Detecting Defecation Premonition from the Acoustic Activity of Bowel Sounds

Shota Miyagawa*, Toshitaka Yamakawa*†, Masayuki Tanabe‡§ and Kazushi Ikeda*
* Nara Institute of Science and Technology, Japan
E-mail: kazushi@is.naist.jp
† Quadlytics Inc.
‡ Salmontech Inc.
§ Kumamoto University, Japan

***Abstract*—This study investigates a method for detecting pre-defecatory signs from bowel sounds to help prevent fecal incontinence. We compared two analytical approaches: a macro-analysis of entire 8-second audio clips and a micro-analysis of individually detected bowel sound events. Using data collected from a single subject, multiple machine learning models were evaluated for their ability to classify audio as either belonging to a specific time window before defecation (ranging from 10 to 270 minutes) or to other times. The results consistently showed that the macro-analysis outperformed the micro-analysis in overall discriminative performance. A CNN-BiLSTM model achieved the highest performance in the macro-analysis, with a peak AUC of 0.76 for the 60-minute pre-defecation window. A subsequent statistical analysis revealed that this performance difference is attributable to a significant increase in the frequency and total duration of bowel sound events prior to defecation ($p < 0.001$, $r = 0.378$). These findings indicate that the quantitative density of bowel sound events is a more dominant indicator for detecting defecation premonition than the qualitative acoustic characteristics of individual sounds. This suggests that focusing on event frequency analysis is a more promising direction for developing non-invasive prediction systems.**

I. Introduction

In today's aging society, fecal incontinence is a serious problem faced by many older adults. This condition not only significantly impairs the physical and psychological quality of life (QoL) of patients but also contributes to an increased burden on caregivers, making it a pressing social issue. If the timing of defecation could be predicted in advance, it would enable planned toilet assistance and care, which is expected to greatly contribute to improving the QoL for both the individuals affected and their caregivers.

Previous efforts to predict defecation have been reported, such as using a pocket-sized ultrasound device to estimate rectal fecal volume [1]. However, this method faces challenges related to the operator's skill and the subjective interpretation of the resulting images. In contrast, this study focuses on bowel sounds, which are acoustic signals that have the significant advantage of being recordable non-invasively and continuously simply by placing a microphone on the abdomen. Bowel sounds are generated by the physiological activity of the digestive tract, specifically the movement of gas and contents within the intestinal lumen [2]. Furthermore, as this data can be acquired as objective digital signals, it is highly compatible with automated analysis.

Prior research on defecation prediction using bowel sounds has shown that a small number of audio samples recorded immediately before and after defecation can be classified by machine learning models [3]. While this result suggests that the acoustic characteristics of bowel sounds may differ before and after defecation, the findings remain limited due to the small sample size. Moreover, that study compared physiologically distinct states (pre- vs. post-defecation) and did not investigate the possibility of capturing earlier premonitory signs within the continuous time leading up to the event, which is essential for practical application.

Therefore, this study aims to clarify what kinds of distinguishable acoustic features appear in bowel sounds during specific time windows before defecation. Our fundamental hypothesis is that as defecation approaches, changes in the intestinal environment, such as the movement and accumulation of feces and increased peristalsis, will manifest as changes in the acoustic characteristics (e.g., timbre) or occurrence patterns (e.g., frequency, rhythm) of bowel sounds.

To capture these characteristic changes, this study establishes two different analytical perspectives and directly compares their effectiveness:

- **Macro-analysis**: Treats the entire recorded 8-second audio clip as a single unit, analyzing global temporal context information such as the frequency and density of bowel sounds contained within it.
- **Micro-analysis**: Precisely extracts individual bowel sound events from the audio clip, eliminates background noise, and performs a detailed analysis of the acoustic characteristics (e.g., timbre) of each event.

Our initial hypothesis was that the micro-analysis, by clearly isolating individual bowel sound events, would more accurately capture subtle changes in timbre compared to the macro-analysis, where background noise is present, thus leading to higher-precision prediction.

This paper reports the results of systematically evaluating the performance of both approaches using multiple time thresholds (i.e., how many minutes before defecation a premonitory sign appears) to test this hypothesis.

979-8-3315-7206-8/25/$31.00

II. Related Work

In state estimation from acoustic signals, a classical and widely used approach involves using statistics (e.g., mean and variance) of Mel-Frequency Cepstral Coefficients (MFCCs) [4] extracted from an entire audio clip as features. This approach has a proven track record in fields such as speech emotion recognition [5]. It has also been employed in prior research on defecation prediction, where MFCC statistics were used as features for classification with SVM and Gradient Boosting. However, this approach has an inherent limitation: it condenses dynamic, time-varying information into static indicators, thus failing to capture the dynamic patterns within the signal.

It is known that bowel sounds consist of multiple acoustic patterns, such as “Single Burst” and “Continuous Random Sound,” and an analysis that preserves the time-series structure is essential to distinguish these features. To address this challenge, this study includes a model combining a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) network as one of the evaluation targets. This architecture aims to capture dynamic changes by having the CNN extract local acoustic features from a spectrogram and the subsequent LSTM learn the temporal dependencies between those features.

Separate from the above, a more micro-level approach also exists, which involves precisely detecting individual acoustic events before extracting features. For example, in a study on classifying Inflammatory Bowel Disease (IBD), an EfficientNet+U-Net segmentation model was first used to automatically detect bowel sound segments, and the mean of MFCCs was then calculated only from these detected segments to be used as features [6]. This method aims for a more precise analysis by limiting the scope to the sound of interest (bowel sounds) and reducing the influence of background noise.

However, the detection models used in these micro-level approaches often require supervised data that has been manually annotated to define events, and there are issues with the subjectivity and reproducibility of these criteria. For instance, the IBD classification study considered events with an interval of less than 100ms to be the same event, while another study defined events as separate if the interval was greater than 10ms [7]. Such inconsistencies in annotation standards not only make it difficult to compare results across studies but also make the annotation process itself a high-cost task.

Considering these challenges, the micro-analysis in this study adopted a statistical Voice Activity Detection (VAD) method that does not require annotation. Specifically, by using a simpler Gaussian Mixture Model (GMM), we aimed for objective and reproducible segmentation, avoiding the issues of subjectivity and cost associated with prior work.

From the perspective of model sophistication, transfer learning, which applies powerful pre-trained models developed in the field of image recognition to audio spectrograms (treated as images), has recently garnered attention. Following this approach, this study also evaluated the effectiveness of applying EfficientNet, a model renowned for its high performance in image recognition [8], to the spectrogram analysis of bowel sounds.

Thus, while both macro-analysis (treating the whole audio clip) and micro-analysis (treating individual events) exist for bowel sound analysis, each has noted advantages and challenges (e.g., loss of temporal information, ambiguity in event definition). It is particularly unclear which approach is more effective for the specific task of detecting defecation premonition. Therefore, this study aims to provide a clear answer by directly comparing these two approaches under the same dataset and conditions to determine which is more effective for detecting defecation premonition.

III. Methods

A. Data Collection

In this study, bowel sounds were collected from a single healthy male subject in his 20s using an electronic stethoscope (AMI-SSS01 series). The sampling frequency was set to 8000 Hz. The stethoscope was placed on the subject’s abdomen, over the sigmoid colon. To minimize the influence of external noise, measurements were conducted in a quiet room while the subject maintained a supine, resting posture. Recordings were performed, in principle, at approximately 30-minute intervals, with three consecutive 8-second recordings taken in each session. However, due to the subject’s availability, recordings were occasionally skipped, meaning the actual measurement intervals were not always uniform. These three consecutive files were managed as a single group.

B. Problem Formulation and Labeling

As mentioned above, the data collected in this study is discontinuous with non-uniform measurement intervals. Because applying rigorous regression or time-series analysis to such data is difficult, we adopted an exploratory approach, formulating the problem as a binary classification task to verify whether a distinguishable difference exists in bowel sounds between a specific time window before defecation and other times.

Specifically, data recorded within X minutes of the defecation time were labeled as the positive class (Class 1), while all other data were labeled as the negative class (Class 0). To explore the optimal time window, this study evaluated performance by varying the threshold X among 10, 30, 60, 90, 120, 150, 210, and 270 minutes.

C. Analysis Approaches and Models

To compare the effectiveness of the macro- and micro-analyses described in the introduction, we constructed models for both approaches. The features, MFCC and Mel spectrogram, were calculated with an FFT size of 320 and a hop length of 128. A band-pass filter (60Hz–2000Hz) was used to focus on the frequency band containing bowel sounds.

1) Macro-analysis (Whole Audio Analysis): This approach uses the entire 8-second audio clip as a single unit of analysis. The combinations of features and models used are as follows:

- **MFCC Statistics + SVM**: 40-dimensional MFCCs were extracted, and their mean and variance were standardized and used as features for classification with an SVM (RBF kernel). The hyperparameters were the default values in scikit-learn.
- **MFCC Time-series + 2DCNN-BiLSTM**: 40-dimensional MFCCs were treated as a time-series and classified using a model composed of a 4-layer CNN and a 2-layer bidirectional LSTM (128 hidden units).
- **Mel Spectrogram + EfficientNet-B0**: A 64-dimensional Mel spectrogram was calculated and resized using linear interpolation to match the input size of the EfficientNet-B0 model. The pre-trained model was then fine-tuned for classification.

2) Micro-analysis (Per-Event Analysis): This approach first detects individual bowel sound events from the audio clip and then analyzes them on a per-event basis.

a) Event Detection: To extract bowel sound events, we used Google's WebRTC Voice Activity Detection (VAD) [9], which does not require manual annotation. Internally, it classifies 10ms audio frames as voiced or unvoiced using a Gaussian Mixture Model (GMM) with log energies from multiple frequency bands as features. As this GMM is updated online and sequentially, stable detection is expected even with the steady background noise assumed in this study. The detection logic marks a segment as starting if 3 or more out of 6 consecutive frames (60ms) are classified as voiced and ending if 3 or more are unvoiced, thereby suppressing transient misclassifications. Fig. 1 illustrates an example of bowel sound event extraction from an actual audio waveform using this method.

b) Model Application: Each individual bowel sound event extracted by the above method was treated as a single unit of analysis and classified using the following models. Since the extracted events have variable lengths, they were zero-padded to match the length of the longest event in the dataset before being input to the time-series models.

- **MFCC Statistics + SVM**
- **MFCC Time-series + BiLSTM**
- **Mel Spectrogram + BiLSTM**

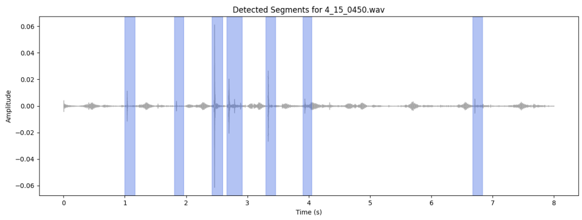

Fig. 1. Example of bowel sound event segmentation by WebRTC VAD.

D. Evaluation Method

a) Model Evaluation: Cross-validation was used to evaluate the generalization performance of the models. Specifically, StratifiedGroupKFold ($n\_splits = 10$) was employed to prevent data from the same recording session (a group of 3 files) from being split into both the training and validation sets, aiming for a more realistic performance evaluation. The evaluation metrics used were Accuracy, AUC (Area Under the Curve), and the F1-Score for the positive class.

For training the deep learning models, the Adam optimizer (initial learning rate: $1e^{-3}$) was used. The learning rate was decayed down to $1e^{-5}$ using a cosine annealing scheduler. Training was conducted for 50 epochs, and the model from the epoch with the minimum validation loss was used for the final evaluation. To address the data imbalance between classes, the class_weight='balanced' option was used for the SVM, while oversampling of the minority class was applied for the deep learning models. Note that methods like SMOTE, which generate new synthetic samples, were not used, as creating data not present in the original recordings was deemed inappropriate for this study's objective of verifying if a difference exists between the two classes.

b) Statistical Evaluation: To investigate the factors behind the performance difference between the macro- and micro-analyses, the Mann-Whitney U test (significance level $p < 0.001$) was used to check for statistical differences in the "number of events" and "total duration of events" per 8-second audio clip between Class 0 and Class 1. The effect size r was also calculated to evaluate the magnitude of the difference.

IV. Results

This section presents the results of capturing defecation premonition from the different analytical viewpoints of "macro" and "micro." First, we compare the classification performance of the two approaches to clarify their respective characteristics. Next, to investigate the factors behind the performance difference, we analyze the statistical characteristics of the acoustic events.

A. Classification Performance by Approach

First, we compared the classification performance of the two approaches (macro-analysis and micro-analysis). Fig. 2 (a), (b), and (c) show the trends in Accuracy, AUC, and F1-Score, respectively, for each model at different positive class definition thresholds.

As seen in the AUC scores in Fig. 2 (b), the macro-analysis (solid lines), which analyzes the entire audio clip, consistently outperformed the micro-analysis (dashed lines), which analyzes on a per-event basis, across all time thresholds. Table I summarizes the peak performance of the main models for each approach at their optimal thresholds.

In the macro-analysis approach, the CNN-LSTM model achieved the highest discriminative performance, with an AUC of 0.76 at the 60-minute threshold. In contrast, the micro-analysis approach generally showed lower performance, with a particular tendency for low AUC scores.

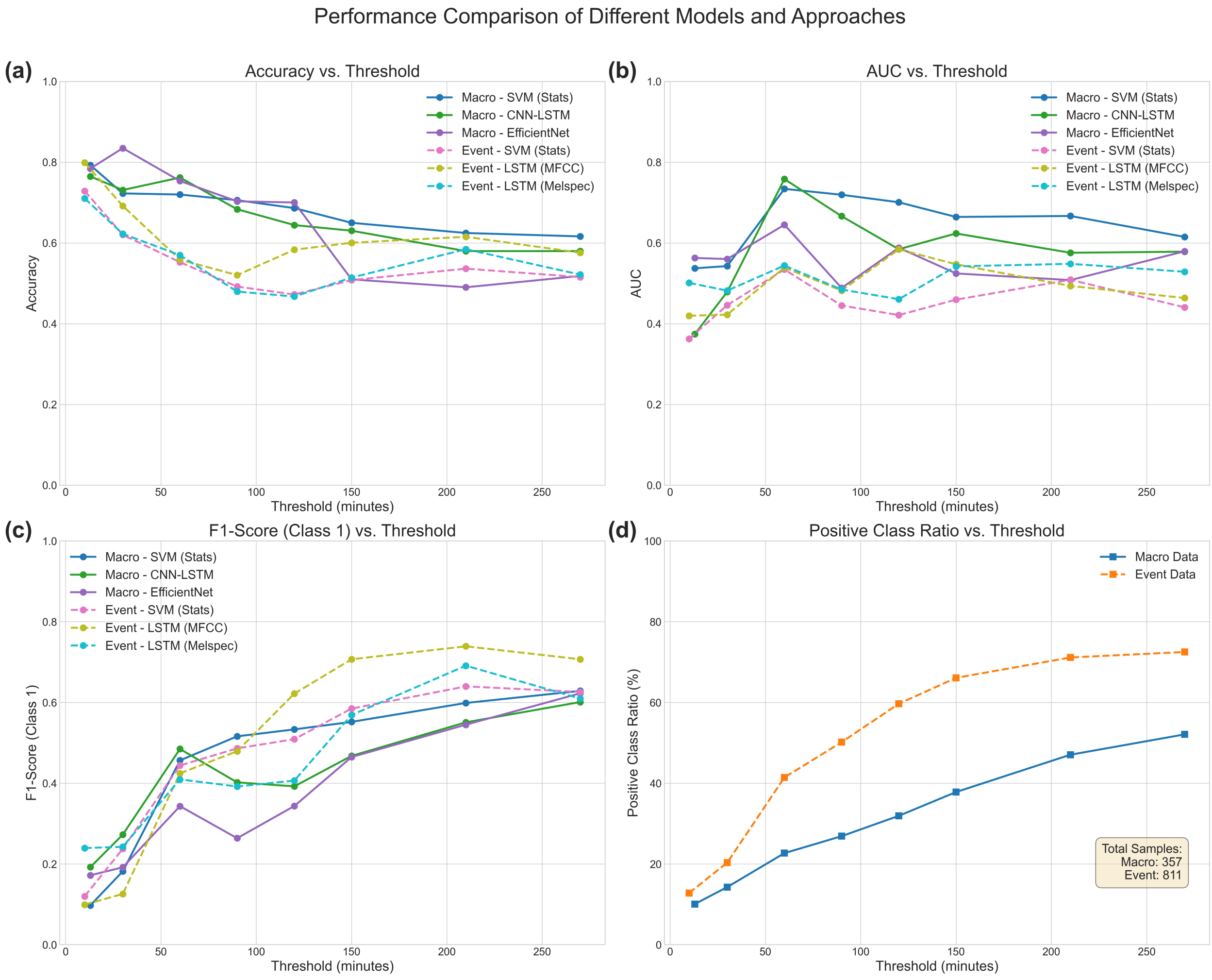

Fig. 2. Trends in (a) Accuracy, (b) AUC, and (c) F1-Score for each model at different time thresholds.

TABLE I
PEAK PERFORMANCE (AUC, F1-SCORE) OF MAIN MODELS FOR EACH APPROACH AT OPTIMAL THRESHOLDS.

Approach	Model	Threshold	AUC / F1-Score
Macro	MFCC+CNN-LSTM	60	0.76 / 0.48
Macro	MFCC stats+SVM	60	0.73 / 0.46
Macro	Melspec+EffNet	60	0.64 / 0.34
Micro	MFCC+LSTM	120	0.58 / 0.62
Micro	Melspec+LSTM	210	0.55 / 0.69
Micro	MFCC stats+SVM	60	0.53 / 0.44

However, a different trend is observed when focusing on the F1-Score in Fig. 2 (c). The two LSTM models used in the micro-analysis (per-event MFCC+LSTM and per-event Melspec+LSTM) began to outperform the other models when the threshold exceeded 90 minutes. Notably, the MFCC+LSTM model recorded the highest F1-Score among all models, 0.74, at the 210-minute threshold, although its AUC remained low at 0.49. This is likely because the F1-Score improved as the number of positive class samples increased with larger thresholds, as shown in Fig. 2 (d), mitigating the data imbalance. This result suggests that the acoustic features of individual events do contain some clues for identifying the positive class. However, the persistently low AUC scores indicate that these features alone are insufficient to clearly distinguish from the negative class, limiting the overall discriminative performance.

B. Statistical Characteristics of Event Frequency

Next, to investigate the background of the performance difference between the two approaches, particularly the superiority of the macro-analysis in AUC score, we compared the number of bowel sound events and their total duration per 8-second audio clip between the classes at the 60-minute threshold setting (Table II).

The results of the Mann-Whitney U test confirmed a statistically significant difference between the classes for both the number of events (Class 0: mean 1.73, Class 1: mean 4.10) and the total event duration (Class 0: mean 0.30s, Class 1: mean 0.77), with $p < 0.001$ for both. The effect size, $r = 0.378$, indicates a medium-sized effect, which is a non-negligible

TABLE II
COMPARISON OF EVENT COUNT AND TOTAL DURATION BETWEEN CLASSES AT 60-MINUTE THRESHOLD.

Metric	C0 (n=275)	C1 (n=82)	p / r
Number of Events	1.73 ± 2.98	4.10 ± 4.57	$< .001$ / 0.378
Total duration (s)	0.30 ± 0.56	0.77 ± 0.91	$< .001$ / 0.378

difference in practical terms.

This result clearly indicates that audio from the period approaching defecation (Class 1) contains statistically significantly more and longer bowel sound events than audio from other times (Class 0). This suggests that the "event density per unit time" is an extremely powerful piece of information for distinguishing between the two classes.

V. DISCUSSION

In this chapter, we discuss the performance difference between macro- and micro-analysis for defecation premonition detection, based on the experimental results.

A. Superiority of Macro-analysis and the Importance of Event Density

The most significant finding of this study is that the macro-analysis of entire audio clips demonstrated higher overall discriminative performance (especially in AUC score) than the micro-analysis of individual events. This result refutes our initial hypothesis that the micro-analysis, by precisely analyzing the sound of interest after removing noise, would achieve higher performance.

The reason for this performance gap is evident from the statistical analysis in Table II. In the state approaching defecation (Class 1), the number and total duration of bowel sound events per unit time increased statistically significantly compared to the normal state (Class 0). This strongly suggests that the temporal context information of "event density," rather than the acoustic characteristics (e.g., timbre) of individual sounds, was the more dominant feature for capturing defecation premonition. It is conceivable that the macro-analysis approach allowed the model to implicitly learn this "event density" by observing the entire 8-second clip, whereas the micro-analysis lost this crucial context by breaking the audio down into individual events.

This observed increase in acoustic activity has a plausible physiological basis. Before defecation, the colon undergoes strong propulsive contractions (High-Amplitude Propagating Contractions, or HAPCs) to move fecal matter toward the rectum. The resulting distension of the rectum is what triggers the urge to defecate [10]. The increased event density captured in this study is likely an acoustic manifestation of this heightened intraluminal activity—namely, the movement of feces and gas caused by strong colonic contractions.

B. Real-time Behavior and Challenges of the Model

To verify how the best-performing macro-analysis model, the CNN-LSTM (at the 60-minute threshold), behaves over time, we applied it to a new 5-day dataset not used in the previous analyses. This dataset also includes records of life events such as meals and waking times.

Fig. 3 plots the model's temporal output. As hypothesized, the model's predictions (Pred=1, red) form clusters concentrated within the 60-minute window preceding each defecation event (highlighted in light red). This suggests that the model has the potential to function not just as a classifier but as a practical premonitory detection system. However, this result also simultaneously reveals two significant limitations for practical application.

- **Limitation 1: False Negatives (Missed Detections in the Pre-defecation Segment).** Many segments immediately preceding defecation are still classified as Pred=0 (no premonition). This is likely because the short 8-second input window makes the model's prediction susceptible to temporary lulls in bowel sound activity, or "silent periods." Extending the analysis window or aggregating predictions across multiple segments might enable more stable premonition detection.
- **Limitation 2: False Positives (Erroneous Detections After Meals/Waking).** Incorrect Pred=1 predictions frequently appear after meals (dotted lines) or waking (dot-dash lines). These are periods when digestive tract motility is known to increase due to postprandial or gastro-colic reflexes. The resulting acoustic activity (event density) likely mimics the pre-defecatory state. To solve this problem, an approach that either captures acoustic patterns specific to defecation, beyond just event density, or incorporates physiological indicators (life logs) like meal and wake times as model features would be necessary.

C. Limitations of Micro-analysis and Future Outlook

On the other hand, the micro-analysis approach showed low classification performance overall, especially in its AUC score. To investigate this result further, we statistically compared the acoustic features (RMS, ZCR, MFCC dimensions, etc.) of individual bowel sound events—averaged over the frames within each event—between the classes, but again, no clear significant difference was found.

These results—namely, the low performance of the classification models and the lack of statistical significance—can be interpreted through multiple factors. The first factor is that our analysis averaged out the temporal changes within a single bowel sound event. Individual bowel sound events may contain dynamic patterns effective for discrimination, such as the sharpness of the sound's attack, and this information may have been lost through averaging.

The second factor is the effect of lumping all "positive class bowel sound events," recorded under different days and physical conditions, together for comparison. While a consistent pattern of changing timbre might exist within a single defecation sequence as it approaches, this pattern could be cancelled out when all events are compared together if the pattern differs from day to day.

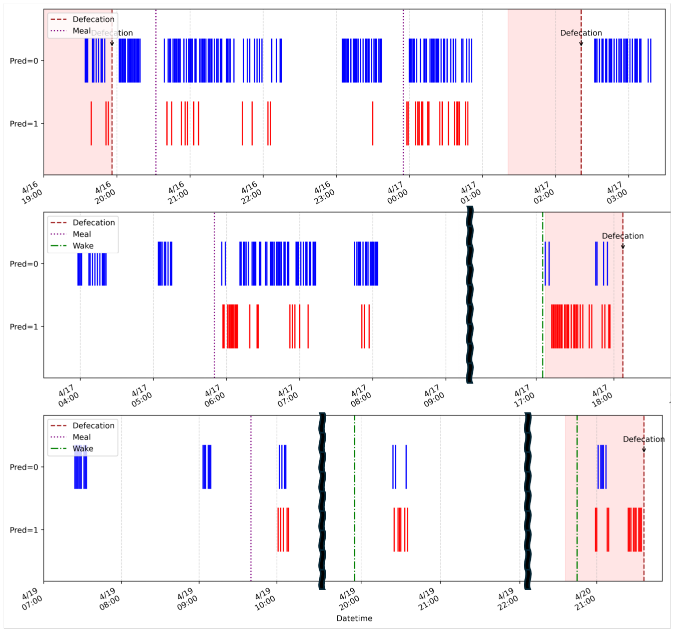

Fig. 3. Time-series visualization of defecation prediction model results. The horizontal axis represents recording date and time, and the vertical axis represents the prediction label by the model. Red dots (Pred=1) indicate samples classified as having defecation premonition, and blue dots (Pred=0) indicate no premonition. Dashed lines represent defecation times, dotted lines represent meal times, and chained lines represent waking times. The period 1 hour before defecation is highlighted in light red.

To test these hypotheses, future work will require collecting data continuously over longer periods and at regular intervals to analyze how the acoustic features of bowel sound events transition over time within a single defecation sequence (e.g., in the 60 minutes before defecation). By enabling the capture of not only event "density" but also the time-series changes in acoustic "quality," we expect to build even more accurate prediction models.

VI. Conclusion

This study aimed to detect defecation premonition through acoustic analysis of bowel sounds, comparing a macro-analysis, which considers entire audio clips, with a micro-analysis focusing on individual sound events. The results demonstrated that macro-analysis achieved higher discriminative performance, primarily due to a significant increase in event frequency as defecation approached, rather than changes in the acoustic characteristics of individual sounds. This finding suggests that emphasizing the frequency of bowel sound events is a promising strategy for defecation prediction.

Nevertheless, this study has several limitations. Since the dataset was limited to a single subject, the generalizability of the findings remains unverified. In addition, real-time analysis indicated that the model also responded to bowel activity unrelated to defecation, such as postprandial or awakening events. These results should therefore be interpreted with caution.

As a baseline investigation, this work provides initial evidence of the feasibility of bowel-sound-based prediction. Future studies should (i) expand the cohort to include individuals of diverse ages, genders, and lifestyles, (ii) develop algorithms that can distinguish defecation-related activity from other physiological phenomena, and (iii) advance sequential modeling of continuously acquired bowel sounds to enable real-time updates and increase confidence in premonition estimation. Addressing these challenges will be crucial to establishing a non-invasive, high-precision technology that can ultimately improve the quality of life of affected individuals.

Acknowledgment

This paper is supported in part by JST [Moonshot R&D] JPMJMS2034, JSPS KAKENHI 22K19833, and KSAC 2024_10.

References

[1] K. Yabunaka, M. Matsumoto, M. Yoshida, *et al.*, "Assessment of rectal feces storage condition by a point-of-care pocket-size ultrasound device for healthy adult subjects: A preliminary study," *Drug Discoveries & Therapeutics*, vol. 12, no. 1, pp. 42–46, 2018.

[2] C. Liu, S. Huang, and H. Chen, "Oscillating gas bubbles as the origin of bowel sounds: A combined acoustic and imaging study," *Chinese Journal of Physiology*, vol. 53, pp. 245–253, 2010.

[3] K. Kishishita and K. Ikeda, "Defecation prediction system using & A mathematical model of bowel sounds," *ICMLC*, 2024.

[4] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.

[5] A. Milton, S. S. Roy, and S. T. Selvi, "Svm scheme for speech emotion recognition using mfcc feature," *International Journal of Computer Applications*, vol. 69, no. 9, pp. 34–39, 2013.

[6] A. Baronetto, S. Fischer, M. F. Neurath, and O. Amft, "Automated inflammatory bowel disease detection using wearable bowel sound event spotting," *Front. Digit. Health*, vol. 7, 2025.

[7] Y. Kutsumi, N. Kanegawa, M. Zeida, H. Matsubara, and N. Murayama, "Automated bowel sound and motility analysis with cnn using a smartphone," *Sensors*, vol. 23, no. 1, p. 407, 2023.

[8] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *ICML*, pp. 6105–6114, 2019.

[9] Google, "Webrtc voice activity detection (vad)," *GitHub repository, https://github.com/wiseman/py-webrtcvad*, accessed July, 2025.

[10] S. Palit, P. J. Lunniss, and S. M. Scott, "The physiology of human defecation," *Digestive Diseases and Sciences*, vol. 57, 2012.