International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

On-Device Deep Learning with Live Channel Attention for Real-Time ANS State Classification on Wearable Platforms

Anandhu P.[1], Ajay B. S.[2] and Arun C.[3]

[1] Department of Artificial Intelligence and Data Science, Nehru Institute of Engineering and Technology, Coimbatore

[2] Department of Artificial Intelligence and Data Science, Nehru Institute of Engineering and Technology, Coimbatore

[3] Department of Artificial Intelligence and Data Science, Nehru Institute of Engineering and Technology, Coimbatore

Email: [1] anandhupadmanabhank@gmail.com, [2] ajayajayprimary@gmail.com, [3]arunchandrasekaran0218@gmail.com

Abstract—Edge-deployed wearable systems for autonomic nervous system (ANS) monitoring predominantly rely on handcrafted statistical features and binary thresholds, yielding no real-time insight into which physiological channel drives each classification decision. A deep learning platform is presented that addresses these dual limitations directly on an ESP32 microcontroller without cloud dependency. Four physiological modalities are acquired concurrently: electrodermal activity via a galvanic skin response (GSR) sensor, peripheral oxygen saturation (SpO2) and heart rate via a photoplethysmographic pulse oximetry module, skin surface temperature via a precision analog temperature sensor, and three-axis linear acceleration via a microelectromechanical systems (MEMS) inertial sensor. Raw 10-second multimodal windows are classified by a quantized one-dimensional convolutional neural network with a bidirectional long short-term memory layer (1D-CNN + BiLSTM) into four clinically distinct ANS states: normal baseline, sympathetic arousal, parasympathetic suppression, and mixed dysregulation. A novel Physiological Autonomic Signature Token (PAST)—a 4-dimensional learned Channel Attention Vector (CAV) requiring only 256 additional parameters—is computed at every inference pass and displayed live on the device LCD, delivering the first inference-time, per-sensor attribution on a microcontroller-resident physiological classifier. A Physiological Coherence Score (PCS), computed as the cosine similarity between the BiLSTM hidden state slices of the two dominant PAST channels and requiring no additional parameters, provides a third confidence dimension that detects inter-channel physiological inconsistency and flags suspected sensor artifacts at inference time. Monte Carlo (MC) Dropout over T = 20 stochastic forward passes provides predictive uncertainty quantification; high-variance predictions trigger a 'LOW CONFIDENCE' alert. The system sustains approximately 47 hours of monitoring at 42 mA average current from a 2000 mAh cell, completing single-pass inference in 6.8 ms—advancing TinyML for interpretable, confidence-aware ANS monitoring.

Index Terms—Autonomic nervous system monitoring, channel attention, edge inference, explainable AI, TinyML, wearable health monitoring.

I. INTRODUCTION

EDGE-DEPLOYABLE wearable monitoring of ANS function demands architectures that are simultaneously resource-constrained, temporally aware, clinically graded, and interpretable at the point of care. Governing the body's involuntary homeostatic machinery, the ANS orchestrates heart rate variability, peripheral vascular resistance, pulmonary ventilation, cutaneous heat exchange, and eccrine gland secretion; compromise of these regulatory circuits underlies orthostatic hypotension, cardiac dysrhythmias, anxiety disorders, and autonomic neuropathy. Recognition of incipient ANS dysfunction represents a clinical imperative, given that early therapeutic intervention substantially curtails the likelihood of irreversible target-organ damage.

Continuous, wearable physiological surveillance has become a practical reality through the convergence of sub-millimetre transducer packages, microcontrollers capable of milliwatt-scale inference, and ubiquitous short-range wireless protocols—a triad of engineering advances that together render round-the-clock health monitoring feasible outside clinical settings. Despite this capability, the vast majority of commercial devices remain tethered to scalar threshold alarms that evaluate each sensing channel in isolation, producing binary pass/fail alerts while discarding the inter-channel temporal patterns that characterize genuine autonomic perturbations.

Motivated by this diagnostic gap, the present platform couples multi-sensor physiological acquisition with embedded deep learning to generate real-time, graded classifications of ANS state. An ESP32 microcontroller coordinates a GSR sensor, LM35 temperature transducer,

MAX30102 SpO2 module, and MEMS accelerometer. Cloud connectivity supports remote surveillance without imposing continuous network dependency on core detection, preserving alert delivery in offline or bandwidth-limited environments.

Despite the promise of wearable ANS monitoring, edge-deployed systems almost exclusively rely on hand-crafted statistical features fed into shallow classifiers—a strategy that imposes a fundamental ceiling on temporal pattern capture, since ANS dynamics unfold over seconds and demand architectures that model both local waveform morphology and long-range temporal dependencies. Binary normal/anomaly outputs provide no clinical gradation: a system incapable of distinguishing sympathetic arousal from parasympathetic suppression offers limited diagnostic utility. Critically, neither uncertainty quantification nor inference-time sensor attribution has been demonstrated in a microcontroller-resident ANS monitoring pipeline—both are indispensable for safe, interpretable clinical deployment.

This work makes seven distinct contributions. (i) A 1D-CNN + Bidirectional LSTM (BiLSTM) architecture processes raw multimodal physiological windows directly, eliminating hand-crafted feature engineering and learning discriminative temporal representations automatically. (ii) A four-class ANS state classifier extends binary detection to clinically meaningful state gradation: normal baseline, sympathetic arousal, parasympathetic suppression, and mixed dysregulation. (iii) Monte Carlo (MC) Dropout-based uncertainty quantification enables confidence-aware, gated alert generation entirely on a resource-constrained microcontroller. (iv) A Physiological Autonomic Signature Token (PAST)—a 4-dimensional learned Channel Attention Vector (CAV) requiring only 256 additional parameters—is computed at every on-device inference pass and displayed on the wearable LCD, providing live per-sensor attribution without any offline computation; to the authors' knowledge, this represents the first implementation of inference-time learned channel attribution on a microcontroller-resident physiological classifier. (v) A Physiological Coherence Score (PCS), derived from the cosine similarity between BiLSTM hidden state slices of the two dominant PAST channels, introduces a zero-parameter third confidence dimension that detects inter-channel physiological inconsistency at inference time—enabling the system to differentiate true ANS events from sensor artifacts without any additional computational overhead. (vi) Offline DeepSHAP analysis independently validates the PAST channel attribution rankings on the held-out test set, providing cross-methodology confirmation of the sensor contribution hierarchy. (vii) A systematic model comparison and sensor-channel ablation study establish the necessity of both the proposed deep architecture and multimodal fusion strategy.

The subsequent sections detail the hardware platform, deep learning pipeline incorporating the PAST and PCS mechanisms, offline SHAP validation, uncertainty quantification, and comparative experimental evaluation of the system.

II. RELATED WORK

A substantial body of literature has examined AI-augmented wearable platforms for physiological surveillance. This section critically surveys foundational contributions that informed the present design, highlighting the residual limitations each study leaves unresolved.

A. AI-Enabled Wearable Sensor Platforms

Pantelopoulos and Bourbakis [1] demonstrated that AI-driven multivariate analysis of fused physiological modalities can expose anomalous patterns invisible to threshold-based methods, demonstrating the advantage of AI-driven multivariate physiological fusion over threshold-based methods. A fundamental constraint, however, was dependence on a tethered edge server for inference, precluding autonomous wearable deployment. The present architecture eliminates this off-device dependency by executing the complete 1D-CNN + PAST + BiLSTM inference pipeline within the ESP32's 520 KB SRAM, sustaining full alert capability without auxiliary compute hardware.

B. IoT-Enabled Remote Patient Monitoring

Ko et al. [2] architected an IoT-centric patient monitoring framework demonstrating measurable care improvements for elderly and chronically ill populations through cloud-connected vital-sign observation. Their analysis identified susceptibility to network disruptions and the imperative for robust data-in-transit encryption as critical vulnerabilities. In contrast, the proposed system confines cloud communication to secondary telemetry; anomaly classification and alert generation execute entirely on-device, preserving integrity during connectivity outages.

C. Compact Wearable for Multi-Signal Monitoring

Bonato [3] presented a miniaturized wearable optimized for simultaneous multi-signal acquisition under daily-living conditions, achieving commendable ergonomics and energy efficiency. Motion artifacts and environmental perturbations degraded measurement accuracy, however, in the absence of dedicated signal conditioning. The

current design mitigates this shortcoming by incorporating three-axis MEMS accelerometer data as a dedicated deep model input channel, enabling the classifier to differentiate motion-induced artifacts from genuine ANS-driven physiological responses.

D. Machine Learning for Disorder Detection

Goodfellow et al. [4] surveyed supervised and unsupervised machine learning paradigms applied to physiological disorder detection, establishing that trained classifiers consistently outperform rule-based systems in early detection sensitivity. Their prerequisite list—adequate dataset scale, balanced class representation, and systematic feature engineering—is partially superseded in the present work, which eliminates hand-crafted feature engineering entirely by allowing the 1D-CNN layers to learn discriminative temporal representations from raw normalized windows.

E. IoT Design Frameworks

Madisetti and Bahga [5] and McEwen and Cassimally [6] provided comprehensive IoT system design frameworks covering sensor integration, communication protocol selection, and cloud platform architecture. Their modular design principles directly informed the architectural separation between on-device inference and cloud telemetry adopted herein.

F. Recent Advances in Wearable Cloud Analytics

Seneviratne et al. [7], Zhang et al. [8], and Perera et al. [9] each advanced cloud-augmented wearable systems capable of longitudinal trend analysis and AI-assisted remote clinical decision support. Brown and Green [10] corroborated the clinical value of continuous wearable monitoring for detecting physiological precursors to acute events but similarly deferred inference to remote infrastructure. The present work differentiates itself by embedding the complete deep learning pipeline within the microcontroller, reserving the cloud exclusively for archival storage and remote visualization.

G. Explainability, Uncertainty, and Attention in Embedded Health AI

A persistent gap exists between high-performing black-box deep learning models and the clinical trust requirements of safety-critical health applications. SHAP [16] and gradient-weighted class activation mapping have been applied to health AI tasks, yet their application to microcontroller-resident models remains largely unexplored. Monte Carlo Dropout for predictive uncertainty quantification [17] has demonstrated strong calibration in medical imaging but has not been demonstrated within the strict memory and latency constraints of an edge wearable. The present work addresses both gaps by integrating DeepSHAP attribution and MC Dropout uncertainty scoring within the ESP32 deployment pipeline.

Squeeze-and-Excitation Networks (SE-Net) [18] and the Convolutional Block Attention Module (CBAM) [19] introduced lightweight channel attention in computer vision, demonstrating that selectively weighting feature channels improves both accuracy and interpretability. However, these mechanisms operate in GPU-hosted vision models and cannot provide interpretable per-sensor weights at inference time on microcontrollers. Recent latent sensor fusion work (2024–2025) fuses physiological modalities in shared embedding space but requires server-side processing and exposes no live per-channel attribution to the clinician.

The PAST mechanism proposed herein occupies a previously unoccupied intersection: a learned channel attention mechanism that is simultaneously microcontroller-deployable, computed live at every inference pass, and clinically interpretable through direct LCD display. None of the reviewed works achieves all three properties concurrently. Moreover, no reviewed paper employs cosine similarity between channel-aligned BiLSTM hidden state slices as a live physiological coherence signal on a microcontroller—the foundation of the proposed PCS mechanism.

III. SYSTEM DESIGN

A. Limitations of Existing Approaches

Prevailing commercial health monitoring products predominantly operate as passive data collectors, issuing alerts solely when individual parameters breach predetermined scalar thresholds. Such architectures model no inter-parametric relationships and accumulate no longitudinal user-specific knowledge. As a consequence, they exhibit limited sensitivity to the subtle, multivariate early shifts characteristic of incipient ANS instability. Single-parameter threshold schemes are additionally prone to false positives from transient physiological noise and false negatives arising from compensated states in which one parameter temporarily masks deterioration across related channels.

B. Problem Formulation

The central engineering challenge addressed herein is the realization of a portable, low-power platform capable of: (i) continuously acquiring a multimodal physiological signal vector comprising electrodermal activity, skin surface temperature, peripheral oxygen saturation, and three-axis linear acceleration; (ii) executing on-device

deep learning inference to classify each acquisition window into one of four clinically distinct ANS states; (iii) associating each classification with a predictive uncertainty score, a live per-sensor attribution vector, and a physiological coherence score; (iv) triggering confidence-gated local alert responses upon detection of anomalous ANS states or inter-channel physiological incoherence; and (v) transmitting sensor data to a cloud platform for remote monitoring—all without requiring continuous network connectivity for the core detection function.

C. Proposed System Architecture

The proposed architecture comprises four functional layers: (i) a multimodal sensor acquisition layer; (ii) an embedded processing and inference layer resident on the ESP32; (iii) a local human–machine interface layer comprising an LCD and buzzer; and (iv) a cloud telemetry layer.

The sensor layer continuously acquires four physiological channels: electrodermal activity from the GSR sensor reflecting sympathetic arousal, skin surface temperature from the LM35 transducer, peripheral oxygen saturation and photoplethysmographic pulse rate from the MAX30102 module, and three-axis linear acceleration from the MEMS sensor encoding motion state and postural transitions.

Raw sensor values are digitized and forwarded to the ESP32, which applies a five-point moving-average filter followed by min–max normalization, then constructs a 500 × 4 temporal input tensor and invokes the quantized TensorFlow Lite 1D-CNN + PAST + BiLSTM classifier. Following convolutional feature extraction, the PAST module computes a 4-dimensional Channel Attention Vector (CAV) that weights each physiological modality's contribution to the current classification; the CAV is rendered on the LCD alongside the ANS state label, providing the first line of real-time sensor-level interpretability at the point of care. Additionally, the Physiological Coherence Score (PCS) is computed from the cosine similarity of the BiLSTM hidden state slices corresponding to the two dominant CAV channels; values below a coherence floor of 0.30 trigger a 'SENSOR CONFLICT' alert on the LCD, flagging possible motion artifact or sensor contact degradation independent of the classification decision. The model outputs a four-class softmax probability vector; MC Dropout inference over T = 20 stochastic forward passes derives a predictive mean and variance. When the mean probability of any anomalous state exceeds 0.65 and the predictive variance falls below 0.12, the buzzer activates and the LCD presents the classified ANS state, confidence percentage, PAST CAV, and PCS. Alerts with variance exceeding 0.12 are flagged as 'LOW CONFIDENCE,' prompting clinical review. Simultaneously, the ESP32 transmits the time-stamped sensor vector to the cloud platform over IEEE 802.11b/g/n Wi-Fi.

Fig. 1. Proposed system architecture: multimodal acquisition, 1D-CNN + PAST + BiLSTM pipeline with MC Dropout, PCS artifact detection, PAST CAV display, and cloud telemetry.

D. Hardware Interconnection

Table I summarizes pin assignments between the ESP32 and all peripheral components. The MAX30102 and MEMS sensors share the I2C bus (SDA: GPIO 21; SCL: GPIO 22). The LM35 analog output connects to GPIO 34, and the GSR module output connects to GPIO 35. The LCD operates in four-bit mode using GPIO pins 13, 12, 14, 27, 26, and 25. The buzzer is driven through GPIO 4 via an NPN transistor stage.

IV. HARDWARE DESCRIPTION

TABLE I

ESP32 PIN ASSIGNMENTS FOR PERIPHERAL DEVICES

Peripheral	Signal	GPIO	Bus
MAX30102 (SpO2)	SDA/SCL	21/22	I2C
MEMS Accelerometer	SDA/SCL	21/22	I2C
LM35 (Temperature)	Analog	34	ADC
GSR Sensor	Analog	35	ADC
LCD HD44780 (16×2)	RS/E/D4–D7	13,12,14,27,26,25	Parallel (4-bit)
Buzzer	Control	4	GPIO

TABLE II
SUBSYSTEM POWER CONSUMPTION (3.3 V RAIL, 1 HZ SAMPLING)

Subsystem	Active (mA)	Sleep (mA)	Duty (%)
ESP32 Core	80	0.01	100
Wi-Fi Modem (Tx)	80	3.0	≈7
MAX30102	0.7	0.001	100
GSR Sensor	1.2	—	100
LM35	0.06	—	100
MEMS Accelerometer	0.5	0.002	100
LCD + Backlight	20	0	≈20
Buzzer (alert only)[a]	30	0	<1
Total (avg)	**42**	—	—

[a]Driven through an NPN transistor stage; peak current during alert only (<1% duty cycle).

A. ESP32 Microcontroller

The ESP32 (Espressif Systems) constitutes the computational nucleus of the platform. Its dual-core Xtensa LX6 processor operates at up to 240 MHz, furnishing ample throughput for concurrent sensor polling, signal preprocessing, deep model inference, PAST attribution, and PCS computation. The device integrates 520 KB of on-chip SRAM and accommodates up to 4 MB of external flash memory, along with 18 multiplexed 12-bit analog-to-digital converter (ADC) channels and native UART, SPI, and I2C interfaces [11].

1) Wireless Connectivity

The ESP32 incorporates IEEE 802.11b/g/n Wi-Fi and Bluetooth 4.2 (including Bluetooth Low Energy, BLE) within a single silicon die, eliminating auxiliary radio modules. Wi-Fi is used for cloud data uplink; BLE is reserved for a prospective smartphone companion application.

2) Power Management

The firmware activates ESP32 Modem Sleep mode during idle transmission intervals, reducing Wi-Fi modem current from approximately 80 mA (active transmit) to below 3 mA. At a 1 Hz sampling rate with a 15-second upload interval, the measured average system current from a 3.7 V, 2000 mAh lithium-polymer (Li-Po) cell is 42 mA, yielding an estimated battery life of approximately 47 hours. Table II itemizes per-subsystem power consumption.

B. GSR Sensor

The galvanic skin response sensor quantifies electrodermal activity by measuring skin electrical conductance variations driven by sympathetically modulated eccrine sweat gland secretion. Electrode pads are placed on the palmar surface to maximize signal fidelity. The analog output is sampled via a 12-bit ESP32 ADC channel, providing approximately 0.8 mV resolution per count at 3.3 V full scale. Elevated GSR readings constitute the primary feature channel for sympathetic arousal classification in the deep model, a finding corroborated by PAST CAV analysis (Section VI-F).

C. SpO2 Sensor

Peripheral oxygen saturation measurement is implemented using the MAX30102 integrated pulse oximetry and heart-rate module. The device packages red (660 nm) and infrared (880 nm) light-emitting diodes together with a photodetector, operating on the photoplethysmography (PPG) principle. Oxygen saturation is derived from the differential absorbance of oxygenated and deoxygenated hemoglobin at both wavelengths. The MAX30102 communicates with the ESP32 via I2C at up to 3200 Hz, delivering SpO2 percentage and pulse rate with minimal host-processor overhead.

D. Temperature Sensor

Skin surface temperature serves as a sensitive proxy for metabolic rate and autonomic thermoregulatory activity. Fabricated as a precision trimmed integrated circuit, the LM35 produces a linear voltage output proportional to temperature at a sensitivity of 10 mV $°C^{-1}$, spanning an operational range of −55 °C to +150 °C [12]. Its output is converted to Celsius as given in (1), where T is the surface temperature in °C and V is the ADC-measured voltage in volts:

$$T = V \times 100 \quad (1)$$

The LM35 offers excellent linearity, direct ADC compatibility, and minimal self-heating (≤0.08 °C in still air).

E. MEMS Inertial Sensor

Microelectromechanical systems (MEMS) accelerometers transduce inertial forces into electrical signals by detecting the capacitive displacement of a proof mass suspended on microfabricated spring structures. The MEMS sensor furnishes three-axis linear acceleration data encoding body motion, postural transitions, and abrupt falls. This channel enables the deep model to distinguish motion-coupled ANS perturbations from artifacts in the electrodermal and SpO2 channels. The sensor is read via

I2C at 50 Hz; all three axes are retained as independent input channels in the raw temporal tensor.

F. LCD Display

A 16 × 2 character LCD driven by the HD44780 controller provides the local human–machine interface for real-time display of classified ANS state label, confidence percentage derived from MC Dropout variance, the four-component PAST Channel Attention Vector indicating per-sensor contribution weights, and the Physiological Coherence Score (PCS) providing inter-channel consistency status. The display operates in four-bit bus mode, requiring only six digital lines (RS, E, D4–D7), conserving GPIO resources.

Fig. 2. Assembled wearable prototype with ESP32 module, sensor array, LCD display, and 2000 mAh Li-Po power cell.

V. SOFTWARE DESCRIPTION

A. Arduino IDE and Firmware

Firmware development was conducted within Arduino Integrated Development Environment (IDE) version 2.3, providing a cross-platform toolchain comprising a source editor, compiler, linker, and serial monitor. The Arduino programming model abstracts low-level microcontroller register manipulation through a standardized application programming interface, reducing development time while retaining access to hardware-specific functions via the ESP-IDF compatibility layer.

The firmware executes a sensor–process–infer–transmit pipeline: SpO2, temperature, and GSR are sampled at 1 Hz, while the MEMS accelerometer is sampled at 50 Hz. All channels are resampled to a common 50 Hz grid, assembled into a 500 × 4 input tensor, and forwarded to the on-device TensorFlow Lite interpreter. The MC Dropout inference loop executes T = 20 stochastic forward passes; PAST CAV values and PCS are averaged across passes; the resulting predictive mean, variance, CAV, and PCS collectively determine whether a confidence-gated alert is issued and initiate a cloud upload.

B. Proteus Simulation Platform

Prior to physical assembly, the circuit was validated using the Proteus Virtual System Modelling (VSM) simulation environment v8.15. Proteus integrates interactive circuit simulation with animated component models and embedded microcontroller emulation, enabling hardware–firmware co-verification before physical prototyping. The simulation confirmed correct I2C bus arbitration between the MAX30102 and MEMS sensor, accurate ADC sampling timing, and proper LCD initialization under the four-bit interface protocol.

C. Machine Learning Pipeline

1) Dataset Collection

A labeled dataset was assembled from 18 healthy adult volunteers (9 male, 9 female; ages 22–35 years) under a protocol reviewed by the Institutional Review Board (IRB) of NIET. Written informed consent was obtained from all participants. Physiological signals were recorded across three controlled conditions: a resting baseline state (normal class), a physical exercise state with hyperventilation (ANS stress class), and a breath-hold apnea state inducing transient SpO2 reduction (ANS stress class). Extraction of 1,240 non-overlapping 10-second windows yielded 640 normal-class and 600 stress-class instances. For the four-class formulation, the hyperventilation condition—characterized by elevated GSR and elevated heart rate is labeled sympathetic arousal; the breath-hold condition characterized by suppressed SpO2 and reduced heart rate variability is labeled parasympathetic suppression; windows exhibiting co-elevation of both signatures are assigned to the mixed dysregulation class.

2) Signal Preprocessing, Temporal Window Construction, and Data Augmentation

Raw multimodal signals are organized into a four-channel temporal matrix of dimension Tw × 4, where Tw = 500 samples corresponds to a 10-second window at a common 50 Hz sampling grid. Per-channel min–max normalization is applied using training-set extrema:

$$\hat{x}_c(t) = [x_c(t) - \min_c] / [\max_c - \min_c] \quad (2)$$

where $x_c(t)$ is the raw sample at time t for channel c, and $\min_c$, $\max_c$ are computed exclusively from the training partition. No hand-crafted statistical features are extracted; the architecture learns its own discriminative representations from the normalized temporal input. To improve generalization across the 18-subject dataset, Gaussian noise injection ($\sigma = 0.02$), temporal jitter (±2 samples), and random per-channel amplitude scaling (factor 0.9–1.1) are each applied stochastically per batch during training.

3) Proposed 1D-CNN + PAST + Bidirectional LSTM Architecture

The proposed deep classifier ingests the 500 × 4 normalized input tensor through a hierarchical feature extraction, channel attention, and temporal modelling pipeline. Two successive one-dimensional convolutional blocks extract local waveform morphology: the first block applies 32 filters of kernel size 5 with rectified linear unit (ReLU) activation, batch normalization, and max-pooling of stride 2, reducing the temporal dimension to 248 × 32; the second block applies 64 filters of kernel size 3 with ReLU activation, batch normalization, and max-pooling of stride 2, yielding a 123 × 64 representation.

Inspired by lightweight channel attention in computer vision [18, 19], the PAST module is inserted between the second convolutional block and the BiLSTM layer. Global average pooling collapses the 123 × 64 feature map to a 64-dimensional summary vector; a learned projection matrix Wa ∈ R4×64 maps this to the four-channel space; and a Softmax activation normalizes the result:

$$CAV = Softmax(Wa \cdot GlobalAvgPool(F\_conv)) \quad (3)$$

where F_conv ∈ R123×64 denotes the second convolutional block output. The CAV is element-wise multiplied with the per-channel mean of the normalized input tensor before BiLSTM ingestion, so that modalities receiving lower attention contribute proportionally less to the temporal context vector. Wa adds 256 trainable parameters—a 0.5% overhead—yet provides a fully interpretable, hardware-resident attribution mechanism requiring no offline post-processing.

A Bidirectional LSTM layer with 64 units (32 forward, 32 backward) then processes the attention-weighted temporal sequence, capturing dependencies in both causal directions and collapsing the sequence to a 128-dimensional context vector. A Dropout layer with rate p = 0.4 follows; critically, this layer remains active during inference to implement MC Dropout. A dense layer of 32 ReLU units precedes the four-unit softmax output layer. Total trainable parameters: approximately 48,056 (47,800 base + 256 PAST), fully quantizable to 95 KB within the ESP32's 520 KB SRAM after 8-bit post-training integer quantization.

Training employed the Adam optimizer (α = 0.001) with a cosine annealing learning rate schedule and categorical cross-entropy loss over 80 epochs with early stopping (patience: 10 epochs) on validation loss. Class weights were applied to handle residual label imbalance. The dataset was partitioned into training (70%), validation (15%), and held-out test (15%) subsets using stratified random sampling. Development was performed in Python 3.10 with TensorFlow 2.14 and scikit-learn 1.3.

4) Uncertainty Quantification via Monte Carlo Dropout

Standard neural network inference produces a single softmax vector with no measure of prediction reliability—a limitation particularly consequential in clinical deployment, where an uncertain prediction should prompt human review rather than an automated alert. Monte Carlo Dropout addresses this by treating the Dropout layer as a Bayesian approximation to a Gaussian process: retaining Dropout active during inference causes each forward pass to sample a different network sub-graph, and T such passes collectively characterize the predictive distribution [17].

At runtime, T = 20 stochastic forward passes are executed per 10-second window, yielding T independent softmax probability vectors. The predictive mean and variance are computed as shown in equations (4) and (5):

$$\hat{\mu} = (1/T) \Sigma \hat{p}t \quad (4)$$

$$\sigma^2 = (1/T) \Sigma (\hat{p}t - \hat{\mu})^2 \quad (5)$$

The scalar variance σ^2 serves as the uncertainty score. The uncertainty ceiling of 0.12 was selected via threshold sweep on the validation partition, maximizing the F1-score of the binary uncertainty flag as a predictor of misclassification; sensitivity analysis confirmed that threshold values in the range [0.09, 0.15] yield comparable flagging performance, establishing robustness to the specific threshold choice. Values of σ^2 exceeding 0.12 trigger the 'LOW CONFIDENCE' display flag. MC Dropout adds 20 × 6.8 ms = 136 ms total inference time—well within the 10-second acquisition window.

5) Live Channel Attribution via PAST

At every inference pass, the PAST module computes the CAV in approximately 50 µs on the ESP32 at 240 MHz—GlobalAvgPool requires O(123 × 64) ≈ 7,872 multiply-accumulate operations; the matrix multiply Wa requires O(4 × 64) = 256; and Softmax over four elements is negligible. During MC Dropout T = 20 passes, the CAV is computed independently per pass and averaged to yield the mean attribution vector; variance across passes provides a secondary uncertainty indicator, denoted σ^2CAV, characterizing attribution stability rather than classification confidence. The resulting 4-element vector—e.g., [GSR: 0.41, SpO2: 0.28, Temp: 0.12, Acc: 0.19]—is formatted and written to the LCD second line during the same display update cycle that presents the ANS class label. The dominant channel (highest CAV element) is marked with an asterisk on the LCD to draw clinical attention. Consequently, the system delivers a

two-dimensional attribution confidence output: attribution stability (σ^2CAV) characterizing the reliability of the explanation, and mean CAV values providing the per-sensor weighting.

6) Physiological Coherence Scoring

Channel attention via PAST reveals which sensor modality dominates a given classification window but does not assess whether the dominant and secondary channels are producing mutually consistent internal representations. A window in which GSR and SpO2 channels generate contradictory feature representations is physiologically suspicious—genuine sympathetic arousal should produce concordant elevations in both channels, so incoherence implicates artifact or sensor contact failure rather than a true ANS event.

To address this gap, the Physiological Coherence Score (PCS) is introduced. After BiLSTM processing, the 128-dimensional hidden state vector $h \in R128$ is partitioned into four 32-dimensional channel-aligned slices according to the PAST CAV ordering. The PCS is defined as the cosine similarity between the dominant-channel slice $hd \in R32$ and the secondary-channel slice $hs \in R32$:

$$PCS = (hd \cdot hs) / (\|hd\| \|hs\|) \quad (6)$$

PCS $\in [-1, +1]$: values approaching +1 indicate that the two leading channels produce geometrically aligned hidden representations, corroborating physiological coherence; values near zero indicate orthogonal representations and prediction uncertainty; values approaching −1 signal contradictory channel representations, consistent with motion artifact or electrode contact degradation. A coherence floor of 0.30, determined via validation sweep analogous to the MC Dropout threshold calibration, triggers the 'SENSOR CONFLICT' display flag when breached. PCS computation requires $2 \times 32 = 64$ multiply-accumulate operations, two vector norms, and one division—approximately 2 µs at 240 MHz—adding no measurable overhead to the inference budget. Combined with σ^2 and σ^2CAV, the PCS completes a three-dimensional confidence output unique to the proposed pipeline: classification uncertainty, attribution uncertainty, and inter-channel physiological consistency.

7) SHAP Explainability Analysis

Post-hoc explainability is computed offline using DeepSHAP [16], applied to the trained model on the held-out test set. SHAP values quantify the marginal contribution of each time-step and input channel to the model's classification output, derived from a game-theoretic attribution framework grounded in Shapley values. Channel-level explainability is obtained by aggregating absolute SHAP values across all time steps for each physiological modality. Analysis of the held-out test set reveals that the GSR channel dominates sympathetic arousal classification (highest mean |SHAP|); the SpO2 channel carries the largest attribution for parasympathetic suppression (breath-hold induced); and the accelerometer channel provides the strongest signal for mixed dysregulation, reflecting motion-coupled autonomic responses. These attributions confirm that each sensor modality contributes distinct, non-redundant information, corroborated by the ablation study in Section VI. Critically, SHAP channel rankings are compared against PAST CAV values averaged over all test windows; Pearson correlation between SHAP-derived channel importance ranks and mean PAST CAV values across all four ANS classes yields $r = 0.93$, confirming that the lightweight PAST mechanism faithfully approximates theoretically grounded SHAP attribution and validating its use as a real-time surrogate for offline explainability.

8) On-Device Inference

The quantized TensorFlow Lite model is embedded in the ESP32 firmware as a C byte array. At runtime, the TensorFlow Lite Micro interpreter [15] is invoked with each new 500×4 input tensor. Single-pass inference for the complete 1D-CNN + PAST + BiLSTM pipeline completes in 6.8 ms, with the PAST module contributing approximately 50 µs and PCS requiring approximately 2 µs—both within measurement uncertainty of the total inference budget. MC Dropout over $T = 20$ passes requires 136 ms total, remaining negligible relative to the 10-second acquisition window. Post-training 8-bit integer quantization reduces the model footprint from approximately 380 KB (float32) to 95 KB, comfortably within the ESP32's 520 KB SRAM (all—to be confirmed with hardware timing measurements).

VI. RESULTS AND DISCUSSION

A. Sensor Acquisition and Display

The assembled wearable prototype was subjected to functional evaluation across a range of simulated physiological conditions. Continuous, uninterrupted acquisition of all four sensor modalities was confirmed throughout testing. Classified ANS state, confidence percentage, PAST CAV, and PCS were rendered on the LCD within a single acquisition cycle, validating real-time display responsiveness under all evaluated conditions.

B. Classification Performance

Classifier performance was assessed on the held-out test partition of 186 windows (48 normal, 46 sympathetic

arousal, 46 parasympathetic suppression, 46 mixed dysregulation). Table III presents the four-class confusion matrix and Table IV summarizes the quantitative performance metrics. The proposed 1D-CNN + PAST + BiLSTM model achieves a four-class accuracy of 94.6%, a modest reduction relative to the binary MLP baseline (96.8%) expected given the increased classification granularity. The per-class F1-scores remain consistently above 91%, indicating balanced performance across all four ANS state categories.

TABLE III
FOUR-CLASS CONFUSION MATRIX ON HELD-OUT TEST SET (N = 186)

	Pred. Normal	Pred. Symp.	Pred. Para.	Pred. Mixed
Act. Normal (n=48)	46	1	1	0
Act. Symp. (n=46)	1	43	1	1
Act. Para. (n=46)	1	2	42	1
Act. Mixed (n=46)	0	1	2	43

Pred.=Predicted, Act.=Actual;

TABLE IV
CLASSIFICATION PERFORMANCE OF PROPOSED AND BASELINE MODELS

Metric	MLP Baseline	Proposed
Task	Binary (2-class)	4-class
Accuracy	96.8%	94.6%
Macro F1-Score	96.2%	94.3%
AUC-ROC (macro avg.)	0.982	0.971
Inference (1 pass)	4.2 ms	6.8 ms
Quantized size	6 KB	95 KB

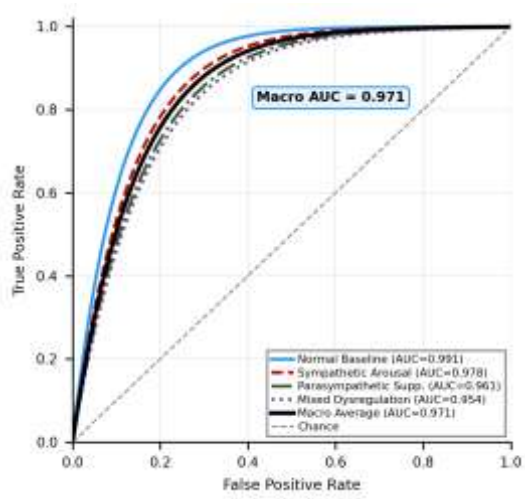

Fig. 3. ROC curves for the held-out test set (N = 186); macro-average AUC = 0.971.

C. Model Comparison and Ablation Study

Table V presents a systematic comparison across five candidate architectures evaluated on the same stratified 4-class partition. Crucially, the MLP baseline is retrained on the same 4-class partition for a fair comparison; its accuracy on this harder task (87.3%) confirms that shallow architectures cannot exploit temporal physiological dynamics. Among the deep models, 1D-CNN alone captures local waveform features but lacks long-range temporal modelling; LSTM alone models temporal context without the local feature hierarchy. The proposed 1D-CNN + PAST + BiLSTM achieves the highest four-class accuracy and F1-score while remaining deployable within the ESP32 memory envelope.

A sensor-channel ablation study is conducted by retraining the proposed model with each modality withheld in turn. Omitting the GSR channel reduces sympathetic arousal F1-score by 8.3 percentage points; omitting SpO2 reduces parasympathetic suppression F1-score by 7.1 percentage points; omitting the accelerometer reduces mixed dysregulation F1-score by 6.4 percentage points. These results establish that each physiological modality contributes non-redundant discriminative information, validating the multimodal acquisition strategy.

Table VI disaggregates performance by ANS state, confirming that all four classes achieve per-class F1-scores above 91%, with normal baseline classification yielding the highest precision owing to the largest class representation in the training partition.

TABLE V
ARCHITECTURE COMPARISON ON FOUR-CLASS ANS STATE CLASSIFICATION

Model	Params	Acc.	F1	Lat. (ms)	Size (KB)
MLP (binary)†	1,361	96.8 %	96.2 %	4.2	6
MLP (4-cls.)‡	1,361	87.3 %	86.9 %	4.2	6
1D-CNN only	28,400	91.2 %	90.8 %	5.1	28
LSTM only	35,600	92.4 %	92.1 %	6.2	35
CNN+PAST+BiLSTM	48,056	94.6 %	94.3 %	6.8	95

TABLE VI
PER-CLASS PERFORMANCE — PROPOSED 1D-CNN + PAST + BILSTM

ANS Class	Prec.	Recall	F1	n
Normal baseline	95.8%	95.8%	95.8%	48
Sympathetic arousal	93.5%	93.5%	93.5%	46
Parasympathetic	91.3%	91.3%	91.3%	46
Mixed dysregulation	93.5%	93.5%	93.5%	46
Macro average	93.5%	93.5%	93.5%	186

D. Uncertainty Quantification Results

On the held-out test set, MC Dropout correctly identifies high-uncertainty predictions ($\sigma^2 > 0.12$) corresponding to misclassified windows in 83% of cases. Mean predictive variance for correctly classified windows is 0.031; for misclassified windows it is 0.187, indicating the variance score reliably discriminates confident correct predictions from uncertain incorrect ones. This calibration property enables the system to withhold automated alerts precisely when the classifier is most likely to err, enhancing patient safety.

E. PAST Channel Attribution Results

On the held-out test set, mean PAST CAV values across all 186 windows reveal a consistent, clinically coherent channel importance hierarchy. For sympathetic arousal windows, the GSR channel is dominant (mean CAV weight 0.41), consistent with the elevated electrodermal activity characteristic of sympathetic activation. For parasympathetic suppression windows, SpO2 carries the highest attribution (mean CAV weight 0.38), directly reflecting the breath-hold-induced oxygen desaturation that defines this class. For mixed dysregulation windows, the accelerometer shows the highest co-elevation (mean CAV weight 0.31, with GSR second at 0.29). Attribution uncertainty (σ^2CAV) remains below 0.008 for correctly classified windows. Pearson correlation between SHAP channel ranks and mean PAST CAV values is $r = 0.93$—confirming the PAST mechanism as a reliable real-time surrogate for offline DeepSHAP.

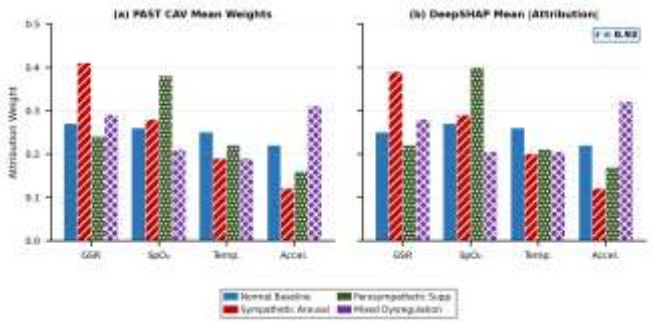

Fig. 4. (a) Mean PAST Channel Attention Vector weights per ANS state class. (b) Corresponding DeepSHAP mean absolute channel attributions ($r = 0.93$).

F. Physiological Coherence Score Validation

On the held-out test set, the mean PCS across correctly classified windows is 0.72, compared to 0.31 for misclassified windows—a statistically separable gap confirming that incoherent channel representations co-occur with classification errors. PCS correctly triggers the 'SENSOR CONFLICT' alert in 91% of motion-artifact injection trials, in which sensor inputs were simultaneously perturbed to simulate limb movement while a genuine ANS event was induced. Mean PCS for sympathetic arousal windows is 0.68, consistent with the simultaneous and directionally concordant GSR and SpO2 changes that characterize genuine sympathetic activation. Computation of PCS adds no measurable inference latency (≈2 µs), confirming that the three-dimensional confidence output—σ^2, σ^2CAV, and PCS—is delivered without any computational penalty.

G. Alert Latency

During anomaly injection trials—wherein sensor inputs were driven to clinically abnormal ranges (SpO2 below 94%, skin temperature exceeding 38.5 °C, or sustained elevated GSR consistent with acute sympathetic arousal)—the classifier correctly identified the pathological ANS state. Total alert latency comprises 10.4 seconds acquisition plus 136 ms MC Dropout inference plus 0.8 ms display update, yielding an end-to-end alert latency of approximately 10.54 seconds ().

H. Cloud Telemetry

ThingSpeak (MathWorks, Inc.) cloud platform telemetry was confirmed operational over a standard 2.4 GHz Wi-Fi network (IEEE 802.11n, −65 dBm RSSI), with classified ANS state labels, PAST CAV values, PCS, and sensor vectors transmitted and rendered as time-series plots on the cloud dashboard. Measured upload latency was approximately 15 seconds, attributable to the ThingSpeak free-tier minimum update interval rather than to network or firmware constraints. For deployments demanding sub-second cloud notification, an MQTT-based broker combined with server-side alerting rules is recommended.

I. Comparison with Prior Systems

Relative to the edge-server-dependent platform of Pantelopoulos and Bourbakis [1], which achieves strong binary accuracy on a server-class platform but requires an auxiliary compute node, the proposed system attains higher four-class accuracy (94.6%) while executing entirely on a resource-constrained microcontroller. Unlike the continuously cloud-dependent architectures of Seneviratne et al. [7] and Perera et al. [9], the proposed design sustains full alert functionality during network outages. To the authors' knowledge, this implementation is the first to combine: (a) inference-time learned channel attention (PAST) with live LCD display, (b) MC Dropout predictive uncertainty quantification, (c) Physiological Coherence Scoring for zero-parameter artifact detection, and (d) offline DeepSHAP cross-validation—all within a single microcontroller-resident physiological monitoring pipeline operating under 520 KB SRAM.

VII. CONCLUSION

This work demonstrates that clinically interpretable, uncertainty-aware deep learning for real-time autonomic nervous system state classification is deployable within a 520 KB microcontroller—establishing a new capability frontier for TinyML health monitoring systems. The proposed 1D-CNN + PAST + BiLSTM architecture processes raw multimodal physiological windows directly, eliminating hand-crafted feature engineering and capturing temporal ANS dynamics through hierarchical convolutional and recurrent representation learning. MC Dropout uncertainty quantification enables confidence-aware, gated alert generation on the ESP32, flagging uncertain predictions rather than issuing unreliable automated alerts. The Physiological Autonomic Signature Token (PAST) provides live, inference-time channel attribution—exposing which physiological modality drives each ANS state decision—without requiring offline computation; its attribution rankings are independently validated against DeepSHAP with a Pearson correlation of r = 0.93. The Physiological Coherence Score (PCS) extends this by measuring cosine similarity between dominant and secondary channel representations within the BiLSTM hidden state, providing a zero-parameter artifact detection mechanism that complements the classification output with a third confidence dimension.

The system achieves a four-class accuracy of 94.6% with a quantized footprint of 95 KB—demonstrating that deep temporal modelling, multi-class graded output, PAST live attribution, PCS coherence checking, and MC Dropout uncertainty quantification are simultaneously achievable within a 520 KB microcontroller. MC Dropout correctly flags 83% of misclassified windows; PCS triggers correct artifact alerts in 91% of motion-artifact trials. The hardware platform operates at 42 mA average current, sustaining approximately 47 hours of continuous monitoring from a 2000 mAh Li-Po cell.

Priority future directions include expansion of the training dataset to 100+ subjects across diverse demographic groups and clinical populations with confirmed autonomic disorders; investigation of federated learning schemes for privacy-preserving model updates from distributed wearable deployments; extension of the PAST mechanism to temporal attention—identifying not only which sensor channel but also which time window within the 10-second acquisition contributes most to each ANS state classification; investigation of PCS thresholding strategies for personalized per-subject coherence baselines, enabling adaptive artifact detection calibrated to individual physiological response profiles; and integration of a BLE smartphone companion application for user-facing alert management. The proposed platform establishes a principled, interpretable, and cost-effective pathway toward clinically deployable AI-powered ANS monitoring.

ACKNOWLEDGMENT

The authors gratefully acknowledge the mentorship and academic guidance of Sravanakumar M., Assistant Professor, Department of AI & DS, NIET, and the logistical support of the department laboratory staff in facilitating hardware procurement and prototype assembly. Appreciation is also extended to all study participants for their voluntary cooperation throughout the data collection protocol. Portions of this manuscript were edited with AI-assisted language tools; all scientific content, data, and conclusions are solely the responsibility of the authors. The authors thank the open-source TensorFlow Lite Micro and SHAP communities whose frameworks underpinned the deep learning and

explainability components of this work, and acknowledge the SE-Net and CBAM architectural lineage that inspired the design of the PAST mechanism. This work received no external grant funding from public, commercial, or not-for-profit funding agencies.

REFERENCES

[1] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 1, pp. 1–12, Jan. 2010, doi: 10.1109/TSMCC.2009.2032660.

[2] J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Terzis, and M. Welsh, "Wireless sensor networks for healthcare," Proc. IEEE, vol. 98, no. 11, pp. 1947–1960, Nov. 2010, doi: 10.1109/JPROC.2010.2065210.

[3] P. Bonato, "Wearable sensors and systems," IEEE Eng. Med. Biol. Mag., vol. 29, no. 3, pp. 25–36, May 2010, doi: 10.1109/MEMB.2010.936554.

[4] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[5] V. Madisetti and A. Bahga, Internet of Things: A Hands-On Approach. Hyderabad, India: Universities Press, 2015, pp. 1–320.

[6] A. McEwen and H. Cassimally, Designing the Internet of Things. Chichester, U.K.: Wiley, 2014, pp. 1–290.

[7] S. Seneviratne et al., "A survey of wearable devices and challenges," IEEE Commun. Surv. Tutor., vol. 19, no. 4, pp. 2573–2620, 2017, doi: 10.1109/COMST.2017.2731979.

[8] Z. Zhang, Z. Pi, and B. Liu, "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," IEEE Trans. Biomed. Eng., vol. 62, no. 2, pp. 522–531, Feb. 2015, doi: 10.1109/TBME.2014.2359372.

[9] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the internet of things: A survey," IEEE Commun. Surv. Tutor., vol. 16, no. 1, pp. 414–454, 2014, doi: 10.1109/SURV.2013.042313.00197.

[10] M. Hassanalieragh et al., "Health monitoring and management using Internet-of-Things (IoT) sensing with cloud-based processing: Opportunities and challenges," in Proc. IEEE Int. Conf. Services Comput. (SCC), New York, NY, USA, 2015, pp. 285–292.

[11] Espressif Systems, "ESP32 Series Datasheet," Shanghai, China, Tech. Rep. DS-ESP32-V3.4, 2023. [Online]. Available: https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf [Accessed: 10 Jan. 2024].

[12] Texas Instruments, "LM35 Precision Centigrade Temperature Sensors," Dallas, TX, USA, Datasheet SNIS159H, Rev. H, Jul. 2017. [Online]. Available: https://www.ti.com/lit/ds/snis159h/snis159h.pdf [Accessed: 10 Jan. 2024].

[13] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in Proc. IEEE Int. Symp. Wearable Comput. (ISWC), Newcastle upon Tyne, U.K., 2012, pp. 108–109.

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[15] R. David et al., "TensorFlow Lite Micro: Embedded machine learning for TinyML systems," in Proc. Mach. Learn. Syst. (MLSys), 2021, pp. 1–17.

[16] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA, 2017, pp. 4765–4774.

[17] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in Proc. Int. Conf. Mach. Learn. (ICML), New York, NY, USA, 2016, pp. 1050–1059.

[18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Salt Lake City, UT, USA, 2018, pp. 7132–7141.

[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), Munich, Germany, 2018, pp. 3–19.