ISSN 2442-6571

Enhancing minority class recognition in cattle monitoring: a robustness analysis of lightweight decision-level fusion

Putri Nayla Sabri [a,1,*], Nisrina Nurhafizhah [a,2], Amrul Faruq [b,3], Acmad Fauzan Hery Soegiharto [c,4], Muhammad Ilham Perdana [a,5]

[a] Informatics Department, University of Muhammadiyah Malang, Malang, Malang, 65144, Indonesia
[b] Electrical Engineering, University of Muhammadiyah Malang, Malang, Malang, 65144, Indonesia
[c] Mechanical Engineering, University of Muhammadiyah Malang, Malang, Malang, 65144, Indonesia
[1] putrinaylasabri@webmail.umm.ac.id; [2] nisrinanuha@webmail.umm.ac.id; [3] faruq@umm.ac.id; [4] achmadfauzan@umm.ac.id;
[5] ilhamperdana@umm.ac.id
* corresponding author

ARTICLE INFO

Article history
Received March 11, 2025
Revised April 15, 2026
Accepted April 17, 2026
Available online May 31, 2026

Keywords
Cattle Behavior Classification
Decision Level Fusion
Edge AI
Imbalanced Data
SMOTE

ABSTRACT

Accelerometer-based monitoring has become an important approach in precision livestock farming, but achieving both high sensitivity for rare, health-relevant behaviors and computational efficiency remains challenging. Deep learning methods can address class imbalance but are often unsuitable for edge deployment due to their computational cost. This study evaluates the robustness of a lightweight decision-level fusion framework for classifying imbalanced cattle behavior using triaxial accelerometer data. To ensure rigorous evaluation, the Synthetic Minority Over-sampling Technique (SMOTE) was applied only in the training feature space to prevent data leakage. Because the dataset is strongly imbalanced, with Salt Licking (SLT) as the minority yet health-relevant class, model performance was assessed using Macro-F1 for global robustness and Recall_SLT for rare-event sensitivity. Individual tree-based models achieved strong results, with the best single model, XGBoost, obtaining 93.70% accuracy and 0.9421 Macro-F1. The proposed soft-voting fusion of Extra Trees, XGBoost, and CatBoost further improved performance to 94.21% accuracy and 0.9447 Macro-F1, with a statistically significant gain over the best single model (Wilcoxon signed-rank test, (p=0.0326). The framework also maintained strong minority-class recognition, with SLT achieving precision = 0.9571, recall = 0.9853, and PR-AUC = 0.9990. These results show that lightweight decision-level fusion can improve robustness and rare-event sensitivity without temporal deep learning, making it suitable for resource-constrained edge monitoring in livestock systems.

1. Introduction

Precision livestock farming (PLF) supports animal health, welfare, and productivity through continuous monitoring and data-driven management [1], [2], [3]. Among sensing technologies, wearable tri-axial accelerometers are widely used for automatic cattle behavior recognition under practical farm conditions [4], [5], [6], [7], [8], [9], [10], [11]. However, robust deployment remains constrained by two issues: limited computational resources on wearable devices and severe class imbalance in real-world behavioral data [12], [13].

Resource-constrained devices limit the deployment of computationally intensive deep learning models despite their strong predictive performance [1], [14], [15], [16]. Lightweight machine learning models using engineered statistical features are therefore more feasible for collar-based, real-time monitoring [1], [17], [18], [19]. At the same time, imbalanced datasets bias models toward dominant classes, reducing sensitivity to rare but important behaviors such as salt licking [5], [20], [21], [22], [23], [24]. This problem is exacerbated when oversampling methods such as SMOTE are applied improperly, causing data leakage [18], [25], [26].

Although ensemble learning and decision-level fusion can improve classification robustness [27], [28], their use in accelerometer-based cattle monitoring remains limited [11], [24], [26]. This study, therefore, proposes a lightweight fusion framework that combines compact statistical feature extraction, leakage-free SMOTE, and soft-voting fusion of Extra Trees, XGBoost, and CatBoost to improve minority-class recognition while preserving computational efficiency [12], [27], [29], [30].

2. Method

The proposed framework consists of six sequential phases: (1) raw data acquisition, (2) sliding-window segmentation, (3) time-domain statistical feature extraction, (4) stratified data partitioning with SMOTE augmentation, (5) benchmarking of lightweight machine learning models, and (6) decision-level fusion for final prediction (Fig. 1). This pipeline was designed to ensure computational efficiency, robust evaluation, and strong classification performance for edge deployment. In addition to overall accuracy, the study emphasizes two evaluation targets: global robustness, measured using Macro-F1, and rare-event sensitivity, measured using Recall_SLT.

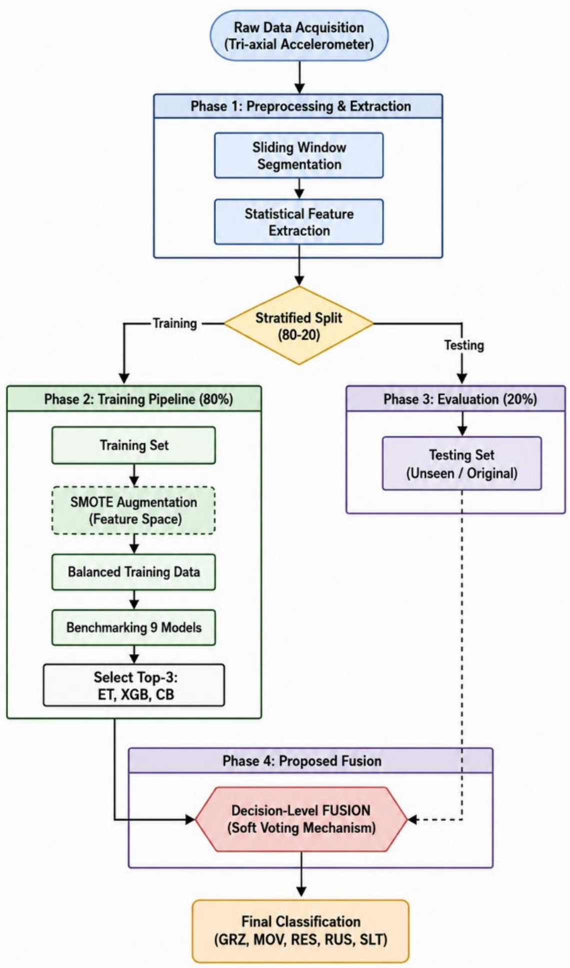

Fig. 1. Diagram overflow.

2.1. Dataset Overview

This study uses the Japanese Black Beef Cow Behavior Classification Dataset [31], which contains tri-axial accelerometer recordings annotated into 13 behavior classes. Data were collected from six Japanese Black Beef cows at the Shinshu University farm in Nagano, Japan, on June 12, 2020, using a 16-bit ±2 g Kionix KX122-1037 accelerometer mounted on the neck at 25 Hz. From 567 min of recordings, 197 min of high-quality labeled data were retained after removing inconsistent segments. Annotation required approximately 69 person-hours. The dataset is limited to a single farm, one day, and six animals.

Classes unsuitable for reliable modeling were excluded. ETC and BLN were removed due to unreliable annotation, while low-frequency classes (DRN, URI, ATT, ESC, BMN, REL, FES, and LCK) were excluded because of insufficient observations [18], [26]. The retained classes were RES, RUS, GRZ, MOV, and SLT, totaling 282,029 samples. RES was the majority class (53.23%), whereas SLT was the minority (3.85%), giving an imbalance ratio of approximately 1:13.8. The original distribution is shown in Fig. 2.

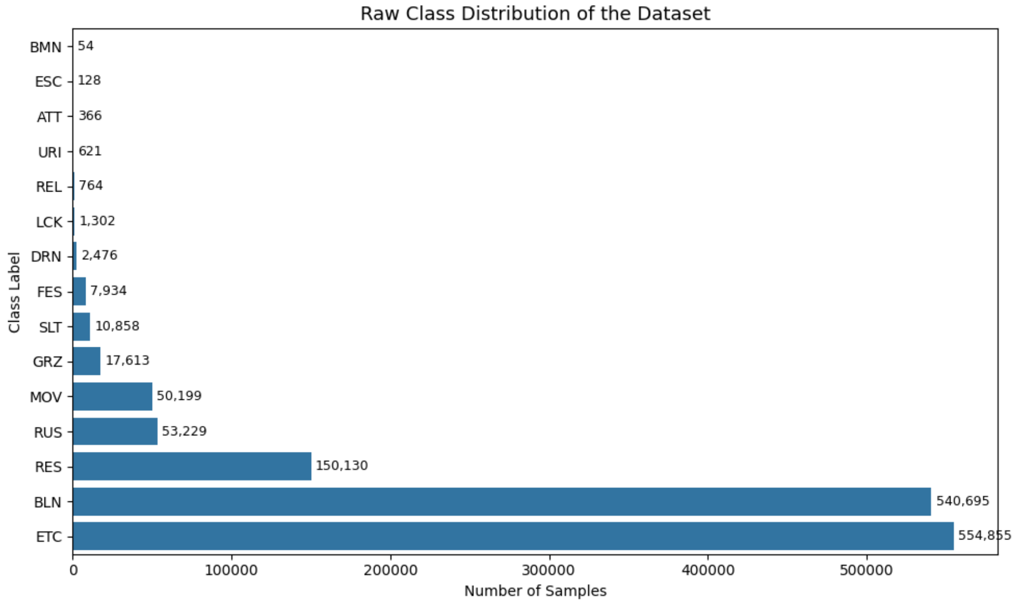

Fig. 2. Data Distribution.

2.2. Data Selection

Classes unsuitable for reliable modeling were excluded to improve label reliability and learning stability. ETC and BLN were removed because of ambiguous or missing annotations, while low-frequency classes (DRN, URI, ATT, ESC, BMN, REL, FES, and LCK) were excluded due to insufficient observations [18], [26]. The retained classes were RES, RUS, GRZ, MOV, and SLT, yielding 282,029 samples (Table 1). RES was the majority class (53.23%) and SLT the minority (3.85%), corresponding to an imbalance ratio of approximately 1:13.8.

Table 1. Selected Class.

Label	Description	Amount	Percentage
RES	Resting (Standing)	150.130	53.23 %
MOV	Moving	50.199	17.79 %
GRZ	Grazing	17.613	6.24 %
RUS	Ruminating (Standing)	53.229	18.87 %
SLT	Salt Licking	10.858	3.85%

2.3. Data Preprocessing and Segmentation

Raw tri-axial accelerometer signals were Z-normalized to reduce inter-subject and sensor-scale variability. The signals were then segmented using a sliding window of 64 samples with 50% overlap (Fig. 3). At 25 Hz, each window represents 2.56 s, providing a balance between temporal sensitivity and feature stability [32]. This process produced 8,805 valid windows (Fig. 4), while preserving class imbalance, with RES = 4,679 windows (53.1%) and SLT = 339 windows (3.8%).

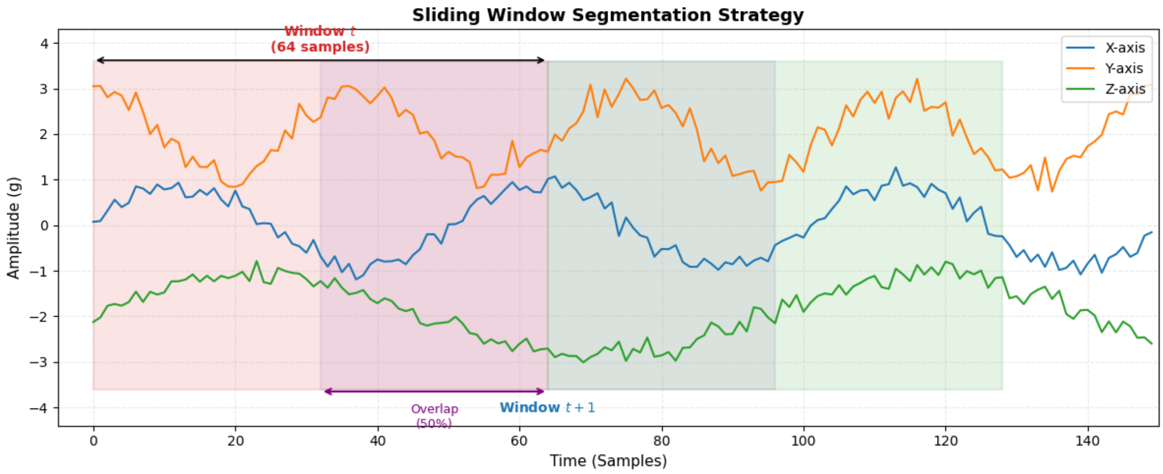

Fig. 3. Sliding window segmentation strategy with a window size of 64 samples and 50% overlap.

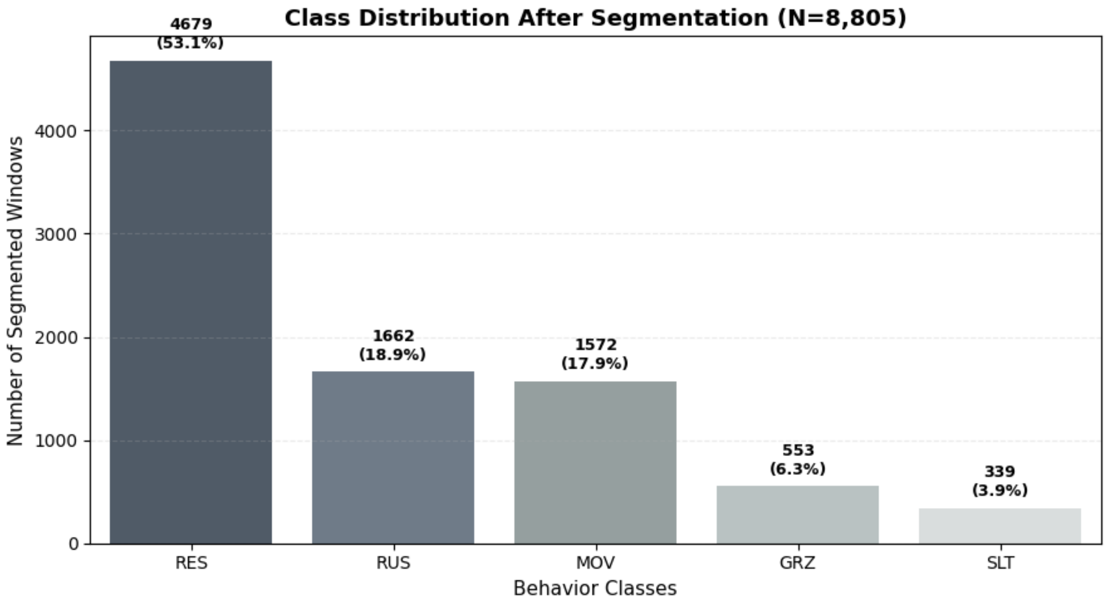

Fig. 4. Distribution of Segmented Data Samples Across Five Target Behavior Classes.

2.4. Statistical Feature Extraction

Each segmented window was converted into a statistical feature vector for lightweight machine learning. In addition to the raw X, Y, and Z acceleration signals, the Signal Magnitude Vector (SMV) was computed as an orientation-independent measure of movement intensity [4], [33], [34], [35]:

$$SMV_t = \sqrt{x_t^2 + y_t^2 + z_t^2} \tag{1}$$

where x_t, y_t, and $z_t$denote the acceleration values along the x-, y-, and z-axes at time t, respectively. The resulting SMV_t provides an orientation-independent measure of movement intensity and is commonly used as a discriminative feature for human activity recognition and motion analysis.

2.5. Time-Domain Statistical Features

Features extracted from X, Y, Z, and SMV produced an initial 36-feature representation (Table 2). Ranked subsets of 10, 14, 18, 21, 23, 25, and 36 features were then evaluated. Because performance improved up to 21 features and plateaued thereafter, the final framework used a compact 21-feature representation.

Table 2. Performance Comparison Across Feature Subsets.

k Features	Best Model	Accuracy	Macro-F1
10	Fusion	0.9284	0.9366
14	XGBoost	0.9296	0.9350
18	Fusion	0.9347	0.9373
21	Extra Trees	0.9341	**0.9413**
23	Extra Trees	0.9341	0.9413
25	Extra Trees	0.9341	0.9413
36	Extra Trees	0.9341	0.9413

2.6. Class Imbalance Handling Strategy

The dataset exhibits a strong imbalance (~13.8:1), which biases classifiers toward majority classes. To address this, SMOTE was applied in the feature space [23], as follows

$$S = A + \lambda(B - A) \tag{2}$$

where $\lambda \in [0,1]$ is a random interpolation factor. To prevent data leakage, the dataset was split into 80% training and 20% testing subsets using stratified sampling, and SMOTE was applied only to the training subset [18], [25], [26]. Before resampling, the training set contained 7,044 samples, including 3,743 RES and 271 SLT samples. After SMOTE, the dataset increased to 18,715 samples, with all five classes balanced at 3,743 samples each, while the testing set remained unchanged at 1,761 samples.

2.7. Train Baseline Model with Lightweight Machine Learning Classifiers

Because livestock monitoring systems operate in resource-constrained environments, this study focuses on lightweight machine learning classifiers rather than deep learning models [14], [15], [36]. The evaluated models include k-NN, Naive Bayes, SVM, Random Forest, Extra Trees, Gradient Boosting, XGBoost, and CatBoost [22], [37], [38], [39], [40], [41]. Based on preliminary benchmarking, Extra Trees, XGBoost, and CatBoost were selected as the base learners for fusion because they provided the best balance between predictive performance and efficiency. Hyperparameter tuning was conducted using Grid Search with 5-fold stratified cross-validation, with SMOTE applied only within the training folds.

2.8. Multi-Model Fusion Strategy

To improve robustness, a decision-level soft-voting strategy was used [42]. Each base classifier outputs posterior probabilities for all target classes, and the final prediction is obtained from their weighted average:

$$\hat{y} = argmax_{c \in C}\left(\sum_{i=1}^{N} w_i \cdot P_i(c \mid x)\right) \tag{3}$$

where N denotes the number of base models, w_i represents the weight assigned to the $i - th$ classifier, and $P_i(c \mid x)$ is the predicted probability of class c given input sample x. Equal weights were initially assigned to all models to maintain fairness and prevent bias toward any single classifier.

The fusion combines Extra Trees, XGBoost, and CatBoost, with optimal weights (0.40, 0.40, 0.20) obtained via grid search (Table 3). This improves robustness by reducing model-specific bias and variance while maintaining computational efficiency.

Table 3. Fusion Weight Optimization Results.

Weight Config (ET, XGB, CAT)	Accuracy	Macro-F1	Selection Status
(0.33, 0.33, 0.34)	0.9387	0.9404	
(0.40, 0.30, 0.30)	0.9398	0.9424	Baseline
(0.50, 0.30, 0.20)	0.9405	0.9426	
(0.40, 0.40, 0.20)	0.9421	0.9447	
(0.30, 0.50, 0.20)	0.9412	0.9431	Selected
(0.40, 0.30, 0.30)	0.9392	0.9420	
(0.30, 0.30, 0.40)	0.9356	0.9391	

2.9. Evaluation Protocol and Performance Metrics

The segmented dataset was divided into training (80%) and testing (20%) subsets using stratified sampling. Hyperparameter tuning was performed using 5-fold stratified cross-validation, with all preprocessing steps that could affect data distribution, especially SMOTE, restricted to the training data. To ensure fair evaluation in the presence of class imbalance, the primary metric was Macro-F1, while Recall_SLT was monitored to assess sensitivity to the minority class. The evaluation metrics are defined as follows [36], [38], [43]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$\text{Precision}_{\text{macro}} = \frac{1}{|C|}\sum_{c\in C}\frac{TP_c}{TP_c+FP_c} \tag{5}$$

$$\text{Recall}_{\text{macro}} = \frac{1}{|C|}\sum_{c\in C}\frac{TP_c}{TP_c+FN_c} \tag{6}$$

$$\text{F1-Score}_{\text{macro}} = \frac{1}{|C|}\sum_{c\in C} 2 \times \frac{\text{Precision}_c \times \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \tag{7}$$

where TP, TN, FP, and FN denote True Positives, True Negatives, False Positives, and False Negatives, respectively, and $|C|=5$ represents the number of behavior classes. All experiments were conducted in Python 3.12 using Google Colab with standard libraries including Scikit-learn, XGBoost, and CatBoost. Since the proposed framework relies on lightweight models and low-dimensional statistical features, GPU acceleration was not required.

3. Results and Discussion

3.1. Effect of Data Balancing Strategy

After preprocessing, the training data remained imbalanced, with RES = 3,743 samples (53.1%) and SLT = 271 samples (3.8%), corresponding to an imbalance ratio of approximately 1:13.8 (Table 4). Applying SMOTE only to the training subset increased the training size from 7,044 to 18,715 samples, balancing all five classes to 3,743 samples each, while leaving the test set unchanged at 1,761 samples. This leakage-free strategy improved minority-class learning without contaminating evaluation. Its effectiveness is reflected in the strong minority performance on the unseen test set, where SLT recall approached 0.99, indicating improved rare-event sensitivity without obvious overfitting to synthetic samples.

Table 4. Class Distribution Before and After SMOTE.

Class	Before	After
GRZ	442	3743
MOV	1258	3743
RES	3743	3743
RUS	1330	3743
SLT	271	3743

3.2. Benchmarking of Lightweight Classifiers

Comparative evaluation on the unseen test set showed that tree-based ensemble models consistently outperformed traditional classifiers. The best individual model was XGBoost, achieving 93.70% accuracy and 0.9421 Macro-F1, followed by Extra Trees with 93.24% accuracy and 0.9370 Macro-F1 (Table 5). CatBoost and Random Forest also performed competitively, whereas SVM, k-NN, and especially Naive Bayes were less effective. These findings indicate that tree-based methods better captured the nonlinear structure of the 21-feature space [44], [45].

Table 5. Performance Benchmarking of Lightweight Classifiers on the Unseen Testing Set.

Model	Accuracy	F1-Macro	Precision	Recall	Training Time (s)
XGBoost	0.936968	0.942052	0.930142	0.955061	1.491307
Extra Trees	0.932425	0.937005	0.924890	0.950035	2.399477
Random Forest	0.927314	0.930840	0.916155	0.947334	8.550886
CatBoost	0.921068	0.927438	0.909419	0.948232	132.533954
Gradient Boosting	0.907439	0.914121	0.893596	0.938440	102.098428
KNN	0.897785	0.902652	0.874589	0.937597	0.006096
SVM	0.850653	0.867130	0.836564	0.910188	13.805009
Naive Bayes	0.651902	0.685319	0.702256	0.778782	0.013022

Table 6 presents the classification performance of the proposed model across the five behavior classes. The model achieved an overall accuracy of 94%, with weighted-average precision, recall, and F1-score all reaching 0.94, indicating balanced and reliable performance on the test dataset. The macro-average F1-score of 0.94 further demonstrates consistent classification capability across all behavior classes despite differences in class distribution.

Table 6. Detailed Classification Report of the Best Lightweight Model (XGBoost).

Behavior Class	Precision	Recall	F1-Score	Support
GRZ	0.93	0.99	0.96	111
MOV	0.86	0.91	0.89	314
RES	0.97	0.93	0.95	936
RUS	0.91	0.96	0.94	332
SLT	0.97	0.99	0.98	68
Accuracy			0.94	1761
Macro Avg	0.93	0.96	0.94	1761
Weighted Avg	0.94	0.94	0.94	1761

Among the individual classes, SLT achieved the highest performance with a precision of 0.97, recall of 0.99, and F1-score of 0.98, followed closely by GRZ with an F1-score of 0.96. The RES class, which contained the largest number of samples (936), was classified accurately with a precision of 0.97 and an F1-score of 0.95, demonstrating the robustness of the proposed model on the majority class. Although the MOV class obtained the lowest F1-score (0.89), its recall of 0.91 indicates that most movement instances were correctly identified. Similarly, the RUS class achieved an F1-score of 0.94, indicating a good balance between precision (0.91) and recall (0.96). Overall, these results confirm that the proposed classifier provides robust and consistent performance across all behavioral categories.

3.3. Evaluation of the Proposed Fusion Framework

Using optimized soft voting, the fusion model combining Extra Trees, XGBoost, and CatBoost achieved the best overall performance, with 94.21% accuracy and 0.9447 Macro-F1. Compared with the best single model, XGBoost (93.70% accuracy, 0.9421 Macro-F1), this corresponds to gains of 0.51 percentage points in accuracy and 0.0026 in Macro-F1. Although modest, the improvement was statistically significant according to the Wilcoxon signed-rank test (p=0.0326), indicating that it was consistent rather than random. The limited numerical gain suggests a ceiling effect because the individual tree-based models already provided strong baselines; however, fusion still improved prediction stability and reduced model-specific errors under imbalanced conditions.

Table 7 presents the Precision–Recall Area Under the Curve (PR-AUC) and Recall at 90% Precision (Recall@90%Prec) for each behavior class. The proposed model demonstrates excellent precision–recall performance, with all classes achieving PR-AUC values above 0.96, indicating strong discriminative capability even in the presence of class imbalance. The SLT class achieved the highest performance, with a PR-AUC of 0.9990 and a Recall@90% Precision of 1.0000, indicating that all SLT instances were correctly identified while maintaining at least 90% precision. Similarly, the GRZ class achieved a PR-AUC of 0.9912 and a Recall@90% Precision of 0.9797, followed by the RES class, which obtained a PR-AUC of 0.9854 and a Recall@90% Precision of 0.9909 despite containing the largest number of samples (936).

Table 7. PR-AUC and Recall@90% Precision for Each Class.

Class	PR-AUC	Support	Recall@90%Prec.
RES	0.9854	936	0.9909
RUS	0.9638	332	0.8853
GRZ	0.9912	111	0.9797
MOV	0.9812	314	0.9639
SLT	0.9990	68	1.0000

The MOV class also exhibited strong performance, achieving a PR-AUC of 0.9812 with a Recall@90% Precision of 0.9639, demonstrating that the model retained high recall while enforcing a strict precision threshold. Among all classes, RUS recorded the lowest PR-AUC (0.9638) and Recall@90% Precision (0.8853), suggesting that this behavior is relatively more challenging to distinguish from the other classes. Nevertheless, the consistently high PR-AUC values across all categories confirm that the proposed fusion model delivers reliable classification performance while maintaining high precision without substantially sacrificing recall.

Model calibration was assessed using Expected Calibration Error (ECE), which yielded 0.1903 (Fig. 5). This indicates a moderate discrepancy between predicted probabilities and observed outcomes, as is common in tree-based ensembles, but it does not materially reduce the model's practical usefulness.

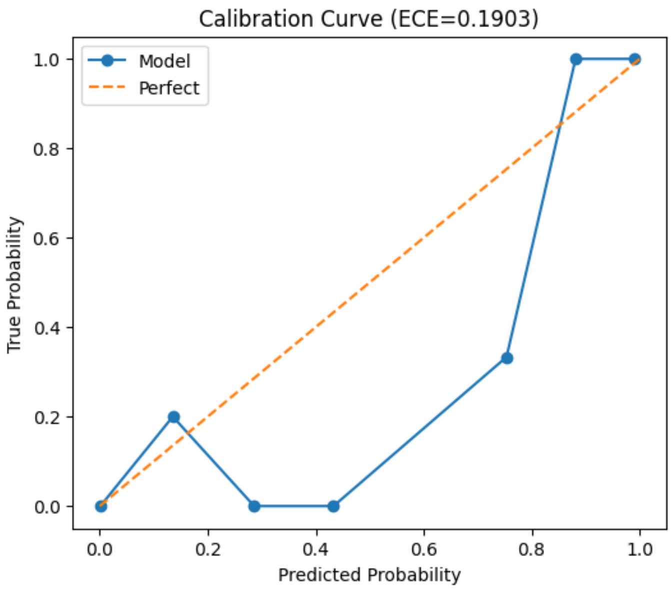

Fig. 5. Calibration Curve (Reliability Diagram) of the Fusion Model.

Table 8 summarizes the detailed classification performance of the proposed fusion model across the five cattle behavior classes. The model achieved an **overall accuracy of 93%**, with a **weighted-average precision of 0.94**, **recall of 0.93**, and **F1-score of 0.94**, indicating robust and balanced classification performance on the test dataset. Furthermore, the **macro-average F1-score of 0.94** demonstrates that the model maintains consistent performance across all behavior categories despite differences in class frequencies.

Table 8. Detailed Classification Report for Fusion Model.

Behavior Class	Precision	Recall	F1-Score	Support
Grazing (GRZ)	0.93	0.99	0.96	111
Moving (MOV)	0.87	0.91	0.89	314
Resting (RES)	0.97	0.93	0.95	936
Ruminating (RUS)	0.91	0.96	0.93	332
Salt Licking (SLT)	0.95	0.98	0.97	68
Accuracy	0.93	0.93	0.93	0.93
Macro Avg	0.93	0.95	0.94	1761
Weighted Avg	0.94	0.93	0.94	1761

Among the individual classes, Grazing (GRZ) achieved the highest recall (0.99) and an F1-score of 0.96, indicating that nearly all grazing instances were correctly identified. Salt Licking (SLT) also exhibited excellent performance with a precision of 0.95, recall of 0.98, and F1-score of 0.97, despite being the smallest class with only 68 samples. The Resting (RES) class, which contained the largest number of samples (936), achieved a precision of 0.97, recall of 0.93, and F1-score of 0.95, demonstrating the robustness of the proposed model on the dominant class. Ruminating (RUS) achieved a balanced performance with an F1-score of 0.93, while Moving (MOV) recorded the lowest F1-score (0.89), suggesting that movement behavior exhibits greater overlap with other activity patterns. Nevertheless, the consistently high precision, recall, and F1-scores across all classes confirm the effectiveness of the proposed decision-level fusion model for multiclass cattle behavior recognition.

Error analysis showed that most misclassifications occurred between kinematically similar classes: 36 RES samples were misclassified as MOV, and 27 RUS samples as RES. In contrast, SLT showed minimal confusion, with only 1 misclassification among 68 test samples, suggesting that this behavior occupied a more distinct region of the feature space. Fig. 6 shows the Precision–Recall (PR) curve for the SLT class.

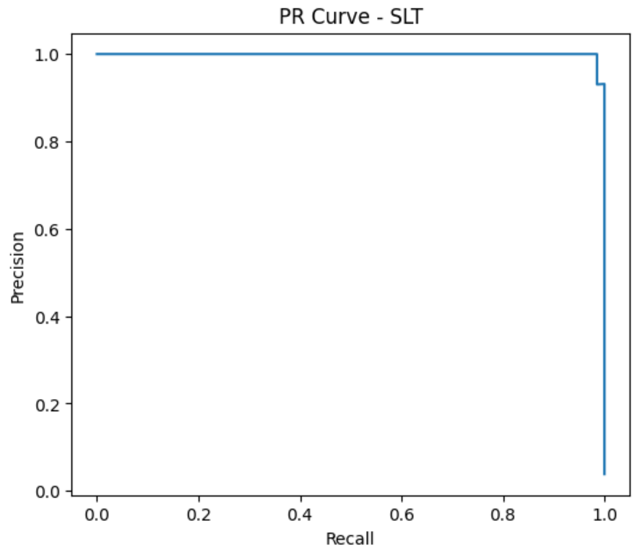

Fig. 6. Precision–Recall Curve for the SLT (Salt Licking) Class.

The curve remains close to the upper-right corner, indicating excellent classification performance with consistently high precision across nearly the entire recall range. This result is consistent with the PR-AUC of 0.9990, demonstrating the proposed fusion model's ability to accurately identify SLT behavior with minimal false positives. The main misclassifications occur between Moving (MOV), Resting (RES), and Ruminating (RUS), while Grazing (GRZ) and Salt Licking (SLT) are recognized with very few errors, demonstrating the effectiveness of the proposed fusion approach (Fig. 7).

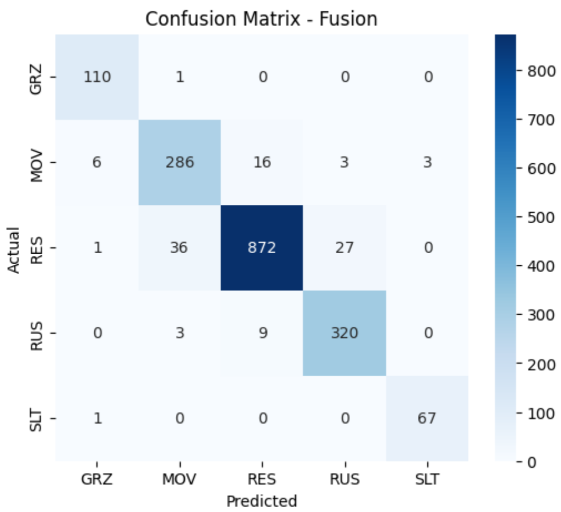

Fig. 7. Confusion Matrix of Fusion Model.

The results demonstrate that the proposed lightweight decision-level fusion framework effectively balances global robustness and minority-class sensitivity while remaining suitable for edge-oriented livestock monitoring.

4. Conclusion

This study proposes a lightweight decision-level fusion framework for cattle behavior classification under imbalanced conditions using tri-axial accelerometer data. The fusion model achieved 94.21% accuracy and 0.9447 Macro-F1, outperforming XGBoost by 0.51 percentage points (p=0.0326), indicating consistent improvement. It also demonstrated strong minority-class performance, where SLT (3.8%) achieved PR-AUC = 0.9990, recall = 0.9853, and precision = 0.9571. The integration of Extra Trees, XGBoost, and CatBoost via optimized soft voting (0.40, 0.40, 0.20) improved robustness and reduced model-specific errors while maintaining balanced performance. Using a compact 21-feature representation, the model achieved efficient inference (~22 ms/sample), supporting real-time deployment in resource-constrained environments. Limitations include the use of a single-farm, single-day dataset and the absence of cross-subject validation due to limited minority samples. Future work will focus on broader validation, embedded deployment, and multimodal integration. Overall, the framework provides an effective and practical solution for robust monitoring of cattle behavior, particularly for rare-event detection.

Acknowledgment

The authors gratefully acknowledge the financial support provided by **Universitas Muhammadiyah Malang** for this research. We also extend our sincere appreciation to **Hiroyuki Ito et al.** for making the *Japanese Black Beef Cow Behavior Classification Dataset* publicly available on Zenodo, which was utilized as the primary data source in this study.

Declarations

Author contribution. P.N. Sabri: Conceptualization, Investigation, Data Curation, Visualization, Writing – Original Draft, Writing – Review & Editing; N. Nurhafizhah: Conceptualization, Investigation, Data Curation, Visualization, Writing – Original Draft, Writing – Review & Editing; A. Faruq and M.I. Pedana: Conceptualization, Methodology, Validation; A.F.H. Soegiharto: Writing – Review & Editing, Supervision. All authors had approved the final version.

Funding statement. This research is supported by Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Muhammadiyah Malang.

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

[1] D. Lovarelli *et al.*, "Development of a New Wearable 3D Sensor Node and Innovative Open Classification System for Dairy Cows' Behavior," *Animals*, vol. 12, no. 11, p. 1447, Jun. 2022, doi: 10.3390/ani12111447.

[2] K. El Moutaouakil and N. Falih, "A comparative study on time series data-based artificial intelligence approaches for classifying cattle feeding behavior," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 1, p. 324, Jan. 2024, doi: 10.11591/ijeecs.v33.i1.pp324-332.

[3] D. Pavlovic *et al.*, "Classification of Cattle Behaviours Using Neck-Mounted Accelerometer-Equipped Collars and Convolutional Neural Networks," *Sensors*, vol. 21, no. 12, p. 4050, Jun. 2021, doi: 10.3390/s21124050.

[4] K. Pongsanun, T. Apirak, L. Aekaluck, C. Sanphet, and I. Chaidate, "Classification of individual dairy cow behaviors using accelerometer, gyroscope, and integrated sensor models," *BMC Vet. Res.*, vol. 21, no. 1, p. 677, Nov. 2025, doi: 10.1186/s12917-025-05092-1.

[5] Á. Michelena, E. Jove, Ó. Fontenla-Romero, and J.-L. Calvo-Rolle, "Evaluation of One-Class Techniques for Early Estrus Detection on Galician Intensive Dairy Cow Farm Based on Behavioral Data From Activity Collars," *ADCAIJ Adv. Distrib. Comput. Artif. Intell. J.*, vol. 13, p. e32508, Dec. 2024, doi: 10.14201/adcaij.32508.

[6] D.-N. Tran, V.-M. Do, M.-T. Vi, and D.-T. Tran, "Robust Real-Time Analysis of Cow Behaviors Using Accelerometer Sensors and Decision Trees with Short Data Windows and Misalignment Compensation," *Comput. Mater. Contin.*, vol. 83, no. 2, pp. 2525–2553, Apr. 2025, doi: 10.32604/cmc.2025.062590.

[7] M. W. Iqbal, I. Draganova, P. C. H. Morel, and S. T. Morris, "Validation of an Accelerometer Sensor-Based Collar for Monitoring Grazing and Rumination Behaviours in Grazing Dairy Cows," *Animals*, vol. 11, no. 9, p. 2724, Sep. 2021, doi: 10.3390/ani11092724.

[8] V. Mehlin, S. Schacht, and C. Lanquillon, "Towards energy-efficient Deep Learning: An overview of energy-efficient approaches along the Deep Learning Lifecycle," pp. 1–18, Feb. 2023, Accessed: Jun. 07, 2026. [Online]. Available: http://arxiv.org/abs/2303.01980.

[9] K. El Moutaouakil and N. Falih, "A design of a smart farm system for cattle monitoring," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 32, no. 2, p. 857, Nov. 2023, doi: 10.11591/ijeecs.v32.i2.pp857-864.

[10] V. Bloch, L. Frondelius, C. Arcidiacono, M. Mancino, and M. Pastell, "Development and Analysis of a CNN- and Transfer-Learning-Based Classification Model for Automated Dairy Cow Feeding Behavior Recognition from Accelerometer Data," *Sensors*, vol. 23, no. 5, p. 2611, Feb. 2023, doi: 10.3390/s23052611.

[11] Z. Yu, Y. Han, L. Cha, S. Chen, Z. Wang, and Y. Zhang, "Design of an intelligent wearable device for real-time cattle health monitoring," *Front. Robot. AI*, vol. 11, p. 1441960, Nov. 2024, doi: 10.3389/frobt.2024.1441960.

[12] G. Ye and R. Yu, "Spatiotemporal Mapping of Grazing Livestock Behaviours Using Machine Learning Algorithms," *Sensors*, vol. 25, no. 15, p. 4561, Jul. 2025, doi: 10.3390/s25154561.

[13] C. Li *et al.*, "Data Augmentation for Inertial Sensor Data in CNNs for Cattle Behavior Classification," *IEEE Sensors Lett.*, vol. 5, no. 11, pp. 1–4, Nov. 2021, doi: 10.1109/LSENS.2021.3119056.

[14] H.-I. Liu *et al.*, "Lightweight Deep Learning for Resource-Constrained Environments: A Survey," *ACM Comput. Surv.*, vol. 56, no. 10, pp. 1–42, Oct. 2024, doi: 10.1145/3657282.

[15] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep Learning for Sensor-based Human Activity Recognition," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–40, May 2022, doi: 10.1145/3447744.

[16] M. M. H. Shuvo, S. K. Islam, J. Cheng, and B. I. Morshed, "Efficient Acceleration of Deep Learning Inference on Resource-Constrained Edge Devices: A Review," *Proc. IEEE*, vol. 111, no. 1, pp. 42–91, Jan. 2023, doi: 10.1109/JPROC.2022.3226481.

[17] B. Hoffman *et al.*, "A benchmark for computational analysis of animal behavior, using animal-borne tags," *Mov. Ecol.*, vol. 12, no. 1, p. 78, Dec. 2024, doi: 10.1186/s40462-024-00511-8.

[18] H. Bragança, J. G. Colonna, H. A. B. F. Oliveira, and E. Souto, "How Validation Methodology Influences Human Activity Recognition Mobile Systems," *Sensors*, vol. 22, no. 6, p. 2360, Mar. 2022, doi: 10.3390/s22062360.

[19] D. Pavlovic *et al.*, "Behavioural Classification of Cattle Using Neck-Mounted Accelerometer-Equipped Collars," *Sensors*, vol. 22, no. 6, p. 2323, Mar. 2022, doi: 10.3390/s22062323.

[20] L. Schmeling *et al.*, "Training and Validating a Machine Learning Model for the Sensor-Based Monitoring of Lying Behavior in Dairy Cows on Pasture and in the Barn," *Animals*, vol. 11, no. 9, p. 2660, Sep. 2021, doi: 10.3390/ani11092660.

[21] S. Mahato, H. Bi, and S. Neethirajan, "Dairy DigiD: a keypoint-based deep learning system for classifying dairy cattle by physiological and reproductive status," *Front. Artif. Intell.*, vol. 8, p. 1545247, Aug. 2025, doi: 10.3389/frai.2025.1545247.

[22] F. A. Nishat, M. F. Mridha, I. Mahmud, M. Alfarhood, M. Safran, and D. Che, "Enhancing Typhoid Fever Diagnosis Based on Clinical Data Using a Lightweight Machine Learning Metamodel," *Diagnostics*, vol. 15, no. 5, p. 562, Feb. 2025, doi: 10.3390/diagnostics15050562.

[23] H. Mansourifar and W. Shi, "Deep Synthetic Minority Over-Sampling Technique," pp. 1–9, Mar. 2020, Accessed: Jun. 07, 2026. [Online]. Available: http://arxiv.org/abs/2003.09788.

[24] X. Zhang *et al.*, "Machine learning for phase prediction of high entropy carbide ceramics from imbalanced data," *npj Comput. Mater.*, vol. 12, no. 1, p. 22, Jan. 2026, doi: 10.1038/s41524-025-01873-2.

[25] H. James, C. Rial, J. Giordano, and D. Erickson, "Advancing standardized cattle behavior classification with a random forest model evaluated across diverse datasets," *Front. Anim. Sci.*, vol. 7, p. 1676504, Jan. 2026, doi: 10.3389/fanim.2026.1676504.

[26] A. Tello, V. Degeler, and A. Lazovik, "Too Good To Be True: accuracy overestimation in (re)current practices for Human Activity Recognition," in *2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, IEEE, Mar. 2024, pp. 511–517. doi: 10.1109/PerComWorkshops59983.2024.10503465.

[27] A. Rojarath and W. Songpan, "Probability-Weighted Voting Ensemble Learning for Classification Model," *J. Adv. Inf. Technol.*, vol. 11, no. 4, pp. 217–227, Nov. 2020, doi: 10.12720/jait.11.4.217-227.

[28] F. Tian, J. Wang, B. Xiong, L. Jiang, Z. Song, and F. Li, "Real-Time Behavioral Recognition in Dairy Cows Based on Geomagnetism and Acceleration Information," *IEEE Access*, vol. 9, pp. 109497–109509, 2021, doi: 10.1109/ACCESS.2021.3099212.

[29] R. Maurya, S. Mahapatra, and L. Rajput, "A Lightweight Meta-Ensemble Approach for Plant Disease Detection Suitable for IoT-Based Environments," *IEEE Access*, vol. 12, pp. 28096–28108, 2024, doi: 10.1109/ACCESS.2024.3367443.

[30] G. Menghani, "Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better," *ACM Comput. Surv.*, vol. 55, no. 12, pp. 1–37, Dec. 2023, doi: 10.1145/3578938.

[31] H. Ito *et al.*, "Japanese Black Beef Cow Behavior Classification Dataset," Zenodo. Accessed: Jun. 07, 2026. [Online]. Available: https://zenodo.org/records/5849025.

[32] Y. Li *et al.*, "Classification and Analysis of Multiple Cattle Unitary Behaviors and Movements Based on Machine Learning Methods," *Animals*, vol. 12, no. 9, p. 1060, Apr. 2022, doi: 10.3390/ani12091060.

[33] T. Mladenova *et al.*, "Evaluation of the Efficiency of Machine Learning Algorithms for Identification of Cattle Behavior Using Accelerometer and Gyroscope Data," *AgriEngineering*, vol. 6, no. 3, pp. 2179–2197, Jul. 2024, doi: 10.3390/agriengineering6030128.

[34] L. Ding *et al.*, "Predicting the Feed Intake of Cattle Based on Jaw Movement Using a Triaxial Accelerometer," *Agriculture*, vol. 12, no. 7, p. 899, Jun. 2022, doi: 10.3390/agriculture12070899.

[35] M. Fatima, O. Rehman, I. M. H. Rahman, A. Ajmal, and S. J. Park, "Towards Ensemble Feature Selection for Lightweight Intrusion Detection in Resource-Constrained IoT Devices," *Futur. Internet*, vol. 16, no. 10, p. 368, Oct. 2024, doi: 10.3390/fi16100368.

[36] T. Shahriar, "Comparative Analysis of Lightweight Deep Learning Models for Memory-Constrained Devices," pp. 1–22, Sep. 2025, Accessed: Jun. 10, 2026. [Online]. Available: http://arxiv.org/abs/2505.03303.

[37] D. Li *et al.*, "Optimized Machine Learning Models for Predicting Core Body Temperature in Dairy Cows: Enhancing Accuracy and Interpretability for Practical Livestock Management," *Animals*, vol. 14, no. 18, p. 2724, Sep. 2024, doi: 10.3390/ani14182724.

[38] S. R. Alve, M. Z. Mahmud, S. Islam, M. A. Chowdhury, and J. Islam, "Resource-Efficient Machine Learning Approaches for Multi-Class Threat Detection in IoT Environments," in *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*, IEEE, Jul. 2025, pp. 1–6. doi: 10.1109/QPAIN66474.2025.11171769.

[39] R. Ma, R. Chen, B. Liang, and X. Li, "A XGBoost-Based Prediction Method for Meat Sheep Transport Stress Using Wearable Photoelectric Sensors and Infrared Thermometry," *Sensors*, vol. 24, no. 23, p. 7826, Dec. 2024, doi: 10.3390/s24237826.

[40] S. Bergen *et al.*, "A review of supervised learning methods for classifying animal behavioural states from environmental features," *Methods Ecol. Evol.*, vol. 14, no. 1, pp. 189–202, Jan. 2023, doi: 10.1111/2041-210X.14019.

[41] K. E. ArunKumar *et al.*, "Predicting dry matter intake in cattle at scale using gradient boosting regression techniques and Gaussian process boosting regression with Shapley additive explanation explainable artificial intelligence, MLflow, and its containerization," *J. Anim. Sci.*, vol. 103, pp. 1–20, Jan. 2025, doi: 10.1093/jas/skaf041.

[42] S. W. A. Sherazi, J.-W. Bae, and J. Y. Lee, "A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome," *PLoS One*, vol. 16, no. 6, p. e0249338, Jun. 2021, doi: 10.1371/journal.pone.0249338.

[43] A. Elkelany, R. Ross, and S. Mckeever, "Identifying Subject Bias in WiFi-based Human Activity Recognition Evaluation Methods," in *The 32nd Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2024)*, 2024, pp. 301–313. Accessed: Jun. 10, 2026. [Online]. Available: https://ceur-ws.org/Vol-3910/aics2024_p78.

[44] L. S. AMBATI, O. El-Gayar, L. S. AMBATI, and O. El-Gayar, "Human Activity Recognition: A Comparison of Machine Learning Approaches," *J. Midwest Assoc. Inf. Syst.*, vol. 2021, no. 1, p. 4, Mar. 2021, Accessed: Jun. 12, 2026. [Online]. Available: https://aisel.aisnet.org/jmwais/vol2021/iss1/4.

[45] M. Gharagozloo, A. Amrani, K. Wittingstall, A. Hamilton-Wright, and D. Gris, "Machine Learning in Modeling of Mouse Behavior," *Front. Neurosci.*, vol. 15, p. 700253, Sep. 2021, doi: 10.3389/fnins.2021.700253.