

Response to Reviewers' Comments

We thank the reviewers for their careful reading of our manuscript and their constructive comments. Below are our point-by-point responses to their comments.

In this response, “Comments of the Reviewer” are presented in black, followed by “Author’s Response” (in red) detailing how we have addressed each comment or providing further explanation where necessary. Additionally, revised text excerpts from the manuscript are provided in blue to show the specific changes made. Locations of these changes are also highlighted in red in the updated manuscript.

Reviewer #1:

Comment 1:

The graphs are not good at all and should be significantly improved. The text on the graphs should be enlarged.

Response:

We appreciate the feedback regarding the figures in our manuscript. We've carefully reviewed and revised all figures to enhance their quality and clarity, including increasing font sizes and adjusting layouts to enhance readability.

We invite you to review the revised figures in our manuscript on [Github](#). We believe these improvements will significantly enhance the overall presentation and understanding of our research.

Revised text:

Figure 1,2,3 has been updated to improve clarity, with enlarged labels and enhanced visual quality to facilitate understanding.

Figure 4 has been replaced with the ROC curve.

Figure 5 has been updated to include threshold 0, 40 and 60. The illegal chinese characters have been converted to english. The color map is set to colorblind friendly.

Comment 2:

The performance of the traditional algorithm, which yielded an accuracy of only 7.2% and an F1 score of 0.135, appears highly unusual and suggests a potential issue. Please provide clarification.

Response:

The low performance of the baseline model (traditional algorithm) highlights the challenges inherent in micro-leak detection when using a single gas flowmeter in a kitchen setting. In this project, approximately 250,000 IoT gas flowmeter devices are deployed. Over the past several years, the city gas supplier has used the baseline model for daily leakage detection. For validation, a subset of positive reports from the baseline model was selected at semi-random intervals for manual door-to-door inspection. These inspections included the use of handheld concentration meters and valve closure pressure retention tests. Our **evaluation dataset** consists of these manually validated positive results, ensuring its reliability.

The baseline model is a logic-based algorithm. Its detection logic is defined as follows: if the duration of continuous and stable gas usage exceeds the corresponding threshold in specific flow intervals, the user is flagged as at risk. The parameter standards currently used in the baseline model are summarized in the table below. These standards were derived from analyzing abnormal pulse patterns in a small sample set.

Security Policies on the Ryder IoT Meter (G2.5):

Flow Interval (pulses/hour)	Duration (hours)	Risk Level
5–10	10	a8
10–30	6	a7
30–60	3	a6
60–120	2	a5
120–300	1	a34

In this table, "duration" refers to the number of hours, "bottom" and "top" represent the hourly pulse counts, and "label" indicates the assigned risk level.

Our proposed model made breakthroughs in the accuracy of the report and the efficiency of the calculation, compared with the baseline model. We developed a physics-based underlying model to generate the training dataset of the seq2seq model used for unsupervised annotation on actual dataset. The performance of our proposed model shows the advantages of using this method.

Revised text:

Section 5, Paragraph 3:

"The baseline model is a logic-based algorithm. Its detection logic is defined as follows: if the duration of continuous and stable gas usage exceeds the corresponding threshold in specific flow intervals, the user is flagged as at risk."

Comment 3:

Please discuss the disadvantages of using this method (if any).

Response:

Thank you for this insightful question. Our project incorporates three interrelated models, each contributing to a comprehensive solution:

- A **physics-based underlying model** that enhances interpretability by simulating datasets used to train the Seq2Seq model.
- A **Seq2Seq generative model** that converts nominal flow data into actual flow estimates, enabling unsupervised annotation for detecting deviations.
- A **CNN-based neural network** that uses the annotated datasets for supervised classification, providing real-time leakage detection.

While this integrated approach provides significant advantages in accuracy and adaptability, it does have some challenges.

The pipeline introduces additional complexity compared to a simple logic-based baseline model. Maintaining this system requires continuous updates to both the training datasets for the CNN model and the simulated datasets for the Seq2Seq model. This demands additional computational resources and expertise.

Our CNN-based model is sensitive to gas appliances equipped with intelligent timing functions. This can lead to false positive (FP) predictions if the pulse time series data is analyzed over a narrow time window (e.g., `window_width = 1` hour). Conversely, using a broad time domain (e.g., `window_width = 24` hours) can be problematic, as certain user behaviors, such as forgetting to close a stove valve, can introduce noise and reduce the system's ability to provide timely risk warnings.

There are also limitations on the use case. The performance of the Seq2Seq model heavily depends on the quality of the simulated datasets generated by the physics-based model. If the simulated datasets fail to accurately represent real-world conditions, the overall detection performance may degrade. This is why we have spent considerable time describing the structure and technical principles of gas flowmeters in Section 3.

Comment 4:

Please make a comparative study in terms of results with other methods from the literature.

Response:

Gas companies often treat customer data as a private asset, resulting in a relatively closed data-sharing culture in the residential and industrial gas monitoring domain. This makes it exceptionally challenging to find labeled datasets publicly available

online, let alone established methods for processing such data. The conservative nature of this environment significantly limits the effectiveness of traditional algorithms (baseline models). We have extensively searched databases such as Google Scholar using keywords like "small leak," "gas flowmeter," and "smart house," but found few closely related scenarios or boundary conditions in published research.

Moreover, our raw data is organized in a highly specific structure, which may require an appendix for clarity. The uniqueness of our data in this application scenario further complicates the task of identifying detailed third-party methods online. For more information on the pulse data from gas flowmeters, please refer to our [notebook](#).

We welcome contributions and encourage the addition of new models to our project. To facilitate this, we have provided access to our evaluation dataset (anonymized by encoding sensitive information). Any new models can be evaluated using this dataset, and we will update the ranking list accordingly as new results become available.

Comment 5:

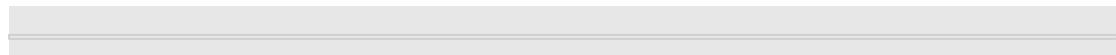
In the title change “based” to “Based”.

Response:

The title has been revised as requested.

Revised text:

Title: "Micro Leak Detection of Natural Gas for Residential Users Based on Gas Flowmeters"



Reviewer #2:

Comment 1:

The authors tackle the interesting problem of detecting micro-leaks within houses: leaks are detected according to a leak threshold Θ . The authors provide an adequate description of the physical model, which can be better revised to actually match the way the formulas were computed within the simulation for generating the data (are you actually using the integrals, or are you performing a discrete time optimization requiring an approximation of the integral to a summation?).

Response:

Thank you for your thoughtful comment and for highlighting the need for clarification regarding our computational approach. In response, we have revised the Section 4.1 to explicitly state that the integrals are approximated using summation for discrete time steps in our simulations. This ensures consistency between the described physical model and the actual computations performed during data generation. You can find more details in our example [script](#).

Revised text:

Section 4.1, Paragraph 3:

"To replicate realistic gas flow patterns, we referred to the typical gas usage of common household appliances, such as water heaters, stoves, and boilers, as outlined in Table~\ref{tab:gas_consumption}. The simulation model was built on pulse metering principles derived from the operational characteristics of membrane gas flowmeters, mapping each device's gas consumption to nominal and actual flow rates under steady-state conditions. The integrals are approximated using summation for discrete time steps in our simulations. "

Comment 2:

Despite the authors are correctly addressing the literature on leak detection, and given that they are also discussing the way to determine the classification at real-time (e.g., whether there is actually a leak or not), the authors provide marginal insight on the real-time classification task, and how their model compares to other state-of-the-art time series classification tasks.

Response:

Thank you for your valuable feedback. We have expanded our discussion of the real-time classification process to provide greater clarity. Our primary focus in this work has been on developing an effective data acquisition and annotation pipeline using the Seq2Seq model, and leveraging a CNN-based neural network as the backbone for time series classification tasks. While our current approach demonstrates reliable performance, we acknowledge the potential for further improvement by exploring more advanced architectures. In future work, we plan to investigate and integrate state-of-the-art time series classification models into our framework and conduct a comprehensive comparative analysis to benchmark performance across various conditions. This iterative approach will allow us to further optimize the classification process and address the broader spectrum of challenges in real-time leak detection.

Revised text:

Section 5, Paragraph 8:

"Compared to baseline model, our approach offers notable advantages. Firstly, the physics-based underlying model improves interpretability and allows the generation of realistic simulated datasets, bridging the gap between synthetic and real-world scenarios. Secondly, the CNN-based classifier excels in detecting leaks under low-flow conditions, where traditional methods often struggle due to insufficient signal strength. Furthermore, we implement a likelihood-based confidence scoring mechanism, normalized between 0 and 100, allowing flexible adjustment of the classification threshold (Θ) to meet varying operational requirements."

Section 5, Paragraph 9:

"A key contribution of this work is the introduction of a low-cost method for unsupervised data labeling of gas flowmeter pulse time series, addressing the scarcity of annotated data—a major limitation for traditional approaches. Even with the simplest CNN model, our methodology achieves remarkable performance in binary time-series classification tasks, highlighting its efficiency and practicality."

Comment 3:

As a side note, most of the figures are completely illegible, as most of the labels are given in Chinese, thus not being suitable for an international audience.

Response:

All figures have been updated with English labels to ensure they are accessible to the international audience.

Revised text:

Figure 1,2,3 has been updated to improve clarity, with enlarged labels and enhanced visual quality to facilitate understanding.

Figure 4 has been replaced with the ROC curve.

Figure 5 has been updated to include threshold 0, 40 and 60. The illegal chinese characters have been converted to english. The color map is set to colorblind friendly.

Comment 4:

Given this, it seems that the authors are providing the classification by estimating the leak through a Seq2Seq model which transforms the nominal flows into the target flow. Then, the pulse model for a real-life sensor is likely attach to the resulting sequence, from which the leak can be detected and the pulses can be sent. Not with standing the former, the authors provide no explicit pipeline to process these, and therefore the previous sentence is mainly resumed from the information being available form the text.

Response:

Thank you for your detailed observation. We have expanded the explanation of our pipeline, explicitly detailing the role of each component in the leakage detection process. Our project incorporates three interrelated models:

- A **physics-based underlying model** improves interpretability and simulates datasets for training the Seq2Seq model.
- A **Seq2Seq generative model** transforms nominal flow data into actual flow estimates for unsupervised annotation.
- A **CNN-based neural network** performs supervised classification on the labeled datasets for real-time leakage detection.

The pipeline now provides a comprehensive and explicit process, connecting each model to its specific task and emphasizing their integration in the overall system.

Revised text:

Abstract:

This paper presents a novel approach to detecting small residential gas leaks by developing and testing a leakage detection algorithm, termed the Weeg algorithm. Our proposed methodology combines a Seq2Seq generative model with a CNN-based classification network, enabling a robust and efficient approach to this task. The evaluation was performed on a dataset containing 86 leakage samples and 1103 non-leakage samples. The Weeg algorithm demonstrated significant improvements, achieving an accuracy of 90.2% and an F1 score of 0.458, while reducing false positive rates compared to conventional logic-based algorithms.

Section 4.3, Paragraph 3:

"This method of unsupervised labeling is particularly advantageous because it eliminates the need for labor-intensive manual labeling of leakage events, which are often rare and difficult to observe directly. Additionally, it enables the generation of large labeled datasets from real-world data, facilitating the training of downstream models for real-time leak detection. These annotations are used to train our primary detection model—a CNN-based neural network designed for binary classification (leakage or no leakage)."

Section 5 Paragraph 4:

"The CNN-based neural network takes nominal flow rates (pulse sequences) as input. Its task is to classify each sequence as either "leakage" or "no leakage." The CNN model produces a likelihood score, which we normalize to a range of 0–100. This score serves as a confidence metric, and we fine-tune the threshold parameter to balance the model performance. When the threshold is set to 0, the proposed model degenerates into the baseline model."

Comment 5:

Still, the authors should consider to better identify how the Seq2Seq model is then associated to the detection of a leak. If they are actually using the pulse model, then the overall task boils down to a binary classification problem, where the binary condition is just associated to the increase of the Θ parameter. Thus, the overall analysis should also reflect the similarity between the forecasted resulting sequence and the expected one (e.g., Dynamic Time Warping), thus remarking the similarity and differences between predicted and expected sequence. This outcome will then motivate the results already provided in Table 2 and 3, which are likely to be associated with the aforementioned pulse detection but, the classification outcome shall be ultimately related to the quality of the resulting time series.

Response:

Thank you for this valuable observation. Indeed, our pipeline integrates the Seq2Seq model as a preprocessing step for unsupervised annotation. This model helps convert observed pulse time series (nominal flow rates) into sequences annotated with potential leakage events. These annotations are used to train our primary detection model—a CNN-based neural network designed for binary classification (leakage or no leakage).

The CNN-based neural network takes nominal flow rates (pulse sequences) as input. Its task is to classify each sequence as either "leakage" or "no leakage." The CNN model produces a likelihood score, which we normalize to a range of 0–100. This score serves as a confidence metric, and we fine-tune the threshold parameter to align predictions with specific business requirements. When the threshold is set to 0, the proposed model degenerates into the baseline model.

To better demonstrate the effectiveness of our classification approach, we have incorporated a Receiver Operating Characteristic (ROC) curve. This provides a clear visual representation of the trade-off between true positive and false positive rates at varying thresholds. The ROC curve complements the results in Tables 2 and 3, which highlight the performance metrics under different parameter settings.

While we recognize the potential of using Dynamic Time Warping (DTW) to assess the similarity between forecasted and expected sequences, this aspect remains unimplemented in the current study. We agree that DTW could provide valuable insights into the alignment and quality of the resulting time series, and we plan to explore this avenue in our future work. Incorporating DTW could further enhance the robustness of the Seq2Seq annotations and improve the overall pipeline.

