Letters in Animal Biology

Journal homepage: www.liabjournal.com

# Identification of estrus cows through their vocalizations using Ensemble Machine Learning

H.M. Rohini [1,2], S. Prabhavathi [1], Doddabasappa Angadi [3]

[1] *Department of Electronics & Communication Engineering, RYMEC Ballari, Karnataka, India*
[2] *Department of Electronics & Communication Engineering, Central University of Karnataka, Kalaburgi, Karnataka, India*
[3] *Department of Mathematical and Computational Sciences, Sri Sathya Sai University for Human Excellence, Navanihal, Okali Post, Kamalapur, Kalaburagi, Karnataka, India*

***Article info***

Received: 09 June 2026
Received in revised form: 30 September 2026
Accepted: 01 October 2026
Published online: 05 October 2026

**Keywords**

Convolutional neural networks (CNN)
Mel-frequency cepstral coefficients (MFCC)
Acoustic waveform
Estrus audio
Non-estrus audio

* *Corresponding author:*
*H.M. Rohini*
*Email: rohinihalakatti@gmail.com*

***Reviewed by:***
*Consent to publish the name of the reviewers could not be obtained*

**Abstract**

Identification of estrus phase is important for effective reproductive management and to improve productivity of dairy cattle. Conventional estrus detection methods, including visual observation and activity-based monitoring, may require substantial labor and can miss estrus events, motivating automated and non-invasive approaches. This study proposes an Ensemble Machine Learning model that investigates cattle vocalizations as an acoustic indicator of estrus. Audio preprocessing included pre-emphasis filtering and data augmentation, followed by extraction of 40-dimensional Mel-frequency cepstral coefficient (MFCC) features. Convolutional neural network architectures, LeNet-5, VGG-16, and ResNet-50, were evaluated individually and integrated using a weighted prediction-averaging strategy. The proposed ensemble achieved an overall classification accuracy of 96.03%, with precision, recall, and F1-score of 97.3%, 95.5%, and 96.5%, respectively, for the estrus class. The results indicate that cattle vocalizations contain discriminative acoustic features connected with reproductive phase and that combining complementary deep-learning architectures can improve classification performance. Hence, the proposed approach provides a non-invasive basis for automated estrus monitoring and may support precision livestock farming.

*This is an open access article under the CC Attribution license (http://creativecommons.org/licenses/by/4.0/)*

## 1. Introduction

Timely and reliable identification of estrus is significant in dairy cattle reproductive management. Accurate identification of the estrus period supports appropriate breeding decisions, improves reproductive efficiency, and can contribute to farm productivity. Conventional detection methods commonly rely on visual observation of behavioral changes or activity monitoring. Visual observation is labor-intensive and may be affected by the duration and intensity of behavioral expression, whereas activity-based models cannot record all estrus-related behavioral or physiological changes. These limitations have encouraged the investigation of automated sensing techniques for continuous and non-invasive estrus monitoring.

Cattle vocalizations provide an additional behavioral signal that may contain information related to reproductive status, earlier studies noted the changes in the frequency and characteristics of cattle vocalizations during estrus. Deep learning-based frameworks that use convolutional neural networks (CNN) and noise-filtering pipelines to classify cattle vocalizations under farm conditions has been made possible by real-time livestock monitoring (Jung et al. 2021). Accuracy has notably increased compared with traditional observation through the implementation of techniques such as machine learning - back propagation neural network (BPNN) in conjunction with dual-channel acoustic tags to identify estrus cows based on vocal characteristics (Wang et al. 2023). Research conducted on farms in the Yogyakarta Special Region of Indonesia revealed that cows vocalize more frequently during the estrus period, which is linked to reproductive physiology (Adi et al. 2020). Hence, automated estrus and calving detection systems emphasize the significance of considering acoustic data with sensor modalities such as pedometers, accelerometers, radio-telemetry, and pressure detectors to enhance robustness and on-farm application (Alves dos Santos et al. 2022). Vocalization in dairy systems has revealed maternal behavior, stress, and state beyond reproduction; and as a function of vocalization, postpartum and cow-calf separation have established the emotional and welfare circumstances in dairy systems (Neave et al. 2024). Bovine Talk demonstrates the prospective role of vocal analytics for precise welfare monitoring by analyzing vocalizations associated with negative emotional states such as isolation (Gavojdian et al. 2024). During estrus or other stress conditions in Holstein and pasture-based dairy cows, an increase in vocal activity with distinct temporal and acoustic patterns have been observed, which supports the potential of these vocal cues in identifying estrus and behavioral assessment (Göncü et al. 2019). Multi-modal data fusion algorithms integrate audio with physiological and behavioral data to improve the classification and interpretation of cow vocalizations (Jobarteh et al. 2025). Assessment of vocal production in postpartum dairy cows plays a significant role in managing maternal stress because vocalizations change distinctly across the stages of estrus, non-estrus, and rest of the reproductive cycle (Green et al. 2021). Acoustic analysis with Mel-Frequency Cepstral Coefficients (MFCCs), formant features, ensemble classifiers, and modern deep-learning techniques are promising, particularly when they are integrated with Hybrid Deep Learning Network (HDLN) to address the limitations of conventional methods (Hiregoud et al 2025).

Though significant advances have been made in acoustic analysis, the current approaches have several limitations. The reliability of acoustic analysis under farm conditions declines due to the influence of environmental noise, animal behavior, management conditions, and recording variability on cattle vocalizations. Previous studies report the investigation of individual machine-learning or deep-learning models, whose performance largely depends on the uniqueness and quality of the available recordings and extracted acoustic features. Furthermore, the limited availability of public and well-annotated datasets of estrus-specific cattle vocalizations challenges the development of highly complex models, with the risk of overfitting due to insufficient training data. Therefore, there is a need for approaches that can exploit complementary representations learned by different model architectures while maintaining robustness when the available dataset is relatively small.

Ensemble learning provides a potential solution to this problem by combining predictions from multiple classifiers rather than relying on a single architecture. Different convolutional neural network architectures can learn complementary patterns from acoustic representations. LeNet-5 provides a relatively compact architecture capable of learning dominant local patterns, whereas ResNet-50 uses residual connections to support deeper feature learning and VGG-16 provides another deep convolutional representation, although its larger model complexity may be challenging when training data are limited. Combining these models through weighted prediction averaging may therefore reduce model-specific errors and provide a more stable classification decision than an individual classifier.

Accordingly, the objective of this study was to develop and evaluate an ensemble deep-learning framework based on LeNet-5, VGG-16, and ResNet-50 for the classification of estrus and non-estrus cattle vocalizations. The framework employs pre-emphasis filtering, audio augmentation, MFCC-based feature-extraction, individual CNN classification, and weighted ensemble prediction. Individual models and the ensemble learning models were evaluated using classification performance measures to investigate whether combining complementary CNN architectures can improve estrus identification. The expected contribution of this work is an acoustic, non-invasive approach for automated estrus identification that can potentially complement existing livestock monitoring systems. The working hypothesis is that estrus and non-estrus cattle vocalizations contain distinguishable acoustic patterns and that an ensemble of complementary CNN architectures can exploit these patterns more effectively than an individual model.

Cattle vocalization recordings considered in this model were retrieved from the publicly available dataset reported by Jung et al. (2021), which contains annotated vocal samples obtained for cattle vocalization categorization research. Acoustics were captured from individual cow by monitoring sensors enabled through microcontroller and microphone, monitoring devices were positioned at a height of 3 m in three different livestock facilities, and estrus call produced from cow was captured from 130 Korean cattle breeding (aged 12 to 40 months). The cattle sounds when engaged in other activities like feeding were collected and labeled as non-estrus call. Fig. 1 shows Flowchart of proposed ensemble machine learning for the identification of estrus cows through their vocalizations.

## 2. Materials and Methods

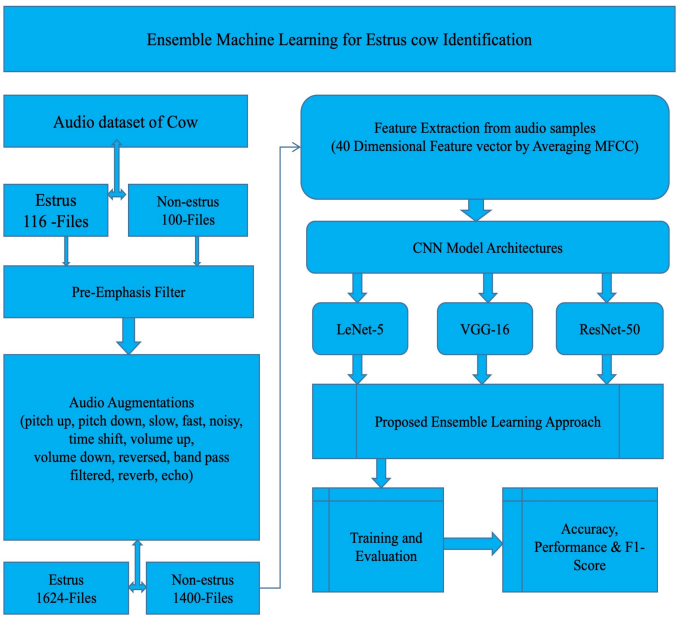


***Fig. 1.*** *Flowchart of proposed ensemble machine learning for estrus cows identification*

Ensemble Machine Learning for estrus cow identification uses the supplementary dataset reported by Jung et al. (2021). The original dataset consisted of 116 estrus and 100 non-estrus cattle vocalization recordings. To increase acoustic variability and improve model robustness, augmentation techniques including pitch shifting, time stretching, background-noise addition, time shifting, volume modification, audio reversal, band-pass filtering, reverberation, and echo were applied to the recordings. This process generated 1,624 estrus and 1,400 non-estrus audio samples.

To prevent data leakage, all augmented samples derived from the same original recording were treated as a single data group during dataset partitioning. Consequently, an original recording and any of its augmented versions were assigned entirely to either the training partition or the testing partition and were never distributed across both partitions. An 80:20 partition was chosen for training and evaluation. This grouped partitioning procedure ensured that the evaluation dataset did not contain augmented versions of recordings represented in the training set, thereby reducing the risk of overly optimistic performance caused by correlated samples. The training configuration included Adam optimizer, and the loss function included as categorical cross-entropy. Training was conducted for multiple epochs until convergence. The evaluation metrics considered were accuracy, precision, recall, and F1-score. Fig. 2 shows a spectrogram of estrus cow audio signals. A pre-emphasis filter is applied, which amplifies high-frequency components while attenuating lower frequencies.

A stretching factor greater than 1.0 speeds up the audio, making it sound faster, whereas a factor less than 1.0 slows it down. Fig. 3 shows a spectrogram of augmented estrus cow signal with time stretch of varying stretch factor. Pitch shifting is used to modify the perceived pitch of the audio signal without modifying its duration, the model is exposed to a wider range of audio variations, making it more robust to real-world conditions.

A positive shift increases the pitch, while a negative shift lowers it, thereby introducing diversity in the dataset and preventing overfitting to specific frequency ranges. Fig. 4 shows a spectrogram of augmented

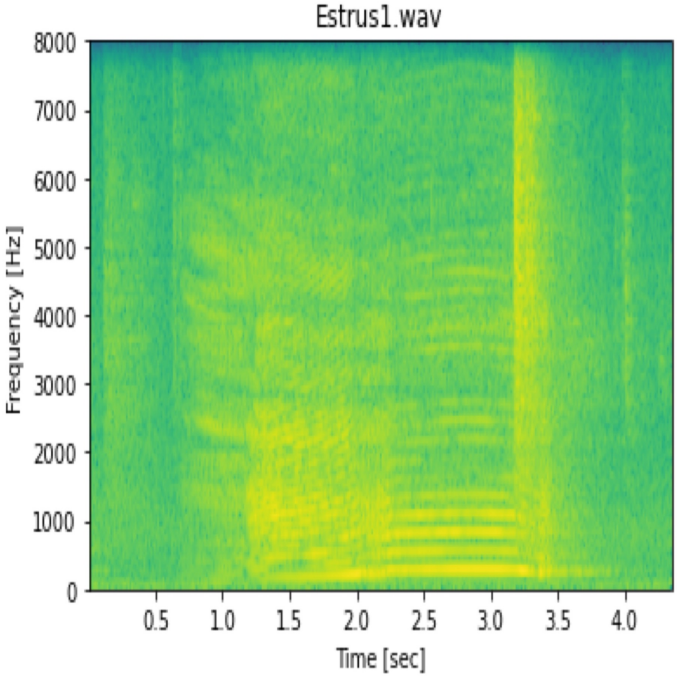


***Fig. 2.*** *Spectrogram of estrus cow audio signal*

estrus cow signal by applying negative and positive pitch shifts. The model is strengthened against changes in the environment by the addition of background noise, time shifting, and volume manipulation, among other methods. Additional improvements, such as band pass filtering, reverberation or echo, and audio reversal, further diversify the dataset and guarantee that the model can manage a range of acoustic conditions.

Fig. 5 shows a spectrogram of augmented estrus cow signal by adding background noise. After applying augmentation, 1624 estrus audio and 1400 non-estrus audio samples were obtained, where 80% of the data was used for training and 20% for evaluation.

The extract features function processes an audio file by loading it using librosa, ensuring it has a fixed length of 2.5 seconds through zero-padding or trimming, and extracting 40 MFCC (Mel-Frequency Cepstral Coefficients) features. To calculate the MFCCs, the function requires parameters like the sampling rate (sr = 22050), FFT window size (n_fft = 2048), and hop length (hop_length = 512). The MFCC values are averaged to provide a 40-dimensional feature vector which can then be used as input for machine-learning models. The proposed ensemble model uses a weighted average method to combine predictions from LeNet-5, VGG-16, and ResNet-50. The weighted averaging equation is:

$$P_{ensemble} = w_1P_1 + w_2P_2 + w_3P_3$$

where $P_1$, $P_2$, $P_3$ are predictions from LeNet-5, VGG-16, and ResNet-50, respectively and $w_1$, $w_2$, and $w_3$ are corresponding model weights

The output from all three individual models were combined based on their unique accuracy, stronger models contribute more to the final prediction. This approach increases classification accuracy, decreases overfitting, and improves generalization. In this study an ensemble model was built by integrating three CNN architectures. LeNet is a small, lightweight CNN designed for picture categorization. VGG is a more advanced design that optimizes feature-extraction by combining numerous convolutional layers. ResNet is a residual network which makes use of skip connections to avoid vanishing gradient concerns while also improving accuracy. The ensemble model built in this study aggregates the outputs of different architectures using weighted averaging, which assigns larger weights to models with better

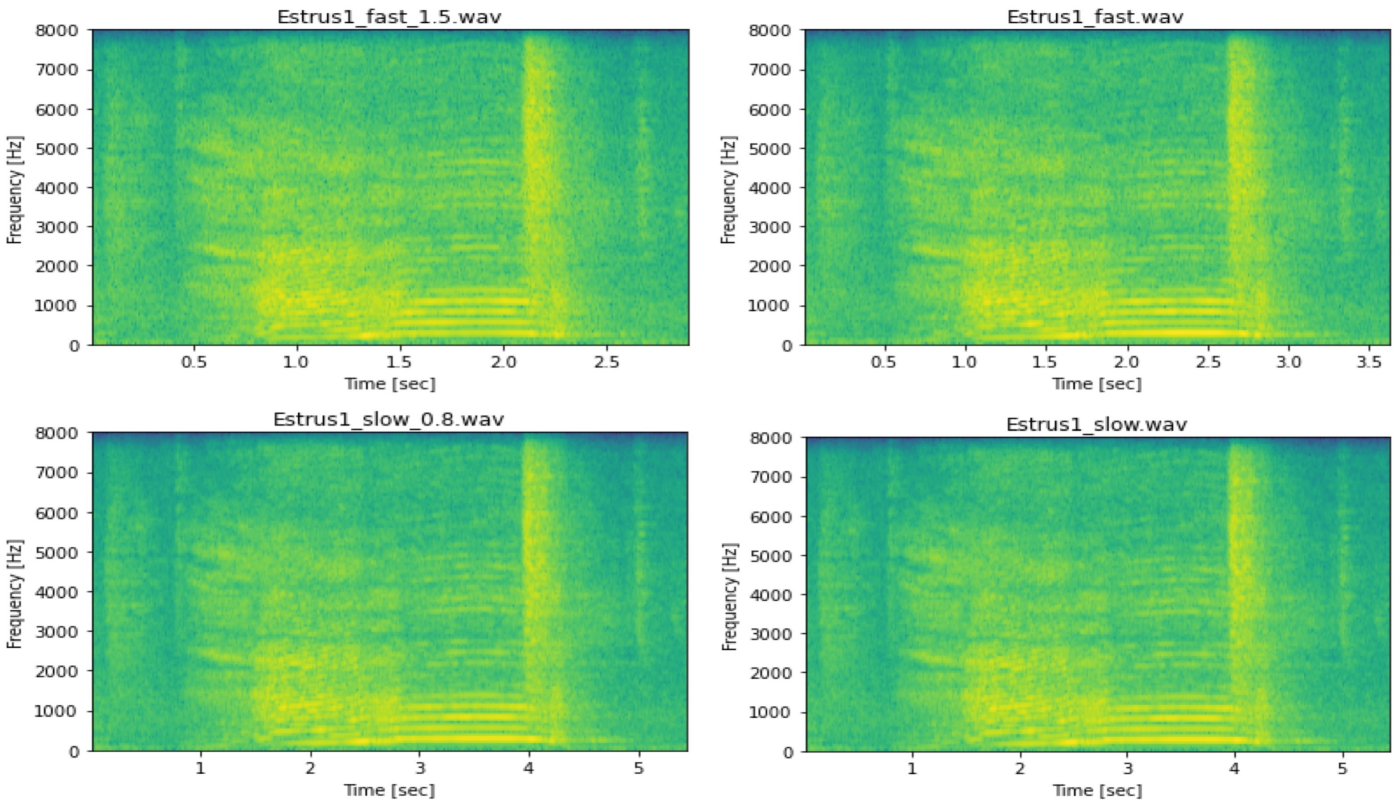


***Fig. 3.*** *Spectrograms of augmented estrus cow audio signals with time stretch of varying stretch factor*

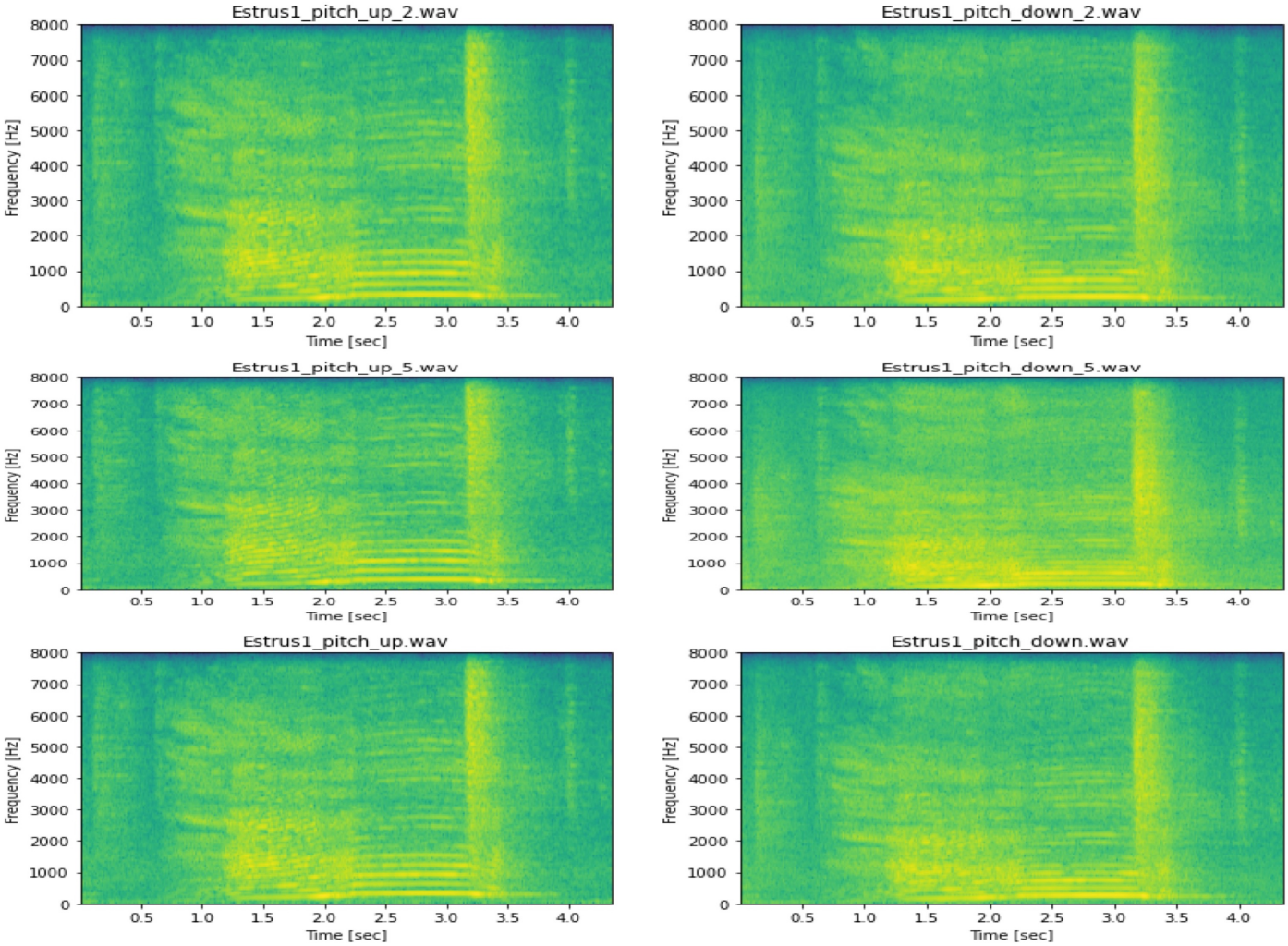


***Fig. 4.*** *Spectrograms of augmented estrus cow audio signals by applying negative and positive pitch shift*

individual performance. The ensemble technique achieves robustness to noise and improved classification accuracy.

## 3. Results

The evaluation of individual LeNet-5, VGG-16, and ResNet-50 models in terms of accuracy and loss metrics for cow vocalizations are depicted in Fig. 6 – 8. Subsequently, ensemble model was compared with individual CNN models to evaluate improvements in estrus detection.

Fig. 9 shows the confusion matrix for testing dataset. Table 1 shows

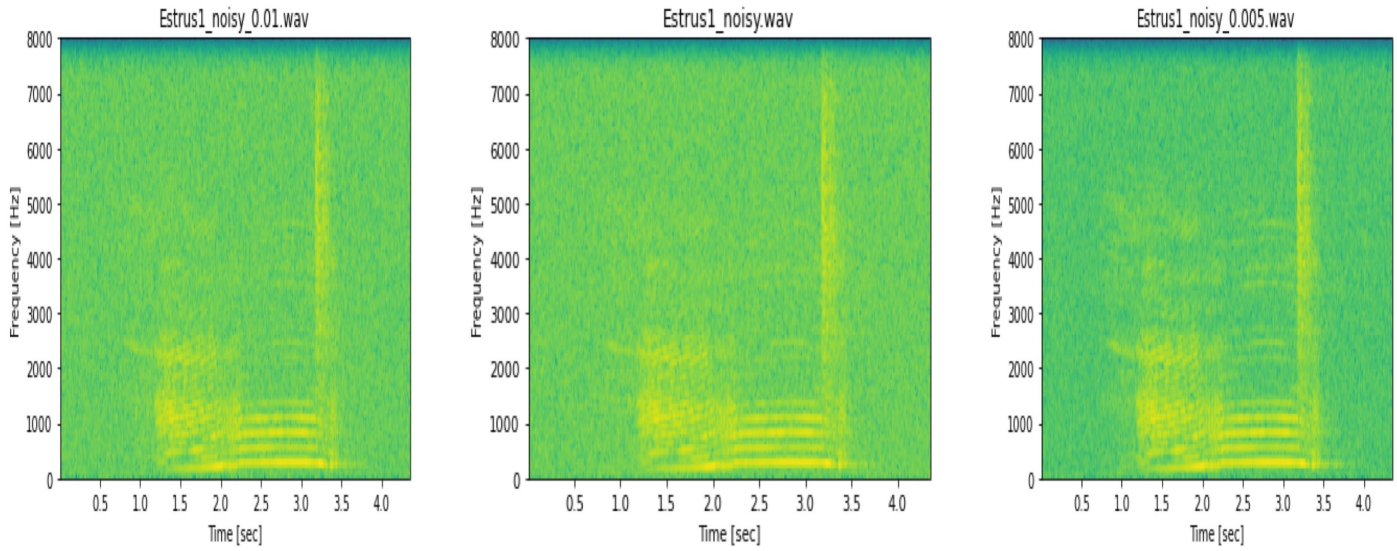


***Fig. 5.*** *Spectrograms of augmented estrus cow audio signals by adding background noise*

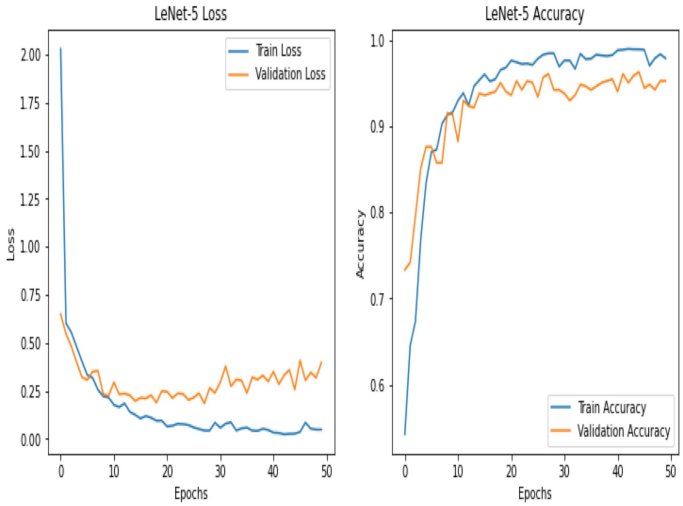


*Fig. 6. LeNet-5 accuracy and loss*

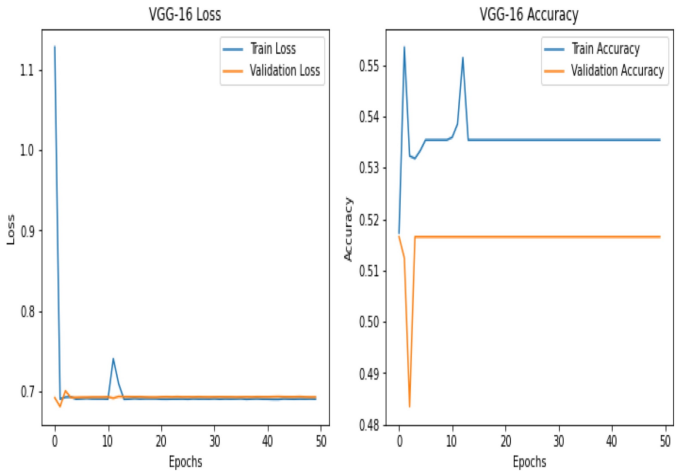


*Fig. 7. VGG-16 accuracy and loss*

accuracy, precision, recall, and F1-score for testing of datasets obtained by implementation of ensemble Model.

The performance of an ensemble model was evaluated using individual deep learning models (LeNet-5, ResNet-50, and VGG-16) and also final ensemble learning framework. Fig. 6 shows that LeNet-5 displays dependable and efficient learning. Validation loss stabilizes at about 0.25-0.35, whereas the training loss decreases from 2.0 to less than 0.1. Training accuracy rises from nearly 55% to 98%, while validation accuracy reaches 94-95%, indicating slight overfitting and strong generalization, even with its simple structure, LeNet-5 effectively captures key acoustic features.

Estrus detection using the VGG-16 model, as shown in Fig. 7, has shown poor learning performance, during training, loss values remain constant at 0.69, whereas training and validation accuracy range between 53-55% and 51-52%, respectively, suggesting underfitting and

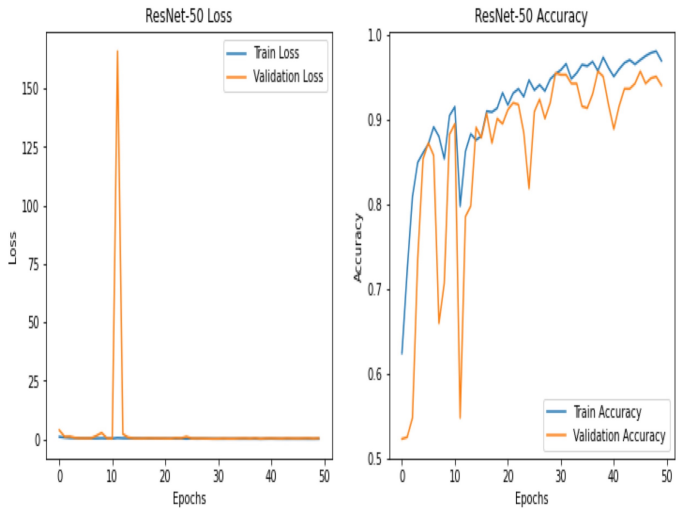


*Fig. 8. ResNet-50 accuracy and loss*

rendering VGG-16 ineffective as a stand-alone classifier which is most likely due to the model's complexity relative to the amount of the dataset. Fig. 8 shows that the ResNet-50 model for estrus detection demonstrated excellent feature-extraction capabilities and speedy convergence. Validation accuracy was stable at 93-95% but training accuracy increases from around 62% to 97-98%. The ResNet-50 model stabilizes in later stages, exhibiting significant learning allowed by residual connections, despite substantial validation loss spikes in the early epochs while model exhibited high discriminative power, particularly for intricate patterns in data.

By combining LeNet-5, ResNet-50, and VGG-Models using weighted averaging, the proposed ensemble learning model achieved an overall accuracy of 96.03% for estrus cow identification. Ensemble model achieved high precision of 97.3%, recall of 95.5%, and F1-score of 96.5% for the estrus class, indicating consistent recognition with minimal false alarms. Similarly, the non-estrus class attained balanced performance with a precision of 94.5%, recall of 96.6%, and F1-score of 95.5%. Accuracy of the methods reported in previous studies and the proposed ensemble model are given in Table 2.

## 4. Discussion

The ensemble learning model using LeNet-5, ResNet-50, and VGG-16 demonstrated strong performance for the recognition of estrus and non-estrus cows using vocalization signals achieved a notable accuracy of 96.03%. Hence, the findings demonstrated that vocal characteristics contain biologically meaningful information associated with reproductive state that can be effectively captured using deep learning approaches (Jung et al. 2021; Röttgen et al. 2018). Cows usually vocalize more frequently during periods of elevated reproductive activity and hormonal changes, increased vocal activity was observed during the estrus phase (Wang et al. 2023; Adi et al. 2020; Göncü et al. 2019). From a biological perspective, during the estrus period, endocrine and behavioral modifications occurring throughout the reproductive cycle

**Table 1. Precision, recall, F1-score and accuracy for estrus cow detection**

| Class | Precision | Recall | F1 - score | Ensemble model accuracy |
|---|---|---|---|---|
| Estrus Cow | 97.3% | 95.5% | 96.5% | 96.03% |
| Non-estrus cow | 94.5% | 96.6% | 95.5% | |

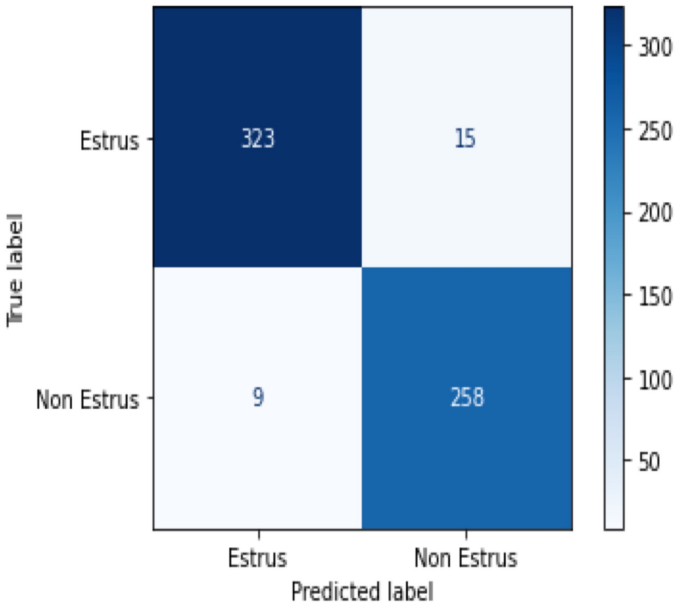


*Fig. 9. Confusion matrix for testing*

cause changes in cow vocalizations. Circulating estrogen concentrations increase substantially during estrus phase of cows, stimulating reproductive behavior and influencing vocal communication patterns, higher estrogen levels are associated with increased activity, restlessness, mounting behavior, and social interactions among cows (Röttgen et al. 2018). Therefore, cows in estrus tend to vocalize more frequently and with greater intensity than non-estrus cows (Röttgen et al. 2018; Schon et al. 2007). The increased vocal activity during estrus constitutes a social signal to communicate reproductive readiness and attract potential mates within the herd. The acoustic properties of vocalizations of cows in estrus are affected by behavioral and physiological changes, making their vocalizations distinguishable from those of non-estrus cows. Hence, classification performance supports the hypothesis that estrus-specific vocal patterns contain discriminative acoustic signatures that can be detected through machine learning techniques (Jung et al. 2021; Wang et al. 2023).

From a technical perspective, a comparative analysis of individual CNN models with ensemble CNN model revealed notable differences in learning capability. Estrus detection using ResNet-50 achieved superior performance among the individual models owing to its residual learning architecture. Residual or skip connections facilitate efficient gradient propagation during training, enabling the network to learn deeper and more discriminative representations without suffering from vanishing-gradient problems. ResNet-50 is useful when processing complex acoustic features extracted from cow vocalizations, resulting in faster convergence and improved generalization performance.

The LeNet-5 model used for estrus detection is smaller and simpler than modern deep learning architectures. The MFCC-based feature vectors used in LeNet-5 provide compact and highly informative acoustic representations, the relatively shallow architecture of LeNet-5 enables it to learn the dominant spectral characteristics of the vocalizations and its lower number of trainable parameters reduces the risk of overfitting when compared with deeper networks trained on limited datasets.

Estrus detection by VGG-16 exhibited poor performance with clear signs of underfitting. Although VGG-16 has demonstrated remarkable success in large-scale classification tasks, its deep architecture contains a large number of parameters and generally requires extensive training data. Even after data augmentation, the effective diversity of the available cow vocalization dataset remained limited relative to the complexity of the model. Hence VGG-16 for estrus cow identification struggled to learn stable acoustic representations and converged toward near-random classification performance. The ensemble learning model using LeNet-5, VGG-16, and ResNet-50 has effectively leveraged the complementary strengths of the individual CNN models. ResNet-50 model for estrus detection contributed strong feature-extraction capabilities while LeNet-5 provided robust generalization, the weighted averaging mechanism combined their predictions to reduce model-specific errors and improve classification stability. Therefore, the ensemble learning model using LeNet-5, VGG-16, and ResNet-50 achieved more consistency than the individual models and demonstrated enhanced robustness to acoustic variability and environmental noise. Results confirm that bio-acoustic monitoring

**Table 2. The comparison of accuracy from proposed ensemble model with previous studies**

| Authors | Method | Results |
|---|---|---|
| Jung et al. (2021) | CNN-based cattle vocal classification with noise filtering | Achieved 94–96% classification accuracy for cattle vocal sounds in real-time farm conditions. |
| Wang et al. (2023) | Machine learning on vocal features using dual-channel acoustic tags | Estrus identification accuracy of 91–93%, outperforming single-channel acoustic systems. |
| Alves dos Santos et al. (2022) | Review of automated estrous and calving detection systems | Surveyed systems reporting 85–95% detection accuracy, with multi-sensor approaches performing best. |
| Gavojdian et al. (2024) | BovineTalk ML framework for affective state detection | Classification accuracy of 88–92% for identifying negative affective states (isolation stress). |
| Green et al. (2021) | Estrous monitoring in pasture-based dairy systems | Estrus detection sensitivity of 85–90% using combined behavioral indicators. |
| Proposed ensemble learning model | Ensemble Machine Learning for estrus cow identification by integrating multiple CNN models—LeNet-5, ResNet-50 and VGG-16 | Overall classification accuracy of 96.03%. High precision of 97.3%, recall of 95.5% and F1-score of 96.5% for the estrus class Similarly, the non-estrus class attains balanced performance with a precision of 94.5%, recall of 96.6% and F1-score of 95.5% |

combined with ensemble learning can serve as an effective, non-invasive, and scalable means for automated estrus recognition in precision livestock farming, while future should investigate larger multi-farm datasets, additional breeds, and multimodal integration of acoustic, behavioral, and physiological signals to further improve system robustness and practical deployment.

## 5. Conclusions

The current study introduced a framework for identification of cows in estrus based on the analysis of their vocalization by integrating different convolutional neural network architectures, including LeNet-5, ResNet-50 and VGG-16. The results revealed that LeNet-5 and ResNet-50 performed well, but VGG-16 had limited learning when employed alone. The individual models complemented each other in feature-extraction skills to increase robustness of classification. Ensemble learning based on weighted averaging improved model stability, with an overall accuracy of 96.03%, precision of 97.3%, recall of 95.5%, and F1-score of 96.5% in identifying vocalizations of cows in estrus, indicating its potential application in precision livestock farming.

## Declarations

**Funding**: This study did not receive funding from any agency

**Conflict of interest**: The authors declare that they have no conflicts of interests of any kind

**Acknowledgements**: The researchers appreciate Topigs Norsvin Farm for providing secondary data from its database for this study and the concerned staff for assisting with data collection and recording

**Data Availability:** Proposed research utilize the supplementary cattle vocalization dataset by Jung et al. (2021). Audio augmentation was implemented to generate additional training samples and processed datasets are available from the corresponding author upon reasonable request

**Ethics approval:** Not applicable

## References

Adi YK, Padeta I, Prihatno SA. (2020). Analysis of cow's vocalization number at the estrus period in the traditional farm in Yogyakarta Special Region, Indonesia. Scholars Journal of Agriculture and Veterinary Sciences 7(4): 77-81. https://doi.org/10.36347/sjavs.2020.v07i04.002

Alves dos Santos C, Landim NMD, de Araújo HX, Paim TP. (2022). Automated systems for estrous and calving detection in dairy cattle. AgriEngineering 4(2): 475-482. https://doi.org/10.3390/agriengineering4020031

Gavojdian D, Mincu M, Lazebnik T, Oren A, Nicolae I, Zamansky A. (2024). BovineTalk: machine learning for vocalization analysis of dairy cattle under the negative affective state of isolation. Frontiers in Veterinary Science 11: 1357109. https://doi.org/10.3389/fvets.2024.1357109

Göncü S, Bozkurt S. (2019). Holstein cow vocalization behavior during oestrus periods. MOJ Ecology & Environmental Sciences 4(6): 276-279. https://doi.org/10.15406/mojes.2019.04.00165

Green AC, Lidfors LM, Lomax S, Favaro L, Clark CEF. (2021). Vocal production in postpartum dairy cows: Temporal organization and association with maternal and stress behaviors. Journal of Dairy Science 104(1): 826-838. https://doi.org/10.3168/jds.2020-18891

Hiregoud V, Reddy RVS. (2025). Enhancing the accuracy of region of interest detection in multi-lesion brain tumors using a hybrid deep learning network. Engineering, Technology & Applied Science Research 15(4): 24181-24187. https://doi.org/10.48084/etasr.10772

Jobarteh B, Mincu-Iorga M, Gavojdian D, Neethirajan S. (2025). Integrating multi-modal data fusion approaches for analysis of dairy cattle vocalizations. Frontiers in Veterinary Science 12: 1704031. https://doi.org/10.3389/fvets.2025.1704031

Jung DH, Kim NY, Moon SH, Jhin C, Kim HJ, Yang JS, Kim HS, Lee TS, Lee JY, Park SH. (2021). Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering. Animals 11(2): 357. https://doi.org/10.3390/ani11020357

Neave HW, Jensen EH, Durrenwachter M, Jensen MB. (2024). Behavioral responses of dairy cows and their calves to gradual or abrupt weaning and separation when managed in full- or part-time cow-calf contact systems. Journal of Dairy Science 107(4): 2297-2320. https://doi.org/10.3168/jds.2023-24085

Röttgen V, Becker F, Tuchscherer A, Wrenzycki C, Düpjan S, Schön PC, Puppe B. (2018). Vocalization as an indicator of estrus climax in Holstein heifers during natural estrus and superovulation. Journal of Dairy Science 101(3): 2383–2394. https://doi.org/10.3168/jds.2017-13412

Schön PC, Hämel K, Puppe B, Tuchscherer A, Kanitz W, Manteuffel G. (2007). Altered vocalization rate during the estrous cycle in dairy cattle. Journal of Dairy Science 90(1): 202–206. https://doi.org/10.3168/jds.S0022-0302(07)72621-8

Wang J, Chen H, Wang J, Zhao K, Li X, Liu B, Zhou Y. (2023). Identification of oestrus cows based on vocalisation characteristics and machine learning technique using a dual-channel-equipped acoustic tag. Animal 17: 100811. https://doi.org/10.1016/j.animal.2023.100811

**Citation**

Rohini HM, Prabhavathi S, Angadi D. (2026). Identification of estrus cows through their vocalizations using Ensemble Machine Learning. Letters in Animal Biology 06(2): 139 – 145. *https://doi.org/10.62310/liab.v6i2.391*