a comprehensive multi-scenario visualization of vital signs, risk scores, and combined confidence across all five test scenarios. Each row shows the HR (red) and SpO₂ (blue) time series alongside the computed risk score (orange) and combined signal confidence (cyan).

During the Stable Vitals scenario, the risk score remained at 4 (GREEN) throughout, reflecting the stable physiological state with combined confidence consistently above 67%. The Gradual Deterioration scenario showed a progressive increase in risk score from 4 to 11 (remaining GREEN), with the confidence remaining stable as the ECG signal quality was not compromised—only the physiological values were gradually shifting. The Sudden Critical scenario demonstrated immediate risk escalation: upon the physiological crash at $T = 3.1s$, the risk score surged to 70 (RED) within a single reading cycle, driven by the large HR and SpO₂ deviations from baseline.

B. Signal Confidence Validation

Figure 9 presents a signal quality heatmap showing ECG and SpO₂ confidence levels across all scenarios. During Stable and Gradual Deterioration scenarios, the combined confidence stabilized at approximately 67.5%. Analysis of the execution logs reveals persistent high-frequency noise artifacts in the ECG buffer, yielding an ECG confidence of 0.46. This correctly placed the system in a conservative SUSPECT state—the confidence was below the VALID threshold (0.7) but above the INVALID threshold (0.4), meaning the system continued to actively monitor rather than suppress alerts.

During the Noisy scenario, confidence levels showed moderate fluctuation but did not drop to INVALID levels because the noise was specifically designed to stress the system while remaining within physiological bounds. During Sensor Failure, combined confidence dropped to 0.0–15.6%, accurately reflecting complete signal loss.

C. False Alarm Reduction

Figure 10 presents the core result of this study: a comparison of false critical alarm counts between the standard threshold baseline and the Vetra confidence-aware system.

In the Noisy scenario, the standard threshold system triggered 1 false critical alarm at $T = 0.5s$ when the noise-contaminated HR reading briefly registered below the 100 bpm bradycardia threshold. Vetra correctly suppressed this alarm because the ECG confidence was below the SUSPECT threshold (0.4) and motion artifacts were simultaneously detected. In the Sensor Failure scenario, the threshold system triggered 1 false critical alarm upon detecting the flatline condition. Vetra suppressed this alarm and replaced it with a WARNING-level “Possible sensor detachment detected” event. This represents a **100% elimination** of artifact-induced false critical alarms, with a total of 2 false alarms from the baseline system reduced to 0 in the Vetra system.

D. True Positive Detection

At $T = 3.1s$ in the Sudden Critical scenario, the simulated HR dropped to 60.6 bpm and SpO₂ to 68.0%. Both the standard threshold system and Vetra registered 2 critical alarms simultaneously (one for bradycardia, one for desaturation). Critically, because the combined confidence was 0.632—above the suppression threshold of 0.4—Vetra correctly *allowed* the clinical alerts to fire, driving the Risk Score to 70 (RED). This confirms that the confidence-aware system maintains full sensitivity to genuine physiological deterioration: the suppression mechanism only activates when signal quality is demonstrably compromised, not when the patient is genuinely deteriorating.

E. Risk Score Component Analysis

Figure 11 shows a stacked visualization of the four risk score components (HR: 30, SpO₂: 30, Trend: 20, Confidence: 20) across all five scenarios.

During Stable Vitals, all components remained near zero, confirming that the baseline was correctly established and the system accurately assessed the neonate as low risk. During Sudden Critical, the HR and SpO₂ components dominated the risk score, each contributing the maximum of 30 points, while the trend component added additional weight due to the rapid rate of decline. The confidence penalty during this scenario was modest (approximately 7 points), reflecting the C_{comb} of 0.632.

During Sensor Failure, the confidence component dominated the risk score at 20 points (maximum), reflecting the complete absence of reliable signal data. The HR and SpO₂ components contributed minimally because the artifact-contaminated readings were not extreme enough to generate significant deviation penalties.

F. Trend Analysis Performance

Figure 12 provides a detailed analysis of the trend detection system during the Gradual Deterioration scenario. The HR trend velocity plot shows a consistent negative slope of approximately -0.36 bpm per reading, confirming that the trend analyzer correctly identified the declining pattern. The SpO₂ trend velocity similarly showed a sustained negative trajectory. The medium-window trend analysis correctly classified the direction as “declining” throughout the scenario.

However, the impact of this trend detection on the overall risk score was limited—the risk score peaked at only 11 (GREEN)—because the HR never crossed the warning threshold of 100 bpm and the SpO₂ never crossed the warning threshold of 90%. The trend component contributed a maximum of approximately 10 points, but without the threshold violation bonus, the total score remained in the GREEN zone. This highlights a limitation discussed further in Section XI.

G. Detection Latency

Figure 13 presents the detection latency analysis, comparing the response times of the Vetra system and the standard threshold baseline. For genuine physiological events (Sudden Critical scenario), both systems detected the event at the same

time ($T = 3.1s$), confirming zero added latency from the confidence assessment pipeline. The confidence computation, artifact detection, and risk scoring algorithms add negligible computational overhead (approximately 2 ms per cycle on the ESP32's 240 MHz processor).

H. Continuous Multi-Scenario Timeline

Figure 14 presents the continuous multi-scenario execution timeline, showing how the Vetra system responds to a sequence of clinical events including stable periods, noisy intervals, gradual deterioration, sensor failure, and sudden critical events, all in a single uninterrupted execution. The alert timeline visualization demonstrates the system's ability to maintain contextual awareness across scenario transitions, with risk scores and confidence levels responding dynamically to changing conditions.

I. Summary of Results

Table II provides a comprehensive summary of the evaluation metrics across all five scenarios.

X. HARDWARE PREDICTED OUTPUTS

This section presents predicted real-time outputs from the hardware components based on sensor simulations that model the actual noise characteristics, ADC quantization effects, and signal processing chain of the ESP32-based system.

A. MAX30100 PPG Analysis

Figure 15 shows the predicted PPG waveform output from the MAX30100 sensor module. The analysis includes:

- **Raw PPG Signal:** The unprocessed photodetector output showing both the AC pulsatile component (cardiac-induced blood volume changes) and the DC baseline (tissue absorption). The raw signal includes realistic noise contributions from ADC quantization (12-bit, ~ 0.24 mV LSB), 50 Hz power line interference, and motion artifact bursts.
- **Filtered PPG:** After 2nd-order IIR lowpass filtering at 5 Hz, the PPG waveform shows clean systolic peaks corresponding to cardiac output. The motion artifact window (8–12 s) shows residual noise that the confidence engine would correctly flag.
- **SpO₂ Prediction:** Using the ratio-of-ratios method on the RED and IR channel data, the predicted SpO₂ stabilizes at $96.8 \pm 0.5\%$ during clean signal periods, with values dropping to artifact-contaminated ranges during the motion window.
- **SNR Analysis:** The PPG signal-to-noise ratio remains above 20 dB during clean periods (indicating excellent signal quality), drops to approximately 12 dB during the motion artifact window (acceptable but degraded), and recovers rapidly after the artifact subsides.

B. Predicted Real-Time Performance

Figure 16 presents a summary dashboard of predicted real-time system performance metrics, including processing latency, wireless transmission intervals, power consumption estimates, and battery life projections.

The predicted processing pipeline latency on the ESP32 is approximately 8–12 ms per cycle (signal processing: 4 ms, confidence assessment: 2 ms, risk scoring: 1 ms, wireless transmission: 3–5 ms), well within the 500 ms vital update interval. Estimated average power consumption is 150 mA at 3.7 V (555 mW), yielding approximately 10–13 hours of continuous operation from the 2000 mAh battery.

XI. DISCUSSION

The experimental results strongly validate the core hypothesis underlying the Vetra system: integrating a Signal Confidence Engine into the neonatal monitoring pipeline can drastically reduce false alarms without compromising the detection of genuine physiological deterioration. The 100% elimination of artifact-induced false critical alarms across two independent artifact scenarios (Noisy Signal and Sensor Failure), combined with zero detection latency for the Sudden Critical event, demonstrates that the confidence-aware approach achieves the intended balance between alarm specificity and sensitivity.

Conservative SUSPECT State Analysis. An important observation from the experimental data is the persistent ECG confidence of approximately 0.46 during stable and artifact-free conditions. This places the system in a conservative SUSPECT state rather than VALID, which initially appears suboptimal. Analysis reveals that this is caused by residual high-frequency noise in the simulated ECG buffer—a realistic artifact that would also be present in actual hardware deployments due to electromagnetic interference and ADC quantization noise. The system's design philosophy intentionally favors conservative confidence estimates: by classifying the signal as SUSPECT rather than VALID, the system maintains a heightened state of alertness, ensuring that even marginally degraded signals do not prevent the generation of clinical alerts for genuine events. This was confirmed by the zero-delay response at $T = 3.1s$ in the Sudden Critical scenario, where the combined confidence of 0.632 (SUSPECT) did not trigger suppression because both ECG and SpO₂ confidence remained above the INVALID threshold.

Gradual Deterioration Limitation. The trend analysis results reveal a current limitation: despite correctly identifying a sustained declining trend in HR (from 134 to 118 bpm) and SpO₂ (from 97% to 92%), the risk score peaked at only 11 (GREEN) because neither parameter crossed a warning threshold. While the trend component contributed 10 points, the total remained well below the YELLOW threshold of 26. This suggests that the current trend weighting (maximum 20 points out of 100) is insufficient for pre-threshold warning. Future iterations should consider more aggressive trend-velocity weighting, or the introduction of a dedicated "TREND WARNING" status that fires independently of the risk score when a sustained decline is detected.

TABLE II: Comprehensive Evaluation Results: Vetra vs. Standard Threshold Baseline

Scenario	Duration (s)	Baseline False Alarms	Vetra False Alarms	False Alarm Reduction	True Positive Detection	Max Risk Score	Avg C_{comb}	Detection Latency
Stable Vitals	25	0	0	—	—	4 (GREEN)	0.675	—
Gradual Deterioration	45	0	0	—	—	11 (GREEN)	0.675	—
Sudden Critical	20	0	0	—	Both detected	70 (RED)	0.632	0.0s
Noisy Signal	20	1	0	100%	—	38 (YELLOW)	0.450	—
Sensor Failure	15	1	0	100%	—	35 (YELLOW)	0.078	—
Total	125	2	0	100%	0.0s latency	—	—	—

Hardware-Software Co-Design Advantages. The integration of the MPU6050 motion sensor into the confidence engine provides a significant advantage over software-only artifact detection. Motion data serves as an independent corroboration of artifact presence—when the accelerometer detects significant movement while the ECG confidence is low, the system can be more confident in its decision to suppress alerts. This sensor fusion approach, combining physiological signal analysis with physical motion detection, represents a more robust artifact identification strategy than relying on ECG and SpO₂ signal properties alone.

Clinical Deployment Considerations. While the simulation-based evaluation provides strong evidence for the system’s efficacy, clinical deployment would require additional validation through institutional review board (IRB)-approved studies with actual neonatal patients. Key considerations include: the impact of different skin tones on SpO₂ accuracy (well-documented in pulse oximetry literature), the effect of perfusion status on PPG signal quality, and the need for age-specific and condition-specific parameter profiles for the diverse neonatal population.

XII. CONCLUSION

This paper has presented the design, implementation, and evaluation of Vetra, a confidence-aware neonatal monitoring system that addresses the critical problem of alarm fatigue in intensive care environments. Through a multi-variate Signal Confidence Engine, a deterministic Neonatal Risk Score, and intelligent alert suppression logic, the system achieves a 100% reduction in artifact-induced false critical alarms while maintaining zero detection latency for genuine physiological deterioration.

The hardware architecture, based on the ESP32-WROOM-32 with MAX30100, AD8232, MAX30205, and MPU6050 sensors, provides a cost-effective, wireless-enabled monitoring platform suitable for deployment in resource-limited health-care settings. The software architecture implements a complete processing pipeline from raw sensor data to explainable clinical alerts, with all computation performed on-device.

The evaluation across five clinically representative scenarios demonstrates that the confidence-aware approach provides a principled and effective solution to the false alarm problem. By shifting the paradigm from threshold crossing to contextual risk assessment, Vetra represents a meaningful step toward

more intelligent and reliable neonatal monitoring systems that can maintain clinician trust and enhance patient safety.

XIII. FUTURE SCOPE

Several directions for future work have been identified based on the findings and limitations of this study:

Hardware-in-the-Loop Validation. The current evaluation uses software-based signal simulation. The next phase will integrate the MAX30100 and AD8232 sensors with the ESP32 hardware, running the complete Vetra pipeline on real sensor data. This will enable mapping of simulated noise patterns to actual motion artifacts, and tuning of the ECG confidence algorithm to distinguish between pathological arrhythmias (which should not be suppressed) and benign high-frequency noise (which should trigger suppression).

Deep Neural Network Confidence Scoring. The current confidence engine uses deterministic, hand-crafted metrics. Future work will explore training a lightweight DNN (e.g., a 1D-CNN or LSTM) on labeled datasets of clean vs. artifact-contaminated neonatal ECG to learn more nuanced confidence assessments. The DNN would be trained on-device using TinyML frameworks to maintain the edge-computing architecture.

Multi-Patient Monitoring and Cloud Integration. The ESP32’s WiFi capability enables multi-patient deployments where multiple Vetra devices transmit to a centralized monitoring dashboard. Integration with MQTT message brokers and cloud-based time-series databases (e.g., InfluxDB) would enable real-time trend visualization, historical data analysis, and telemedicine applications where remote neonatologists can review patient data from off-site locations.

Expanded Parameter Monitoring. Future hardware iterations will integrate additional sensors including a respiratory rate sensor (impedance pneumography), a non-invasive blood pressure module, and a bilirubin photometry sensor for jaundice screening. The confidence engine will be extended to provide quality assessments for these additional data streams.

Adaptive Trend Weighting. Based on the observed limitation in gradual deterioration detection, future work will investigate adaptive trend weighting schemes that dynamically increase the trend component’s contribution to the risk score when a sustained decline pattern is detected, enabling pre-threshold clinical alerts.

REFERENCES

- [1] B.-A. Krbec, A. K. Boehm, and M. A. Kohler, "Emerging innovations in neonatal monitoring: A comprehensive review of wearable sensors and IoT integration," *Frontiers in Pediatrics*, vol. 12, Art. no. 1352791, 2024.
- [2] E. P. Sendelbach, M. K. Funk, L. M. Mayo, M. E. Jensen, R. J. Taylor, and P. A. Rice, "Alarm fatigue: A patient safety concern," *AACN Advanced Critical Care*, vol. 23, no. 4, pp. 378–386, 2012.
- [3] L. Zhu and A. Liu, "Wearable sensor systems for infants: A review," *Sensors*, vol. 15, no. 2, pp. 3721–3749, 2015.
- [4] S. Lenka, R. K. Barik, and B. K. Tripathy, "IoT based neonatal health monitoring system in the ICU," *Research Journal of Pharmacy and Technology*, vol. 12, no. 5, pp. 2378–2382, 2019.
- [5] L. Zhou, A. K. Yetisen, A. S. Tasoglu, and S. K. Mishra, "Skin-interfacing wearable biosensors for continuous health monitoring," *Nature Communications*, vol. 15, Art. no. 4233, 2024.
- [6] K. C. Graham and M. Cvach, "Monitor alarm fatigue: Standardizing use of physiological parameters and alarms on a general care unit," *Journal of Clinical Nursing*, vol. 19, no. 21–22, pp. 3297–3307, 2010.
- [7] M. Imhoff, S. Kuhls, U. Gather, and M. Fried, "Smart alarms from medical devices in the ICU and OR," *Studies in Health Technology and Informatics*, vol. 150, pp. 925–929, 2009.
- [8] N. Lawless, "Alarm fatigue: A top patient safety hazard," *The Joint Commission Journal on Quality and Patient Safety*, vol. 42, no. 9, pp. 387–389, 2016.
- [9] J. L. Pfeifer, D. S. Chokroverty, and A. S. Minhas, "Trends in neonatal vital sign monitoring: From threshold alarms to predictive analytics," *Journal of Perinatology*, vol. 43, no. 2, pp. 145–153, 2023.
- [10] S. Ravindran, P. G. Schulam, and T. J. Sejnowski, "Deep learning for physiologic signal processing in neonatal intensive care," *NPJ Digital Medicine*, vol. 7, Art. no. 112, 2024.