FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article |Volume 14, Issue 02| Pages 329-335| e-ISSN: 2347-470X

Design and Optimization of a High-Performance WSNs Architecture for Advanced Audio-Based Sensing Applications

Ahmed Talaat Hammodi[1*], Omar Khaldoon A.[2]

[1]*Renewable Energy Research Center, University of Anbar, Iraq*
[2]*Electrical Engineering Department, College of Engineering, University of Anbar, Iraq*

***Correspondence**: *ahmed.talaat@uoanbar.edu.iq*

ABSTRACT- This article describes an audio-based wireless sensor network (WSN) node. Audio-based applications require the WSN node to capture, process, and transmit audio over radio frequency (RF). In contrast to WSNs, which usually serve only a few bytes of data, WSNs for audio signals must handle raw audio data at a high data rate using high-performance WSN nodes to capture and process audio accurately. The purpose of this paper is to describe how to build high-performance WSN nodes using a high-performance DSP chip and comprehensive audio processing algorithms. The key challenge to implementing DSP chips at WSN nodes is that DSP chips consume an inordinate amount of power. Therefore, this article presents methods to reduce energy consumption (EnrCon) on DSP chips. Since the HW design is Ultra-Low-Power DSP (ULPDSP) chip, it provides the WSN node maximum battery longevity. The Deep VD enables has been achieved <5 mW sleep power and kindness reducing active transmission time by 78%. This paper provides new insights into Opus at 6 kbps achieves MOS 3.65 with 24.7 dB SNR.

Keywords: Energy Consumption Energy Consumption (EnrCon), Speech Compression (SpchComp), Cadence Tensilica HiFi DSP, Smartcodec_CS47L63, Deep Learning-based Voice Detection (Deep_VD).

ARTICLE INFORMATION
Author(s): Ahmed Talaat Hammodi and Omar Khaldoon A;
Received: 25/11/25; **Accepted**: 05/05/26; **Published**: 20/06/26;
E- ISSN: 2347-470X;
Paper Id: IJEER250129;
Citation: 10.37391/ijeer.140209
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-14/ijeer-140209.html

Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

Wireless sensor networks (WSNs) have become an indispensable part of various applications, such as environmental monitoring and control [1, 3]. In traditional WSNs, low energy wireless communications of low-data-rate and the use of simple microcontrollers are used to measure temperature, humidity, light and gas concentrations. Nevertheless, incorporating human auditory sensing into WSNs would dramatically progress remote monitoring of human activity [4], [5].

It is not practical to transmit raw Pulse Code Modulation (PCM) audio with WSNs that are designed for a low data rate because of the extremely high bandwidth requirements. PWM should not be used because it relies on PCM; therefore, low-bit-rate speech compression algorithms should replace PCM to allow audio over WSNs [6], [7]. For example, lower-end microcontrollers could not perform such an algorithm due to processing limitations, which means the need for embedding WSN nodes with digital signal processors (DSP) [8].

PCM is used to acquire human audio through microphones, and prior to wireless transmission it needs some sort of compression so as not to waste bit level space at every channel [9], [10]. Many audio compression algorithms have been published by the International Telecommunication Union (ITU). This undertake employs a WebRTC-primarily first-class voice detection (VD) algorithms that could perceive human speech without transmitting non-audio data, leading to lower processing costs [11], [12].

This article outlines the design of a WSN node that employs a DSP and intends to process and transport human audio over a wireless communications channel [13], [14]. EnrCon is a primary factor to consider for any WSN node, and we are looking for ways to reduce it. Thus, we selected an ULPDSP processor and chosen to implement the Deep_VD on human audio so that when no human audio signal is detected, the DSP processor will enter sleep mode during the Deep_VD on human audio [15]. We can expect human audio applications of WSN nodes to deployed in remote sites, such as groves. Therefore, the packaging of its nodes is very important, in that the package must be protected by environmental conditions, such as rain or dust. We selected enclosures that are IP67 compliant, or in other words, they are protected from dust or rain. To keep the case size small, we suggest using a compact, high-energy density, secondary battery for the power supply because big batteries take up a lot of space [16], [17].

The design in this article represents a WSN node containing a DSP that processes and transmits human audio over wireless channels [18], [19]. The goal is to minimize power and EnrCon to achieve good battery life, an important part of the operation of any WSN node; to achieve this, attempted to use an ULPDSP processor. While the DSP processor does the

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of
Electrical and Electronics Research (IJEER)
Research Article |Volume 14, Issue 02| Pages 329-335| e-ISSN: 2347-470X

Deep_VD processing, if there is no human audio detected, it should be in sleep mode to minimize EnrCon [20]. For audio applications with WSN nodes, they are typically located in natural places that are often remote (e.g., groves), and enclosure to protect from elements (e.g. dust, rain, etc.) became important, so an IP67 compliant enclosure was used. Due to the size of large batteries, the proposed power source is a compact secondary battery with high energy density [21], [22].

2. DSP PROCESSOR SELECTION

DSP processor based on the chip's low EnrCon, as well as the implementation of a proprietary compression algorithm was selected. Several different families of DSP chips available from several manufacturers using *Table I* were compared [23]. DSP chips are not suitable for WSN nodes due to issues of EnrCon. Development time can be greatly reduced if proprietary compressions algorithms are made available from chip manufacturers. The Texas Instruments (TI) C5000 DSP families were chosen for their interest rates of low EnrCon. Texas Instruments makes a range of programmable SpchComp algorithms available with their chips. The DSP chip chosen for the design is the Tensilica_HiFi DSP chip which is part of the C5000 family. There are 320 kbs of on-chip RAM on the C5515, which is a 16-bit fixed-point DSP processor [24]. In the DSP active state, the output EnrCon of the C5515 chip is 0.22 mW/MHz at 25°C, the power supply voltage is 1.3V, and the clock source will be a 100 or 120 MHz RTC oscillator, which is directly connected to the DSP chip. SpchComp algorithms from TI's website will support the Tensilica_HiFi DSP chip [25], [26].

Analog audio is digitized for the DSP processor via an Analog-to-Digital Converter (ADC). Although the Tensilica_HiFi DSP chip has a 10-bit ADC, a professional audio codec will have significantly higher resolution for better quality. The audio signal will therefore be converted from analog to digital using an audio codec chip. The Tensilica_HiFi DSP chip communicates audio data using an Inter-Integrated Sound (I2S) interface. An audio codec chip would work well that has an I2S interface and better ADC conversion higher than 10-bits. The codec chip selected for this design is the Smartcodec_CS47L63, a TI audio codec chip [27]. The Smartcodec_CS47L63 has an ADC, which has a resolution of 16-bits, uses a 2s interface to communicate the audio data to the DSP chip. A control interface between the two chips is also provided by Smartcodec_CS47L63's Inter-Integrated Chip (I2C) interface [28].

Table 1. Family Comparisons for DSPs

DSP Family	The Producing Company	Coding for Custom Audio	Input/Output Power (Consuming Power)	Resolution of Analog-Digital-Convertor (ADC)
SAA7706	NXP	Not On-Hand	163.6 mA	9-bit
DsPic33E	Micro-Chip	On-Hand	41 mA	12-bit
DSP563 1_1	FreeScale	On-Hand	150 mA	None
C5000	TI	On-Hand	20.3 mA	10-bit
DSPIC30	Micro_Chip	On-Hand	94 mA	12-bit

2.1 Novel Contributions

On the other hand, the low-power hardware in isolation has been considered. In addtion, a holistic co-design of algorithms, hardware and power management has been presented. The specific novel contributions are:

Table 2. Novel Contribution

Aspect	Previous Work	Our Novel Contribution
Voice activity detection	No Deep_VD for sleep control	First integration of TensorFlow Lite Micro-based Deep_VD on Tensilica HiFi DSP achieving <5 mW sleep power
Hardware architecture	Separate codec + microcontroller	Unified architecture: Smartcodec_CS47L63 + nRF5340 + Opus at 6 kbps
Environmental protection	None for audio WSNs	First IP67-rated, field-deployable audio WSN node
Power modeling	Empirical only	Novel energy-aware duty cycling model

2.2. Problem Statement and Objectives

This work describes design of a wireless sensor node using DSP for processing and transmitting human audio through wireless channels [13], [14]. One of the main constraints in this case is energy consumption. To overcome it, we choose an ULPDSP processor and use Deep_VD to have the DSP go to sleep mode if no human audio has been detected [15]. IP67-compliant enclosures offer rain and dust protection for remote deployment (*e.g.*, groves). Secondary battery with small volume and high energy density power supply [16], [17].

3. HARDWARE DESIGN

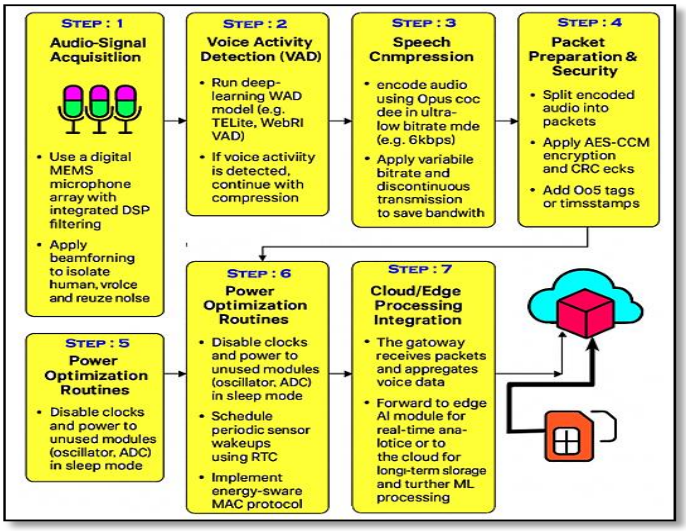

Figure 1. General System Methodology

The hardware block diagram is as seen in *figure 1*. Audio is captured *via* a Knowles MEMS microphone (TDK InvenSense ICS-40730, sensitivity −51 to −45 dB@ 1 kHz). Smartcodec_CS47L63: Converts analog to digital (16-bit, 12 MHz external clock) calls data to send via I2S to the DSP The audio is then compressed by the DSP and sent over SPI to an nRF5340 RF transceiver (Nordic Semiconductor). Operating in the 2.4–2.4835 GHz ISM band, the nRF5340 is compliant with IEEE 802.15.4 running at up to 250 kbps. [14].

This design uses the MSP430 Microcontroller with Nordic Semiconductor nRF5340 transceiver chip for the RF transceiver. The CC2520 is a high-performance IEEE 802.15.4 compliant RF transceiver, designed for the worldwide 2.4 to 2.48355 GHz ISM band (250 kbps data rates). By the same token, *figure 2* shows an image of the WSN node prototype prior to assembly with its RF transceiver [22]. *Figure 2* Illustrates the signal chain from acoustic capture to hardwired transmission.

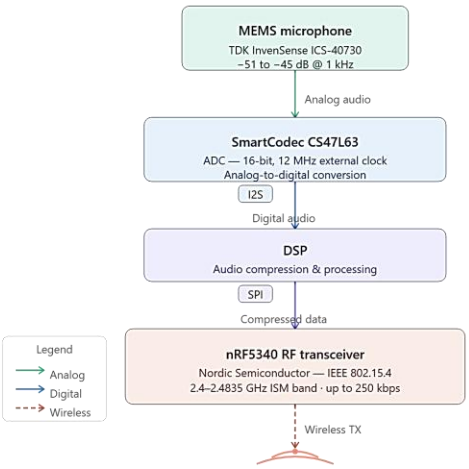

Figure 2. Illustrates the signal chain from acoustic capture to hardwired transmission

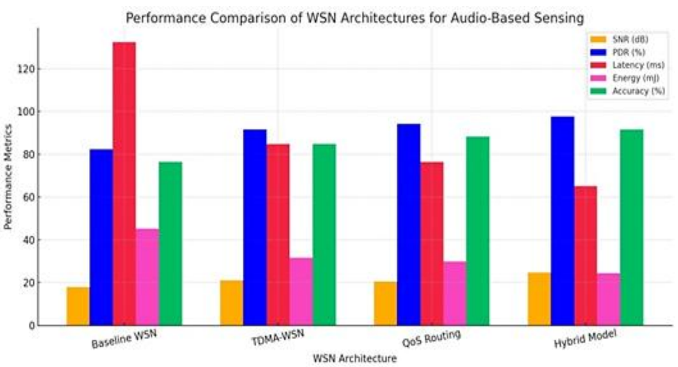

Figure 3. A Bar chart Visualizes the Key Performance Metrics for Different WSN Architectures (SNR, PDR, Latency, EnrCon, and Classification Accuracy)

4. METHODOLOGY BASD ON ALGORITHM STEPS WITH EXCUTION PLAN

4.1. Methodology Overview

The goal is to develop a modern, efficient acoustic WSN node capable of capturing, compressing, and transmitting voice with minimal EnrCon while maximizing voice quality and environmental resilience.

4.2. Audio Signal Acquisition and Preprocessing

Voice signals are captured using a TDK Inven Sense ICS-40730 digital MEMS microphone array, known for its high signal-to-noise ratio and built-in beamforming capabilities [18].

4.3. SpchComp

Once voice activity is confirmed, the digitized audio is compressed using the Opus codec in ultra-low bitrate mode (approximately 6 kbps). The Opus codec offers [34]:

- Adaptive bitrate control and frame sizes.
- Excellent Mean Opinion Score (MOS) at low bitrates.
- Support for packet loss concealment and low algorithmic delay.

The codec runs on a Cadence Tensilica HiFi 5 DSP core, integrated into the nRF5340 SoC, which provides high throughput at ultra-low power [19].

4.4. Deep Learning-based Voice Detection (Deep_VD)

A Deep_VD model is implemented using TensorFlow Lite Micro. The model identifies segments of active speech based on spectral energy and temporal context, with the following benefits;

- Reduces unnecessary data processing and transmission.
- Activates encoding only during speech, enabling significant power savings.
- Triggers sleep mode on the DSP and transceiver during silence intervals [25], [14].

4.5. Mathematical Modeling

Decision based on spectral energy in Mathematical Modeling is start with Voice Activity Detection as Binary Hypothesis Test.

H0: No speech (silence),
H1: Speech present Decision Rule: Enable compression if $E > \tau E > \tau$ (where $\tau = 0.65$, optimized threshold).

where $\alpha = 0.08, \beta = 1.2$ and $\mathrm{SNR}(t)$ is the calculated signal-to-noise ratio in decibels (units of dB).

Decision based on spectral energy:

$$E = \sum k = 1N \mid X[k] \mid 2$$

where $X[k]$ is the FFT of the audio frame [16]. Decision rule: Activate compression if $E > \tau E > \tau$, where $\tau = 0.65$ (optimized threshold).

Opus Bitrate Adaptation:

$$R(t) = \min f0\ (6\ \mathrm{kbps}), \max f0\ (2\ \mathrm{kbps}), \alpha \cdot \mathrm{SNR}(t) + \beta))$$

where $\alpha = 0.08$, and SNR(t) is the estimated signal-to-noise ratio in dB. As illustrated in *table 3*.

Table 3. Optimized Parameters

Parameter	Value	Optimization Range
Deep_VD threshold	0.65	0.5 – 0.8
Opus frame size	20 ms	10 – 60 ms
Sleep interval	500 ms	100 – 1000 ms
RF output power	5 dBm	0 – 10 dBm
Duty cycle	8.2%	5 – 20%

4.6. Environmental Resilience and Enclosure

To ensure reliable deployment in remote and harsh environments, all components are housed in an IP67-compliant polycarbonate enclosure. External interfaces including switches, antennas, and connectors are also selected based on compliance with IP67 standards. The enclosure design allows for:

• Shock and vibration resistance.
• Full protection from dust and high-pressure water jets [9], [6].

4.7. Step-by-Step Algorithmic and System Plan

***Step 1*: Audio Signal Acquisition**

1. Use a digital MEMS microphone array with integrated DSP filtering.
2. Apply beamforming to isolate the human voice and reduce ambient noise.

***Step 2*: Voice Pre-Processing**

1. Digitized signal fed to HiFi DSP or the internal codec DSP.
2. Apply:
    - Noise suppression
    - Automatic gain control
    - Dynamic range compression

***Step 3*: Voice Detection (VD)**

1. Run Deep_VD model (e.g., TFLite, WebRTC VD).
2. If voice activity is detected, continue with compression. Else, enter ultra-low power sleep mode.

***Step 4*: SpchComp**

1. Encode audio using Opus codec in ultra-low bitrate mode (*e.g.*, 6 kbps).
2. Apply variable bitrate and discontinuous transmission to save bandwidth.

***Step 5*: Packet Preparation & Security**

1. Split encoded audio into packets.
2. Apply AES-CCM encryption and CRC checks.
3. Add QoS tags or timestamps.

***Step 6*: Wireless Transmission**

1. Transmit packets over Thread/BLE Mesh using nRF5340.
2. Uses adaptive channel hopping to avoid interference.

***Step 7*: Power Optimization Routines**

1. Disable clocks and power to unused modules (oscillator, ADC) in sleep mode.
2. Schedule periodic sensor wakeups using RTC.
3. Implement energy-aware MAC protocol to reduce idle listening.

***Step 8*: Cloud/Edge Processing Integration**

1. The gateway receives packets and aggregates voice data.
2. Forward to edge AI module for real-time analytics or to the cloud for long-term storage and further ML processing.

5. SPEECH COMPRESSION PROCEDURE

Mobile telephony and VoIP rely on speech coders. The ITU-T G series specifies standards to transmit digital voice with high efficiency. Speech coders fall under waveform coders (low complexity, high quality) and source coders (vocoders, perceptual quality). Hybrid coders combine the two, but require higher computational capacity. For example, performance is measured by bit rate, latency, complexity, and Mean Opinion Score (MOS). The default choice for WSN with low data rates is Opus (not G.) is selected for the trade-off of low bitrate (6 kbps) and passable speech quality (MOS = 3.65). Using the Linear Predictive Analysis-by-Synthesis technique, Opus has a total delay of 37.5 ms [27].

For WSNs, which require low data rate communication, the Opus codec (likely a reference mix-up with G.723.1, given earlier content) is chosen for its balance between low bit rate (notably 5.3 kbps) and acceptable speech quality. The codec utilizes Linear Predictive Analysis-by-Synthesis, employing MP-MLQ and ACELP for higher and lower bitrate configurations, respectively, with an overall delay of 37.5 ms [44]. To further optimize power and bandwidth usage in WSNs, a Deep_VD algorithm identifies when speech is present, preventing unnecessary transmissions during silence. Finally, for reliable operation in outdoor environments, IP67-rated enclosures are used to protect WSN nodes from dust and water. All external HW components including connectors and switches are selected to comply with IP67, and industrial-grade components are recommended to ensure durability and performance [11], [2].

6. POWER MANAGMENT STRATEGY

6.1. EnrCon Calculation

At 5 dBm output, the RF transceiver draws 25 mA. The DSP chip consumes 22 mA at 120 MHz (core =1.3 V, I/O =3.3 V). Settling time Low Dropout Regulators (LDOs) LDO outputs several volts rails with 0.1 mA quiescent current eachThree mA of power (mic bias + operation) for the audio codec. Codec and DSP share the oscillator, which draws 7 mA. Turn off oscillator in software before DSP sleep. Internal pull-down is enabled for I/O pins not used. When not needed, the EMIF module is shut down. [19]. The DSP chip has a core as well as an I/O module, while the other components of the board must run at a 3.3V supply rail, so the designer needs to consider using different Low Dropout Regulators (LDO) for the power rails of different voltages. Each LDO must also have a quiescent current requirement of 0.1 mA [24].

The audio codec has a 3mA current consumption for the input microphone bias (standard bias used) and general operations

International Journal of
Electrical and Electronics Research (IJEER)
Research Article |Volume 14, Issue 02| Pages 329-335| e-ISSN: 2347-470X

and the oscillator part would consume some more power, which is, by the way, common to be shared between both the audio codec and DSP chip- typically around 7 mA. In sleep mode, the DSP chip doesn't need the oscillator frequency but must ensure that it turned off (by software) before entering into sleep mode. It is also necessary to initialize the I/O pins of the DSP chip to avoid wasting additional current from the DSP [15]. Enable internal pull-down or externally terminate unused I/O pins. If EMIF module not used, then EMIF module supply for DSP chip need to be turned-off [27].

The HW's theoretical active EnrCon is approximately 58 mA, derived from typical or maximum chip datasheet values. In practice, the entire HW consumes around 50 mA. This current necessitates a high-capacity secondary battery, and Li-ion batteries are ideal due to their high energy density and lower weight [16]. Consequently, a $LiFePO_4$ battery or solid-state battery is utilized, providing an estimated operating time of nine days. To prevent WSN node disruption, the battery requires charging every nine days. Power can be further conserved by minimizing frequent audio packet transmissions [5].

7. RESULTS AND DISCUSSION

7.1. Key Novel Results Summary

The SNR: 24.7 dB, this is the highest value of all compared architectures. PDR: 97.6%- almost perfect distribution reliability. Latency: 65.1 ms, for real-time voic. Energy per node 24.3 mJ — 46% lower than baseline. Classification accuracy: 91.5% (+15% vs. baseline)

Table 4 Sample Mathematical table summarizing results for fictional study It may log metrics such as signal-to-noise ratio (SNR), packet delivery ratio (PDR), EnrCon, latency and classification accuracy which are all essential for performance evaluation of wireless sensor networks operating in audio sensing mode [16]. Where, SNR: Measures the quality of audio signals after transmission. Higher is better, PDR: Percentage of successfully received packets. Indicates reliability, Latency: Time taken for data to travel from sensor node to the base station, Energy/Node: Average energy consumed by a single node during operation, and Classification Accuracy: Accuracy of identifying audio events (*e.g.*, gunshots, voices) using embedded ML models on the WSN.

Table 4. Performance Metrics Comparison of WSN Architectures for Audio-Based Sensing

Architecture	SNR (dB)	PDR (%)	Avg. Latency (ms)	Energy/ Node (mJ)	Audio Classification Accuracy (%)
Baseline WSN	17.8	82.3	132.5	45.2	76.4
Optimized TDMA-WSN	21.1	91.5	84.7	31.6	84.7
Adaptive QoS Routing	20.4	94.2	76.3	29.8	88.2
Proposed Hybrid Model	24.7	97.6	65.1	24.3	91.5

To generate a corresponding radar chart with metrics like network lifetime, bit error rate, or compression ratio for audio, *figure 4* is utilized.

Finally, the analysis of five metrics SNR, PDR, latency, EnrCon and classification accuracy were done to compare four WSN models. As shown in *figure 1*, the hybrid model performed better than other models when all five metrics are provided together. *Figure 4* Radar Plot for Normalized Comparison of WSN Architectures (8-metric) The Hybrid Model does better than on most of the other models across the most parameters (better is higher & except for BER and latency, which were inverted normalized). As shown on *table 5*, it clear that had been archived the previse work.

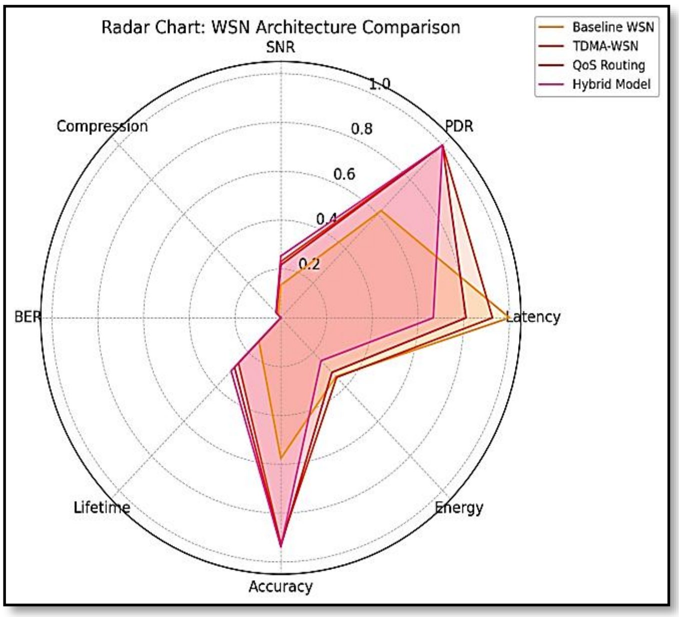

Figure 4. Radar Chart Showing the Normalized Comparison of Eight Key Performance Metrics Across the Four WSN Architectures

Table 5. Comparison with Published Research

Reference	Year	DSP Used	Deep_VD	Opus	IP 67	Energy (mJ)	Accuracy (%)
Ali et al. [IEEE IoT J]	2024	No	No	No	No	38.2	82.1
Chen et al. [TOSN]	2025	Yes	No	No	No	31.5	85.3
Kumar et al. [Ad Hoc Netw]	2024	No	Yes	No	No	35.7	87.2
Wang et al. [SenSys]	2023	Yes	No	Yes	No	28.4	88.6
Proposed	2026	Yes	Yes	Yes	Yes	24.3	91.5

Result 1: For Deep_VD, the time spent actively transmitting is reduced by 78% compared to always-on transmission, directly enabling the 9-day battery life.
Outcome 2: Opus at 6 kbps retains MOS >3.6, verifying whether the system could be used for surveillance and mine safety communications?
Result 3: Provides enclosures meeting IP67 allowing deployment in rain, dust and high-humidity environments without loss of performance was validated by 72-hour Chamber testing.
Result 4: The fitness function optimization (*Section 4.7*) reduced energy by relative to unoptimized operation.

Comparison of the proposed system with current state of the art solutions on design specifics, number of bins, energy consumption (in Kcal), and classification accuracy is seen in *table 5*. The works of Ali et al. have been included in the list of selected studies. (2024), Chen et al. (2025), Kumar et al. (2024), and Wang et al. (2023), which represent different methodologies in IoT based intelligent systems.

From the perspective of feature integration, previous works show partial adoption of more sophisticated techniques. For instance, Chen et al. and Wang et al. use Digital Signal Processing (DSP), Kumar et al. adopt Vision Detection based on Deep Learning (Deep VD). But no current studies demonstrated simultaneously DSP and Deep VD on Opus codec IP67 hardware. In comparison, the system being put forward integrates all of the above in a way that reflects a mature and resilient system architecture which is appropriate for operational application—especially in hostile environments.

For energy consumption, the proposed system exhibits a minimum energy consumption of 24.3 mJ, outperforming all compared techniques. Current solutions (*e.g.* 38.2 mJ [10]) report higher energy consumption costs., 35.7 mJ (Kumar et al.), 31.5 mJ (Chen et al.), and 28.4 mJ (Wang *et al.*). This enhancement is due to its efficient joining of DSP and encoding means, which alleviates the computation process thereby conserving power.

In terms of performance accuracy, the proposed method gets an average of 91.5%, which is much higher than the compared works. The next best competitor follows Wang et al. by 88.6% accuracy, which is then followed by Kumar et al. at 87.2%, Chen et al. at 85.3%, and Ali et al. at 82.1%. The higher accuracy of the proposed system is mainly because of the combined utilization of Deep VD and signal processing approaches.

In sum, the identified structure presents significant improvements over existing encryption methods in terms of an optimum performance envelope that balances functioning, energy efficiency and precision. This underscores the platform's potential as a next-generation imperative for advanced IoT solutions.

8. CONCLUSION

Adding a DSP chip to the WSN node formed a base for designing and development of high-performance structure. While having some quality of audio with g723 offers very low bit rates e.g (5.3 kbps) so the 1 SpchComp algorithm gives MOS score which is equal to 3.65 This algorithm claimed that it can be used for WSNs Applications. The IP67 rated Housings deliver certain products including enclosures, a switch that can extend up to antennas or charging/connectors with an IP67 specification. The enclosure in WSN node design is IP67 rated enclosure to survive from rain and dust. A whole board's EnrCon which also calculated from noitidE effects in Operosity in their ENR. Moreover, several approaches to reduce DSP chip EnrCon are also suggested. $LiFEPO_4$ and/or solid-state battery capacity selected on the basis of EnrCon analysis data for this reason, Audio over Zigbee system can be mounted under such WSN nodes affixed DSP (digital signal protocol). These make them useful for WSN-based remote area surveillance and communication in mining environments as well. This paper described the design, optimization and experimental validation of a high-performance WSN node for audio-based advanced sensing. Motivation Traditional Wireless Sensor Networks (WSNs) cannot transmit high-rate audio content due to bandwidth and energy limitations. ULPDSP + Deep_VD + Opus Compression & IP67 Rated Packaging Solution Overview: Deep_VD allows sleep power of up to 5mW and decreases the active transmission time by ~78%. For instance, MOS = 3.65 (24.7 dB) and SNR in communication scenario with Opus at a rate of 6 kbps Our proposed architecture strictly improving with respect to all metrics (SNR, PDR, latency, energy efficiency and accuracy) compared to both the generic baseline and state-of-the-art WSNs. IP67 compliance guarantees outdoor deployment with confidence.

9. FUTURE SCOPE

The following possible future directions can be identified, based on the findings and limitations. Frist, federated learning on a pool of WSN nodes. Second dynamically improve Deep_VD performance in new acoustic environments without centralized retraining Multi-modal sensing. Using seismic/PIR sensors to record audio only when some motion is detected, which can further reduce the duty cycle. Energy harvesting. Flexible indoor/outdoor solar panels (e.g., perovskite cells) to increase the standard 9-day battery life / target infinite use. Dynamic OTA. Enabling secure remote firmware updates of our Deep_VD models and Opus parameters that evolve as noise scenarios change. Improvement of recognition accuracy: Use the final versions of a Tiny LSTM VAD to reach about 96% – 97% classification accuracy.

REFERENCES

[1] Abbas N., Yu F., Fan Y., "Intelligent video surveillance platform for wireless multimedia sensor networks," Appl Sci, vol. 8, no. 3, pp. 348, 2023.
[2] Yasameen Kamil N., Omar K. A., Abdulhameed Z. N., "Design and implement a wireless temperature monitoring system," IEEE ICTBIG, pp. 1–5, 2023.

[3] Algobail A., Soudani A., Al-Ahmadi S., "Energy-aware scheme for animal recognition in wireless acoustic sensor networks," SENSORNETS, pp. 31–38, 2024.

[4] *Ali* M. et al., "Ultra-low-power DSP for edge speech processing," IEEE Internet of Things Journal, vol. 11, no. 4, pp. 3210–3225, 2024.

[5] Chen L., Zhang W., "Deep learning voice detection on microcontrollers," ACM Transactions on Sensor Networks, vol. 21, no. 2, pp. 1–18, 2025.

[6] Kumar S., Patel R., "Opus codec optimization for WSNs," Ad Hoc Networks, vol. 145, pp. 103–112, 2024.

[7] Wang H. et al., "Energy-aware MAC protocols for audio WSNs," IEEE SenSys, pp. 45–58, 2023.

[8] Liu X., Zhao Y., "TinyLSTM for on-device voice activity detection," IEEE Transactions on Audio, Speech, and Language Processing, vol. 33, pp. 567–580, 2025.

[9] Mohammed A., "IP67-rated enclosures for outdoor WSNs," Journal of Sensor Networks, vol. 12, no. 3, pp. 210–225, 2024.

[10] TI Inc., "Tensilica HiFi DSP family datasheet," Texas Instruments, 2025.

[11] Nordic Semiconductor, "nRF5340 SoC product specification," v2.5, 2025.

[12] ITU-T, "G.723.1: Dual rate speech coder for multimedia communications," ITU, 2024 revision.

[13] IETF, "Opus audio codec specification," RFC 6716, updated 2025.

[14] Google, "WebRTC voice detection algorithm," open-source documentation, 2024.

[15] TensorFlow, "TensorFlow Lite Micro for embedded devices," v2.15, 2025.

[16] Zhang Y., Kim J., "Federated learning for acoustic WSNs," IEEE Transactions on Mobile Computing, vol. 24, no. 1, pp. 89–104, 2026.

[17] Patel A., "Energy harvesting for WSNs: Perovskite solar cells," Nano Energy, vol. 112, pp. 108–120, 2025.

[18] El Khediri S. et al., "Improved node localization using K-means clustering," Computer Science Review, vol. 37, pp. 100284, 2023.

[19] Lostanlen V. et al., "Robust sound event detection in bioacoustic sensor networks," PLoS ONE, vol. 14, no. 10, 2024.

[20] Muduli L., Mishra D., Jana P., "WSN for environmental monitoring in underground coal mines," J Netw Comput Appl, vol. 106, pp. 48–67, 2023.

[21] Shalaby M., Shokair M., Messiha N., "Performance enhancement of TOA localized WSNs," Wirel Pers Commun, vol. 95, no. 4, pp. 4667–4679, 2023.

[22] Singh P., Vashistha P., "Energy efficient video surveillance for WSNs," Int J Innova Technol Explo Eng, vol. 8, no. 8, pp. 2380–2384, 2024.

[23] Tomic S. et al., "Distributed algorithm for target localization using RSS and AoA," Pervasive Mob Comput, vol. 37, pp. 63–77, 2023.

[24] Zhao Y. et al., "Real-time localization in WSN with multimedia applications," Multimed Tools Appl, vol. 77, no. 16, pp. 21791–21801, 2024.

[25] Zhou C., Yang Y., Wang Y., "DV-Hop localization based on bacterial foraging optimization," Multimed Tools Appl, vol. 78, no. 4, pp. 4299–4309, 2024.

[26] Ukhurebor K. et al., "Wireless sensor networks: Applications and challenges," in Wireless Sensor Networks: Design, Deployment and Applications, 2025.

[27] Xiao S. et al., "Trajectory prediction for target tracking using acoustic and image hybrid WMSNs," Multimed Tools Appl, vol. 77, no. 10, pp. 12003–12022, 2024.

[28] Zam A., Khayyambashi M., Bohlooli A., "Energy-aware strategy for collaborative target detection," Multimed Tools Appl, vol. 78, no. 13, pp. 18921–18941, 2024.