

Domain-Aware Gate Selection for Acoustic Sensor Networks in Internet-of-Sounds Leak Monitoring

Ayodeji Ologunmeta*, Raghunandan Kalibhat†, Yafan Zhao*, Somasundar Kannan*, Lina Mohjazi†, Nazila Fough†

*Robert Gordon University, Aberdeen, United Kingdom

†University of Glasgow, Glasgow, United Kingdom

Abstract—Domain-aware gate selection is necessary for acoustic sensor networks in the Internet of Sounds (IoS) because the most suitable first-stage routing strategy can vary across acoustic domains. This paper presents an edge-oriented adaptive acoustic inference framework for water-pipeline leak monitoring, in which sensor nodes classify leak/no-leak events locally and activate a stronger CNN only when the first-stage decision is insufficient. Two routing strategies are evaluated: a learned tiny-CNN confidence gate and a supervised non-neural spectral entropy-flux gate trained on labelled spectrogram summaries. The scope is local water-pipeline acoustic monitoring rather than general sound classification or a completed offshore underwater-acoustics deployment. The framework is evaluated on the controlled Aghashahi hydrophone dataset and the HK buried noise-logger dataset using source-file grouped splits, five random seeds, calibration diagnostics, model parameter and MAC counts, strong-model activation rate, measured software-level CPU-latency proxy, paired uncertainty analyses, and zero-shot domain-shift evaluation. On the controlled hydrophone dataset, the learned cascade at $\tau = 0.70$ closely matches the always-on strong CNN descriptively, achieving $F1 = 0.9424 \pm 0.0102$ with 15.95% strong-model activation, 69.3% MAC saving, and 38.6% measured CPU-latency proxy saving. On the buried dataset, learned confidence is less effective, whereas the spectral entropy-flux cascade at $\tau = 0.75$ achieves $F1 = 0.7892 \pm 0.0360$ with 47.49% activation and 27.0% measured latency saving. Paired uncertainty analyses detect no statistically significant F1 difference between this cascade and the always-on strong CNN. Zero-shot transfer from the controlled to the buried domain remains weak, demonstrating the need for target-domain validation before deployment. The findings support domain-aware gate selection for resource-conscious acoustic sensing and embedded IoS inference.

Index Terms—Internet of Sounds, acoustic sensor networks, water pipeline leak detection, edge acoustic inference, adaptive inference, confidence calibration, spectral entropy-flux gate.

I. INTRODUCTION

The Internet of Sounds (IoS) is emerging as a practical extension of the Internet of Things in which sound-capable devices are used to sense, acquire, process, exchange, and act on sound-related information [1], [2], [3], [4], [5]. In this setting, acoustic sensors are not only passive recording devices, but can also support local decision-making at the edge. This is important for monitoring applications where continuous sensing is required, but communication bandwidth, power availability, and onboard computation remain limited. Smart audio sensing at the edge is therefore relevant for anomaly detection and condition monitoring because part of

the acoustic analysis can be performed close to the source rather than relying only on continuous raw-data transmission [6]. Water pipeline leak monitoring is a useful application area for this type of sound-aware sensing. Leaks can generate acoustic signatures that propagate through pipe walls, water columns, or the surrounding medium, making acoustic sensing one of the most widely studied approaches for leak detection and localization in pressurized pipeline systems [7]. Existing studies have investigated leak monitoring using hydrophones, accelerometers, pressure sensors, noise loggers, signal-processing methods, classical machine learning, and deep learning [7], [8], [9], [10], [11], [12], [13], [14], [15]. Public datasets also support reproducible evaluation. The Aghashahi benchmark provides controlled leak and no-leak measurements using hydrophones, accelerometers, and dynamic pressure sensors [8], [9], while the Hong Kong buried water-distribution dataset provides field-style acoustic data collected from real in-service and buried water networks [10], [11]. Recent acoustic leak detection research has mainly focused on classification accuracy, localization, data augmentation, interpretability, and model design. Generative modelling has been used to improve acoustic leak detection under data-limited conditions [12], data-centric studies have examined how framed acoustic data and augmentation affect deep learning performance [13], interpretable deep learning has been explored for understanding leak-related acoustic cues [14], and transformer-based models have been proposed for acoustic leak detection in water distribution networks [15]. Underwater and marine acoustic leakage studies further show the broader relevance of acoustic monitoring for leak, bubble, and gas-release events [16]–[18]. These studies confirm the importance of acoustic leak monitoring, but most treat inference as a direct classification or localization task where one selected model is applied uniformly to all samples. Uniform inference may not be ideal for IoS and embedded monitoring systems. A small acoustic model or signal-domain gate may be sufficient for many easy no-leak or clear leak segments, while uncertain or acoustically difficult segments may require a stronger model. This idea is related to adaptive inference and cascade learning, where early-exit or conditional-computation methods reduce unnecessary processing by allowing easy inputs to be handled at lower computational cost [19]–[23]. However, confidence-triggered and signal-feature-based acoustic routing have not been sufficiently compared for water pipeline leak

monitoring under controlled and buried acoustic conditions, especially with source-file grouped splitting, calibration reporting, strong-model activation rate, latency proxy, and seed-level uncertainty. This paper addresses that gap as a domain-aware gate-selection study for IoS-oriented water pipeline acoustic monitoring. It does not claim a new generic cascade algorithm. Instead, it asks which first-stage routing strategy is justified under which acoustic condition when the same strong-CNN refinement stage is available. Two gate types are evaluated: a learned tiny-CNN confidence gate and a supervised spectral entropy-flux gate trained on labelled spectrogram summaries. The main contributions are as follows: (i) acoustic leak routing is formulated as a gate-selection problem for edge acoustic sensing; (ii) controlled hydrophone and buried noise-logger domains are evaluated using source-file grouped splits with zero source-file overlap between partitions; (iii) learned confidence routing and supervised entropy-flux routing are compared using F1-score, activation rate, calibration, MACs, measured software-level CPU latency, and paired seed-level statistics; and (iv) a practical validation protocol is provided for choosing learned routing, spectral routing, or direct strong inference from target-domain evidence.

II. RELATED WORK

A. IoS Acoustic Sensor Networks and Leak Monitoring

IoS acoustic sensor networks extend sound processing beyond entertainment and musical applications by allowing acoustic events to be sensed, classified, and routed through networked devices [1]- [6]. For infrastructure monitoring, acoustic leak detection has been studied using signal processing, classical machine learning, and deep learning. Hu et al. [7] review acoustic leak localization in pressurized pipelines and show that leak-generated sound and vibration can propagate through pipes and surrounding media. The Aghashahi benchmark [8], [9] supports controlled laboratory evaluation, while the HK dataset [10], [11] represents field-style buried water-network recordings. These datasets are useful because they provide contrasting acoustic domains: a controlled hydrophone testbed and a buried noise-logger environment. These studies motivate the edge-acoustic inference question addressed here: given a local water-pipeline acoustic sensor, when is a low-cost gate sufficient and when should a stronger classifier be activated under controlled and buried acoustic domains?

B. Adaptive Inference and Calibration

Early-exit networks, conditional computation, and cascaded inference have been widely studied [19]- [23]. BranchyNet and related early-exit systems show that high-confidence samples need not always pass through the full model [19]. Conditional deep learning and learned cascades similarly reduce unnecessary computation [20]- [22]. For cascades, confidence calibration is central because overconfident first-stage models may suppress escalation even when they are wrong [21]. This paper therefore reports Expected Calibration Error (ECE) and a temperature-scaling diagnostic [25] instead of treating softmax confidence as automatically reliable.

C. Spectral Entropy and Spectral Flux

Spectral entropy and spectral flux are used here as acoustic descriptors for a supervised first-stage gate. Spectral entropy follows Shannon entropy applied to normalized spectral power [26] and has also been used as a spectral audio feature in robust speech recognition [28]. It summarizes whether acoustic energy is concentrated in a few frequency bins or spread across the spectrum. Spectral flux measures frame-to-frame spectral change and is commonly used as a spectral novelty descriptor [27]. In leak monitoring, these descriptors are meaningful because leaks can introduce turbulent, broadband, and time-varying acoustic behaviour. The paper does not claim that entropy or flux is a universal leak detector; it tests whether labelled entropy-flux summaries can provide a useful routing signal in a target acoustic domain.

III. METHODOLOGY

A. Acoustic Segments and Learned Confidence Cascade

Each recording is segmented into two-second windows and converted to a normalized log-magnitude spectrogram of shape $1 \times 129 \times 124$. The learned gate uses a tiny CNN to produce a leak/no-leak probability vector. If the maximum probability is at least threshold τ , the tiny prediction is accepted. Otherwise, the sample is escalated to a stronger CNN:

$$\hat{y}_i = \begin{cases} \arg \max_c p_t(c | x_i), & \text{if } \max_c p_t(c | x_i) \geq \tau, \\ \arg \max_c p_s(c | x_i), & \text{otherwise.} \end{cases} \quad (1)$$

Here, x_i is a spectrogram segment, $c \in \{0, 1\}$ denotes the binary class index, where $c = 1$ represents leak and $c = 0$ represents no leak. p_t is the tiny-CNN probability vector, and p_s is the strong-CNN probability vector. The threshold τ is swept from 0.50 to 0.90.

B. Supervised Spectral Entropy-Flux Gate

The spectral gate replaces the first-stage CNN with hand-crafted acoustic features and an L_2 regularised logistic regression gate. For each spectrogram frame, the positive pseudo-power spectrum is normalized over frequency bins. Entropy summarizes spectral spread and flux summarizes temporal spectral change:

$$H(t) = - \sum_f q(f, t) \log q(f, t) \quad (2)$$

$$q(f, t) = \frac{P(f, t)}{\sum_f P(f, t)} \quad (3)$$

$$SF(t) = \sum_f [q(f, t) - q(f, t-1)]^2 \quad (4)$$

Eight features are used: mean, standard deviation, maximum, and minimum entropy; and mean, standard deviation, maximum, and total flux. These statistics capture central tendency, dispersion, frame-level extremes, and accumulated spectral movement without selecting features on the test set. The logistic gate uses $C = 1.0$, lbfgs solver, balanced class

weights, $max_{iter} = 2000$, and one fitted model per seed on the same training split used for the CNN models. At test time, its probability confidence is thresholded in the same way as the learned gate.

C. Complexity, Latency, and Evaluation Protocol

The tiny CNN has two Conv-BN-ReLU-MaxPool blocks with 3×3 kernels and channels 8 and 16, followed by adaptive average pooling and a two-class linear head. The strong CNN uses the same lightweight convolutional block family with channels 16, 32, and 64. This design deliberately keeps both branches within a constrained IoS-style acoustic-CNN family so that the experiment isolates gate-selection behaviour rather than comparing unrelated large audio backbones. The measured spectral-gate latency includes entropy-flux extraction and logistic prediction. Apple M3 CPU latency is reported only as a reproducible software-level proxy; it is not embedded energy evidence. MACs and activation rates remain the more hardware-independent quantities. No neural architecture search or exhaustive architecture optimisation was performed; the channel widths and depth tiers were fixed before the routing experiments so that the study isolates gate-selection behaviour rather than claiming an optimised tinyML architecture.

$$T_C(\tau) = T_g + r(\tau) T_s \quad (5)$$

$$S_T(\tau) = 100 \left[1 - \frac{T_C(\tau)}{T_s} \right] \quad (6)$$

$$r(\tau) < 1 - \frac{T_g}{T_s} \quad (7)$$

T_g is the first-stage gate latency, T_s is the strong-model latency, and $r(\tau)$ is the strong-model activation rate. Unless otherwise stated, F1-score is reported as binary leak-class F1, with the leak class treated as the positive class. This is stated explicitly because the Aghashahi split is leak-heavy, while the HK buried split is more balanced. All main results use five random seeds. Paired Wilcoxon tests and bootstrap confidence intervals are used for seed-level differences. Calibration is assessed with 10-bin ECE [25]. The operating thresholds highlighted in the abstract and results were selected from validation-set F1-activation behaviour and then evaluated on the held-out test split; they were not chosen post-hoc from test performance. Because the Aghashahi validation set is also leak-heavy, its selected threshold may favour leak-class performance over balanced no-leak/leak behaviour; this is treated as part of the dataset-specific operating-point selection rather than as a universal threshold. Latency values were measured under the same Apple M3 CPU software environment and used only for relative comparison, not as embedded energy evidence.

D. Datasets and Source-File Grouped Splitting

The Aghashahi dataset is used as the controlled water-distribution hydrophone domain. It provides leak and no-leak recordings from a laboratory water-distribution testbed

with hydrophone, accelerometer, and pressure modalities; this paper uses the hydrophone signal because it is closest to an IoS acoustic sensing node. The hydrophone subset contains 120 source files and 9,072 two-second spectrogram segments. The source-file grouped split contains 5,744 training segments from 76 files, 1,475 validation segments from 20 files, and 1,853 test segments from 24 files. The corresponding no-leak/leak counts are 1,171/4,573 for training, 280/1,195 for validation, and 433/1,420 for testing. The HK buried dataset represents a field-style water-distribution noise-logger domain. It contains acoustic measurements collected from buried water-distribution infrastructure, where leak signatures can be mixed with hydraulic, soil-coupled, and ambient noise. The processed subset contains 140 source files and 1,530 labelled acoustic segments. The source-file grouped split contains 1,022 training segments from 98 files, 249 validation segments from 21 files, and 259 test segments from 21 files. The corresponding no-leak/leak counts are 521/501 for training, 140/109 for validation, and 139/120 for testing. Both datasets are treated as binary leak/no-leak acoustic event detection tasks. All train, validation, and test partitions are grouped by original source file so that neighbouring windows from the same raw recording do not appear in different partitions. The overlap check confirmed 0 shared source files for train-validation, train-test, and validation-test comparisons in both datasets. This source-file grouped split is a key validity control because it reduces the risk that segment-level leakage inflates test performance.

TABLE I
SOURCE-FILE GROUPED SPLITS AND BINARY CLASS DISTRIBUTION FOR THE PUBLIC WATER-PIPELINE ACOUSTIC DATASETS

Dataset / Split	Seg.	Files	No-leak	Leak
Agh-Train	5744	76	1171	4573
Agh-Val	1475	20	280	1195
Agh-Test	1853	24	433	1420
HK-Train	1022	98	521	501
HK-Val	249	21	140	109
HK-Test	259	21	139	120

TABLE II
MODEL/GATE COMPLEXITY, MODEL WEIGHT STORAGE, AND MEASURED SOFTWARE-LEVEL CPU LATENCY

Gate/Model	Params	MACs	CPU ms (Agh/HK)	Weights KB (F32/Int8)
Spectral gate	8 feats.	N/A [†]	0.209 / 0.201	N/A
Tiny CNN	1,330	5.72M	0.403 / 0.362	5.2 / 1.3
Strong CNN	23,650	38.87M	0.888 / 0.789	92.4 / 23.1
CNN total	24,980	—	—	97.6 / 24.4

The spectral gate is not a neural MAC-array model. Its entropy-flux extraction uses reductions, logarithms and frame differences, followed by a two-class logistic prediction; therefore it is reported through measured CPU latency rather than CNN-equivalent MACs. Weight memory is estimated from stored parameters only. The CNN weights total row includes the tiny and strong CNN weights only; the spectral logistic coefficients are negligible compared with CNN weights. Runtime

activations, buffers, and deployment framework overhead are excluded, so these values should not be interpreted as complete embedded memory use.

IV. RESULTS

A. Learned Tiny-CNN Confidence Gate

Table III shows that the learned cascade is most useful on the controlled Aghashahi hydrophone dataset. At $\tau = 0.70$, it reaches $F1 = 0.9424 \pm 0.0102$ with only 15.95% strong-model activation, corresponding to 69.3% MAC saving and 38.6% measured CPU-latency proxy saving. This selected operating point closely matches the always-on strong CNN descriptively. On HK buried data, the learned cascade improves over the tiny baseline, but latency saving disappears as activation rises. At $\tau = 0.75$, the learned cascade still improves over tiny-only F1 but activates the strong model for 65.56% of samples, producing negative measured saving because the gate overhead is added to frequent strong-model calls.

TABLE III

LEARNED-GATE FIVE-SEED PERFORMANCE AND LATENCY-SAVING PROXY. NEGATIVE SAVE% MEANS THE CASCADE IS SLOWER THAN ALWAYS USING THE STRONG CNN.

Data	Mode	F1 $\pm$ std	Act. (%)	Save (%)
Agh	Tiny	0.9337 ± 0.0087	0.00	—
Agh	Strong	0.9404 ± 0.0130	100	0.0
Agh	Cas. 0.70	0.9424 ± 0.0102	15.95	38.6
Agh	Cas. 0.75	0.9428 ± 0.0112	23.27	31.3
HK	Tiny	0.5648 ± 0.0782	0.00	—
HK	Strong	0.7697 ± 0.0710	100	0.0
HK	Cas. 0.60	0.6854 ± 0.0447	19.54	34.6
HK	Cas. 0.70	0.7238 ± 0.0564	52.66	1.5
HK	Cas. 0.75	0.7434 ± 0.0544	65.56	-11.4

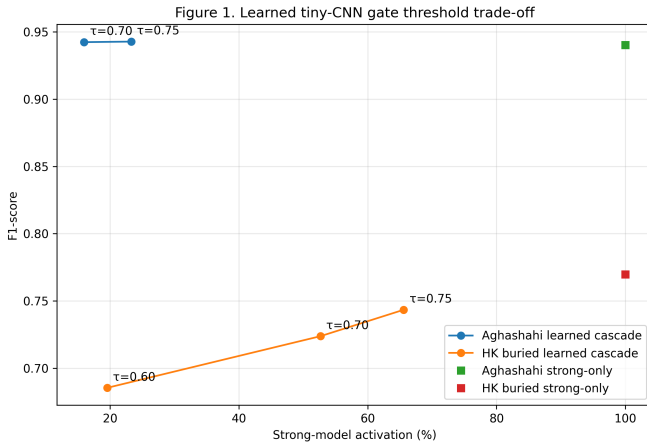


Fig. 1. Learned tiny-CNN gate threshold trade-off. The controlled dataset has an efficient operating region near $\tau = 0.70$, while HK buried data requires high activation and loses the latency-saving advantage.

B. Spectral Entropy-Flux Gate

On Aghashahi, the spectral gate approaches strong-CNN F1 only when activation becomes very high. As seen from Table

IV at $\tau = 0.75$, it reaches $F1 = 0.9297$ compared with the strong model $F1 = 0.9404$, but it activates the strong model for 92.77% of samples and gives -16.3% measured saving. It is therefore not useful as an efficiency-motivated gate on the controlled hydrophone dataset, even though its final F1 is close to the strong model. The result is different on HK buried data. At $\tau = 0.75$, the spectral-gate cascade reaches $F1 = 0.7892 \pm 0.0360$ with 47.49% activation and 27.0% measured latency saving. Paired testing does not establish superiority or equivalence: mean $\Delta F1 = +0.0195$ with bootstrap 95% CI $[-0.0077, 0.0551]$. The spectral-only standard deviation is 0.0000 because the logistic gate is a deterministic function of fixed labelled feature summaries; the five seeds affect CNN training and strong-branch predictions but do not alter the spectral-only decision when the same training features are used.

TABLE IV

SPECTRAL ENTROPY-FLUX GATE RESULTS WITH MEASURED GATE LATENCY. SPEC.-ONLY ROWS USE 0% STRONG-MODEL ACTIVATION; NEGATIVE SAVE% MEANS THE CASCADE IS SLOWER THAN ALWAYS USING THE STRONG CNN

Data	Mode	F1 $\pm$ std	Act. (%)	Save (%)
Agh	Spec. only	0.6352 ± 0.0000	0.00	76.5
Agh	Spec. 0.60	0.7987 ± 0.0075	44.47	32.0
Agh	Spec. 0.75	0.9297 ± 0.0125	92.77	-16.3
Agh	Strong	0.9404 ± 0.0130	100	0.0
HK	Spec. only	0.7407 ± 0.0000	0.00	74.5
HK	Spec. 0.60	0.7571 ± 0.0153	18.53	55.9
HK	Spec. 0.70	0.7642 ± 0.0286	36.29	38.2
HK	Spec. 0.75	0.7892 ± 0.0360	47.49	27.0
HK	Strong	0.7697 ± 0.0710	100	0.0

TABLE V

FROZEN SPECTRAL-FEATURE ABLATION ON HK BURIED DATA AT $\tau = 0.75$. GATE F1 IS THE NON-NEURAL GATE ALONE; CASCADE F1 INCLUDES STRONG-CNN REFINEMENT.

Feature	Gate F1	Casc. F1 $\pm$ std	Act.%
Entropy only	0.6966	0.7712 ± 0.0516	58.30
Flux only	0.7036	0.7798 ± 0.0562	72.59
Entropy + flux	0.7407	0.7892 ± 0.0360	47.49

The ablation retained the source-file grouped training split, balanced $C = 1$ logistic regression, and the frozen $\tau = 0.75$ operating point for all feature sets. On HK buried data, the combined gate gives the highest gate-only and cascade F1 while requiring the lowest activation. Its mean cascade-F1 gain over entropy-only was +0.0180 with bootstrap 95% CI $[-0.0046, 0.0468]$, and over flux-only was +0.0093 with CI $[-0.0071, 0.0364]$. Because both intervals cross zero, these F1 differences are descriptive rather than statistically significant; the lower routing burden is the clearer practical result. On Aghashahi, entropy-only, flux-only, and combined gates reach cascade F1 values of 0.9415, 0.9235, and 0.9297 but activate the strong CNN for 97.19%, 96.55%, and 92.77% of test samples, respectively. Thus, none provides a useful efficiency operating point in the controlled domain, reinforcing selection of the learned gate there. For reproducibility, the standardized

HK combined-gate coefficients for mean, standard deviation, maximum, and minimum entropy were 0.465, 0.196, -1.620, and 0.865; the corresponding mean, standard deviation, maximum, and total flux coefficients were 0.914, -0.944, -0.083, and 0.914.

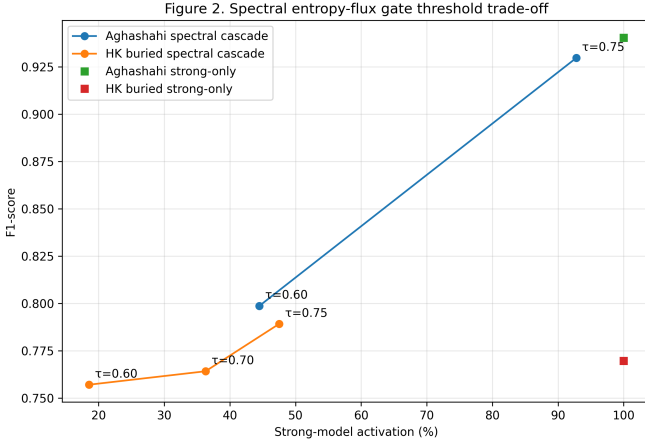


Fig. 2. Spectral entropy-flux gate threshold trade-off. The gate is weak on Aghashahi but provides the best HK operating point at $\tau = 0.75$, where the paired F1 difference from the strong model remains inconclusive while measured latency is lower.

C. Calibration, Statistical Testing, and Domain Shift

Temperature scaling was fitted on validation logits and evaluated per seed on the test split. On Aghashahi, it changed the tiny-model ECE from 0.0575 ± 0.0146 to 0.0666 ± 0.0161 , with unchanged F1. On HK buried data, it worsened ECE from 0.1222 ± 0.0265 to 0.1849 ± 0.0199 , again without changing F1. Temperature scaling is therefore reported as a diagnostic rather than adopted in the final gate-selection results. The calibration and spectral-gate results jointly explain the domain difference: learned CNN confidence is adequate on Aghashahi but less useful on HK buried data, where entropy-flux features provide a more suitable routing signal. The HK tiny raw ECE range across seeds was 0.0860 to 0.1538. To reflect the modest HK test size, a 100,000-resample stratified source-file-blocked bootstrap resampled the 10 leak and 11 no-leak test files separately and averaged F1 over the five frozen seeds. The selected spectral cascade achieved $F1 = 0.7892$ with 95% CI [0.5956, 0.9468]; its paired $\Delta F1$ relative to the strong CNN was $+0.0195$ with 95% CI [-0.0509, 0.0987]. The broad intervals reinforce cautious interpretation.

TABLE VI
CALIBRATION AND PAIRED UNCERTAINTY ANALYSES.

Item	Value
Agh tiny ECE (raw/temp)	$0.0575 \pm 0.0146 / 0.0666 \pm 0.0161$
HK tiny ECE (raw/temp)	$0.1222 \pm 0.0265 / 0.1849 \pm 0.0199$
HK learned 0.75 – Strong	-0.0263 ; CI [-0.0475, -0.0093]
HK spectral 0.75 – Strong	$+0.0195$; CI [-0.0077, 0.0551]
HK spectral 0.75 file boot-strap	0.7892 ; CI [0.5956, 0.9468]; $\Delta = +0.0195$; CI [-0.0509, 0.0987]

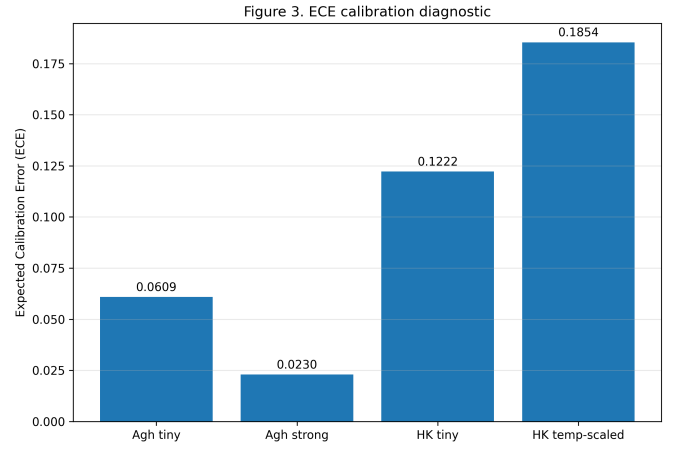


Fig. 3. ECE calibration diagnostic. HK buried confidence calibration is weaker, and validation-fitted temperature scaling worsens mean test ECE rather than improving the routing signal

TABLE VII
ZERO-SHOT CNN DOMAIN-SHIFT RESULTS

Mode	F1	Precision	Recall	Act. (%)
Tiny ext.	0.4292	0.4243	0.4342	0.0
Strong ext.	0.6493	0.5235	0.8548	100
Cas. 0.70	0.4668	0.4336	0.5055	15.29
Cas. 0.90	0.5751	0.4880	0.7000	42.55

Zero-shot transfer from Aghashahi to HK buried represents a cold-start IoS deployment scenario, where a model trained in a controlled hydrophone environment is applied directly to a new buried acoustic site without labelled target-domain calibration. Table VII transfers the trained CNN models only: the Aghashahi-trained tiny and strong CNNs are evaluated directly on HK buried data without retraining. The result remains weak, with the best listed cascade reaching $F1 = 0.5751$ at 42.55% activation. Precision also shows the domain-shift behaviour more clearly: the zero-shot strong model has high recall (0.8548) but modest precision (0.5235), suggesting leak-biased prediction under target-domain mismatch. The spectral logistic gate is not reported as a zero-shot transfer result because its coefficients encode source-domain class-conditional entropy-flux distributions; applying those coefficients directly to HK would mainly reflect distribution mismatch in the handcrafted feature space rather than a validated target-domain routing policy. In practical IoS deployment, a small labelled target-domain validation set is therefore needed before selecting learned routing, spectral routing, or direct strong inference.

V. DISCUSSION

A. Domain-Aware Gate Selection

The dual-gate comparison yields a more specific conclusion than a single-strategy cascade study would allow. Learned confidence routing is the better strategy on controlled hydrophone data because the tiny CNN is already strong and reasonably

calibrated. In contrast, HK buried data is less favourable to learned confidence routing; the tiny CNN is weak, and escalation must be frequent. The spectral entropy-flux gate provides an acoustic-specific alternative that is not universally better but is more suitable for the buried dataset. This interpretation is also supported by the cross-domain metrics: tiny-CNN F1 falls from 0.9337 ± 0.0087 on Aghashahi to 0.5648 ± 0.0782 on HK, mean raw ECE rises from 0.0575 to 0.1222, and strong-model activation at the selected learned-cascade operating points increases from 15.95% to 65.56%. The HK challenge is therefore not merely lower final accuracy; the tiny CNN also provides a weaker and less efficient confidence signal for routing.

The different behaviour is grounded in the acquisition contexts. Aghashahi is a controlled hydrophone testbed, while HK buried data is obtained from field-style buried water-distribution recordings [8]- [11]. Buried noise loggers in pressurised distribution mains can capture leak-generated stress waves together with soil-coupled and hydraulic ambient noise [7], [10]. This can make entropy and flux more separable as routing descriptors in the HK buried setting: no-leak segments may have variable broadband entropy from ambient conditions, while leak segments can introduce stronger frame-to-frame spectral change associated with turbulent flow. The result is an observed dataset-specific finding, not an assumption about all buried leak recordings.

B. Efficiency-Motivated Interpretation

The IoS interpretation is efficiency-motivated rather than a claim of measured embedded energy saving. Apple M3 CPU latency is used as a software-level proxy, while MACs and activation rates remain the more hardware-independent evidence. From $r(\tau) < 1 - T_g/T_s$, the learned-gate break-even activation is $1 - 0.403/0.888 = 54.6\%$ on Aghashahi and $1 - 0.362/0.789 = 54.1\%$ on HK. The spectral-gate break-even activation is higher: $1 - 0.209/0.888 = 76.5\%$ on Aghashahi and $1 - 0.201/0.789 = 74.5\%$ on HK. This means the cheaper spectral gate retains positive latency saving over a wider activation range. Consequently, the HK spectral gate remains below break-even at 47.49% activation and gives 27.0% measured latency saving, whereas the learned gate at 65.56% activation has crossed its approximately 54.1% HK break-even point and gives negative measured saving. A three-tier extension could place an intermediate student CNN between the current tiny and strong models. Such a design would require two independently validation-selected escalation thresholds and platform-specific latency profiling; because it was not evaluated here, it remains a future architecture rather than evidence for the present two-tier system.

C. Reproducibility and Gate-Selection Protocol

The source-file grouped split is a key validity control. It prevents neighbouring windows from the same raw recording appearing in different partitions, reducing the risk of segment-level leakage. The resulting gate-selection protocol is

explicit and target-domain driven. The target-domain IoS gate-selection protocol is: (1) split acoustic recordings by source file; (2) train the learned tiny-CNN gate, strong CNN, and supervised spectral logistic gate on the training split only; (3) evaluate validation F1, ECE, strong-model activation, latency saving, and break-even activation; (4) select learned routing only when tiny-CNN confidence is accurate, calibrated, and requires low activation; (5) select spectral routing when learned confidence is weak but entropy-flux routing gives acceptable F1 with activation below break-even; and (6) use direct strong inference when both gates require near-break-even activation or show a clear F1 deficit. For an IoS network, this protocol should be applied per site or per sensor group rather than as one global threshold shared by all nodes. Different hydrophone or noise-logger nodes may observe different pipe materials, hydraulic conditions, installation depths, background noise, sensor ageing, and seasonal acoustic drift. A small labelled target-domain validation set is therefore needed before deployment, and re-validation should be triggered after sensor replacement, major hydraulic regime change, seasonal noise shift, or sustained calibration degradation. This makes the method a practical IoS edge-inference protocol, not only an offline cascade experiment.

D. Limitations

Several limitations remain. First, the latency values are not embedded-hardware energy measurements. They were obtained under the same Apple M3 CPU software setup for relative comparison; thread affinity, power draw and deployment-level scheduling were not controlled. A deployment study should repeat the full learned-gate and spectral-gate evaluation on a representative low-power platform and report latency, memory use, runtime activations, and energy per segment. Second, the strong CNN is a wider model from the same lightweight acoustic-CNN family rather than a large pretrained audio backbone; this was intentional for constrained IoS-style comparison, but it limits claims against stronger audio classifiers such as pretrained mobile or PANN-style models. Third, the HK buried test set contains only 259 segments from 21 files. A stratified source-file-blocked bootstrap gave a broad 95% interval of [0.5956, 0.9468] for the selected spectral-cascade F1, so the five-seed result should be interpreted as indicative rather than definitive. Fourth, the spectral gate is supervised and must be calibrated on labelled target-domain data; its HK buried result does not imply that the same logistic coefficients transfer to another network or sensor placement. Fifth, a distilled-gate cascade was not evaluated in the reproducible run set, although compact distillation is a natural future baseline [24]. Finally, the ablation compares entropy-only and flux-only feature families but does not isolate all eight statistics individually; its five-seed paired intervals also do not establish statistically significant cascade-F1 superiority for the combined gate.

VI. CONCLUSION

This paper evaluated domain-aware gate selection for acoustic sensor networks in Internet-of-Sounds water pipeline leak monitoring. The focus is water pipeline acoustic data and local edge inference, not general sound recognition or a completed offshore underwater-acoustics deployment. The learned tiny-CNN confidence cascade is effective on the controlled hydrophone dataset, where it closely matches the always-on strong CNN descriptively while reducing activation and measured CPU-latency proxy. On the HK buried dataset, learned confidence routing is less effective, but a supervised spectral entropy-flux gate provides a stronger acoustic-specific routing signal. At $\tau = 0.75$, no statistically significant F1 difference is detected between the spectral-gate cascade and the strong model, while the cascade reduces measured software-level latency by 27.0%. The zero-shot transfer results also show that cross-domain deployment without target-domain calibration is unreliable, so even a small labelled target-domain validation set is important before field use. The practical outcome is an IoS gate-selection protocol: validate the first-stage gate on target-domain acoustic data using F1, calibration, activation rate, break-even latency, and paired seed-level uncertainty before choosing learned routing, spectral routing, or direct strong inference.

REFERENCES

- [1] L. Turchet et al., "The Internet of Sounds: Convergent trends, insights and future directions," *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11264-11292, 2023, doi: 10.1109/JIOT.2023.3253602.
- [2] L. Turchet, G. Fazekas, M. Lagrange, H. S. Ghadikolaei, and C. Fischione, "The Internet of Audio Things: State of the art, vision, and challenges," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10233-10249, 2020, doi: 10.1109/JIOT.2020.2997047.
- [3] C. Stewart, F. Fough, N. Fough, and R. Prabhu, "Optimizing Energy Efficiency in Underwater Acoustic Networks Through Machine Learning Classifiers," in *2024 31st IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2024, pp. 1-2.
- [4] C. Stewart, N. Fough, N. Erdogan, and R. Prabhu, "Performance and Energy Modelling for a Low Energy Acoustic Network for the Underwater Internet of Things," in *2023 IEEE International Workshop on Metrology for the Sea: Learning to Measure Sea Health Parameters (MetroSea)*, pp. 110-115, 2023.
- [5] C. Stewart, B. Ayaz, N. Fough, and R. Prabhu, "Towards the Underwater Internet of Things for Subsea Oil and Gas Monitoring," *Internet of Things*, vol. 30, p. 101510, 2025.
- [6] M. Antonini, M. Vecchio, F. Antonelli, P. Ducange, and C. Perera, "Smart audio sensors in the Internet of Things edge for anomaly detection," *IEEE Access*, vol. 6, pp. 67594-67610, 2018, doi: 10.1109/ACCESS.2018.2877523.
- [7] Z. Hu, S. Tariq, and T. Zayed, "A comprehensive review of acoustic based leak localization method in pressurized pipelines," *Mechanical Systems and Signal Processing*, vol. 161, Art. no. 107994, 2021, doi: 10.1016/j.ymssp.2021.107994.
- [8] M. Aghashahi, L. Sela, and M. K. Banks, "Benchmarking dataset for leak detection and localization in water distribution systems," *Data in Brief*, vol. 48, Art. no. 109148, 2023, doi: 10.1016/j.dib.2023.109148.
- [9] M. Aghashahi, L. Sela, and M. K. Banks, "Dataset of leak simulations in experimental testbed water distribution system," Mendeley Data, Version 1, 2022, doi: 10.17632/tbrnp6vrnj.1.
- [10] A. Fares, I. A. Tijani, Z. Rui, and T. Zayed, "Leak detection in real water distribution networks based on acoustic emission and machine learning," *Environmental Technology*, vol. 44, no. 25, pp. 3850-3866, 2023, doi: 10.1080/09593330.2022.2074320.
- [11] I. A. Tijani et al., "Acoustic based data acquisition for leak detection of water distribution networks," Mendeley Data, Version 1, 2022, doi: 10.17632/hkn8mxcjyz.1.
- [12] R. Liu, T. Zayed, and R. Xiao, "Advanced acoustic leak detection in water distribution networks using integrated generative model," *Water Research*, vol. 254, Art. no. 121434, 2024, doi: 10.1016/j.watres.2024.121434.
- [13] Y. Wu et al., "Advancing deep learning-based acoustic leak detection methods towards application for water distribution systems from a data-centric perspective," *Water Research*, vol. 261, Art. no. 121999, 2024, doi: 10.1016/j.watres.2024.121999.
- [14] Z. Xu et al., "Interpretable deep learning for acoustic leak detection in water distribution systems," *Water Research*, vol. 273, Art. no. 123076, 2025, doi: 10.1016/j.watres.2024.123076.
- [15] R. Liu, T. Zayed, R. Xiao, and Q. Hu, "Time-Transformer for acoustic leak detection in water distribution network," *Journal of Civil Structural Health Monitoring*, vol. 15, no. 3, pp. 759-775, 2025, doi: 10.1007/s13349-024-00845-2.
- [16] Y. Xie et al., "Time-frequency distribution map-based convolutional neural network model for underwater pipeline leakage detection using acoustic signals," *Sensors*, vol. 20, no. 18, Art. no. 5040, 2020, doi: 10.3390/s20185040.
- [17] H. Pan et al., "Underwater acoustic technology-based monitoring of oil spill: A review," *Journal of Marine Science and Engineering*, vol. 11, no. 4, Art. no. 870, 2023, doi: 10.3390/jmse11040870.
- [18] K. Yang, T. Zhou, J. Hui, and C. Xu, "Passive acoustic bearing estimation of underwater gas leak using single vector hydrophone," *Applied Acoustics*, vol. 233, Art. no. 110623, 2025, doi: 10.1016/j.apacoust.2025.110623.
- [19] S. Teerapittayanon, B. McDanel, and H. T. Kung, "BranchyNet: Fast inference via early exiting from deep neural networks," in *23rd International Conference on Pattern Recognition*, 2016, pp. 2464-2469, doi: 10.1109/ICPR.2016.7900006.
- [20] P. Panda, A. Sengupta, and K. Roy, "Conditional deep learning for energy-efficient and enhanced pattern recognition," in *Proc. DATE*, 2016, pp. 475-480, doi: 10.3850/97839815370790819.
- [21] S. Enomoto and T. Eda, "Learning to cascade: Confidence calibration for improving the accuracy and computational cost of cascade inference systems," in *Proc. AAAI Conference on Artificial Intelligence*, vol. 35, no. 8, pp. 7331-7339, 2021, doi: 10.1609/aaai.v35i8.16900.
- [22] X. Wang, Y. Luo, D. Crankshaw, A. Tumanov, F. Yu, and J. E. Gonzalez, "IDK Cascades: Fast deep learning by learning not to overthink," in *Proc. UAI*, 2018, pp. 580-590.
- [23] H. Rahmath P., V. Srivastava, K. Chaurasia, R. G. Pacheco, and R. S. Couto, "Early-exit deep neural network-A comprehensive survey," *ACM Computing Surveys*, vol. 57, no. 3, Art. no. 75, pp. 1-37, 2024, doi: 10.1145/3698767.
- [24] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," presented at the *NIPS Deep Learning and Representation Learning Workshop*, Montreal, QC, Canada, 2014.
- [25] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. ICML*, 2017, pp. 1321-1330.
- [26] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423, 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.
- [27] S. Dixon, "Onset detection revisited," in *Proc. International Conference on Digital Audio Effects (DAFx)*, Montreal, Canada, 2006, pp. 133-137.
- [28] H. Misra, S. Ikbil, H. Bourlard, and H. Hermansky, "Spectral entropy based feature for robust ASR," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, QC, Canada, 2004, vol. 1, p. I-193, doi: 10.1109/ICASSP.2004.1325955.