Identifying Indian Cattle Behaviour Using Acoustic Biomarkers

Ruturaj Patil, Hemavathy B, Sanat Sarangi, Dineshkumar Singh,
Rupayan Chakraborty, Sanket Junagade and Srinivasu Pappula
TCS Research, Tata Consultancy Services Limited, India
patil.ruturaj, hemavathy.b, sanat.sarangi, dineshkumar.singh, rupayan.chakraborty, sanket.junagade,

Keywords: Cattle Vocalisation, Acoustic Feature Recognition, Livestock Behaviour, OpenSMILE, Deep Learning.

Abstract: A system to recognise sounds from some major cattle breeds commonly found in India and linking them to intents reflecting specific behaviour along with associated needs is proposed. Cattle breeds in India consist of a mix of indigenous and exotic breeds where Sindhi, Sahiwal, and Gir make up a significant fraction of the indigenous breeds. Exotic breeds are Jersey and Holstein Friesian. Vocalisation from the animals in this cattle group is used to create a sound dataset comprising 120 utterances for over six intents where the intents were labelled by domain experts familiar with the animals and their behaviour. MFCCs and OpenSMILE global features from the audio signal with 6552 properties are used to model for intent recognition. The dataset is scaled and augmented with four different methods to 870 cattle sounds for the six classes. Two model architectures are created and tested on data for each method independently and with all of them together. The models are also tested on unseen cattle sounds for speaker independent verification. An accuracy of 97% was obtained for intent classification with MFCCs and OpenSMILE features. This indicates that behaviour recognition from sounds for Indian cattle breeds is possible with a good confidence level.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

Livestock are a critical component of the global food production ecosystem. Behaviour and health of livestock must be monitored to ensure their welfare, productivity, and profitability (Orihuela, 2021). Controlling cattle in large herds can be difficult for farmers, as monitoring the behaviour of each individual animal in a sizeable group takes time and requires constant attention (Herlin et al., 2021). Cattle monitoring can be facilitated by understanding their behaviour from vocalisation. Some work has been done to classify sounds of animals and birds based on their distinct characteristics (Lin et al., 2018), (Schröter et al., 2019), (Thakur et al., 2019), (Briggs et al., 2013), (Esposito et al., 2023). There are studies that discuss sound classification in human and environmental context (Song et al., 2021), (Gong et al., 2022), (Wang et al., 2021), (Jindal et al., 2021). Dairy-cattle in India largely consists of a mix of indigenous and exotic animal breeds. Indigenous breeds include Gir, Sindhi and Sahiwal while the exotic breeds are Jersey and Holstein Friesian (HF). Traditional livestock monitoring involves regular visual inspection of the cattle by farmers based on their experience and knowledge. They check for indicators of disease or discomfort, as well as irregularities in behaviour and health. In modern livestock monitoring procedures, body sensors are attached to the animal's body and vital signs such as heart rate, body temperature, and respiration rate are monitored (Sharma and Koundal, 2018). Cameras are utilized to observe the animal's behaviour and detect any indicators of illness or discomfort (Wu et al., 2023). GPS collars are used to track the animal's location in real time and to monitor its activity and health. Livestock behaviour monitoring could be significantly enhanced if a more natural and non-invasive approach involving deeper interpretation of cattle sounds and conversations in their corresponding ecosystems could be part of the interventions. This would even strengthen the body of work on human-cattle interaction systems going forward.

We present an intelligent cattle monitoring system based on the interpretation of cattle behaviour using their vocalisation. The system involves (a) creating a corpus of cattle sounds by collecting data from the field for the scope of interest, (b) preparing a dataset in which the utterances and intents for each animal are extracted with certain segmentation and preprocessing steps, and (c) developing models that work on specific features to analyses these sounds in order to identify the cattle intent or behaviour. Handcrafted audio

Patil, R., B, H., Sarangi, S., Singh, D., Chakraborty, R., Junagade, S. and Pappula, S.
Identifying Indian Cattle Behaviour Using Acoustic Biomarkers.
DOI: 10.5220/0012576800003654
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 594-602
ISBN: 978-989-758-684-2; ISSN: 2184-4313
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

deep model based embeddings are explored to extract meaningful features for training a deep sound classification architecture. For the handcrafted features, we extract a 6552 dimensional vector that consists of statistics (mean, standard deviation, skewness, kurtosis, extremes, linear regressions, etc.) of several Low-Level Descriptors (LLDs) (e.g., Zero Crossing Rate (ZCR), Root Mean Square (RMS) energy, Fundamental frequency (F0), Harmonic to Noise Ratio (HNR), Mel-Frequency Cepstral Coefficients (MFCCs)) from each audio sample (Eyben et al., 2010). MFCC features are also extracted separately. Next, deep learning techniques are used to train the model to recognise various sorts of cow sounds and connect them with specific behaviours. Using simply vocalisation, our cattle monitoring system can assist in monitoring cattle behaviour, health, and anomalies. It is important to note that, to monitor cattle activity and health, no additional equipment such as multiple sensors, GPS collars, or cameras will be required. This research article intends to contribute to the development of innovative cattle monitoring tools that can assist farmers in optimising herd management practices and improving animal welfare.

The main contributions of the paper are as follows. (1) To the best of our knowledge, a sound corpus for indigenous dairy cattle breeds is created to be used in an intelligent cattle monitoring system for the first time in India. (2) The exploration of a diversified set of state-of-the-art features in an intelligent cattle monitoring system is done for the first time, with an aim to understand which one shows promise for cattle-sound classification. (3) A framework with a deep learning model, trained using only a handful of real-life audio recordings, with an extension of sample space through diversified augmentation techniques, shows good performance (accuracies around 88%) even in subject independent testing scenario. This suggests that it is possible to build generalised models with a small dataset. The rest of the paper is organized as follows. In Section 2, we discuss briefly the prior art. Section 3 describes different methods of feature extraction and the deep learning model architecture. Section 4 presents the dataset, data augmentation, experimental details, results and analysis. In Section 5, we conclude our work.

2 RELATED WORK

In (Jung et al., 2021a), a deep-learning based model for classifying four different cattle vocal sounds, including oestrus call, food anticipating call, cough sound, and normal call is proposed. Authors used Korean native cattle with 897 cattle voice samples. The developed model achieved a final classification accuracy of 81.96%. In (Sattar, 2022), authors used Multiclass Support Vector Machine (MSVM) classifier with contextual acoustic features to group 270 cattle records into four classes such as oestrus call, food anticipating call, cough sound, and normal call. The system achieves 84% classification accuracy. In the above papers, authors studied only four behaviours with a single breed and in a limited age group. In (Green et al., 2019), authors analysed 333 calls of 13 HF non-pregnant virgin heifers using Praat DSP package v.6.0.31, through calculation of both oscillograms and spectrograms. Vocalisation is grouped into five classes, as in oestrus, anticipating feed, denied feed access, physically isolated from conspecifics, and physically and visually isolated from conspecifics. In (Ikeda and Ishii, 2008), authors examined how a Japanese black cow's vocalisation changed under two psychologically demanding circumstances: being hungry before eating and being taken away from her calf during weaning, and found that the cow's vocalisation varied in both cases. In (Jung et al., 2021b), authors built convolutional neural network (CNN) models to categorise the vocalisation of laying hens and calves where the cattle-sound classes include cattle call for isolation of colleagues, oestrus, pain by parturition, food anticipation, maternal after parturition, cough, weaning and growing calf food anticipation call, and neonatal calf call. In (Green et al., 2019), (Ikeda and Ishii, 2008) and (Jung et al., 2021b), authors studied a single exotic breed with a limited age group. In the prior work, study on behaviour or intent identification of cattle breeds in India and especially for indigenous breeds is missing. The scope in terms of breed, age-group, or classes is small, so there is a limited view into generalisation of the outcomes. Furthermore, subject independent evaluation of the proposed models are not carried out.

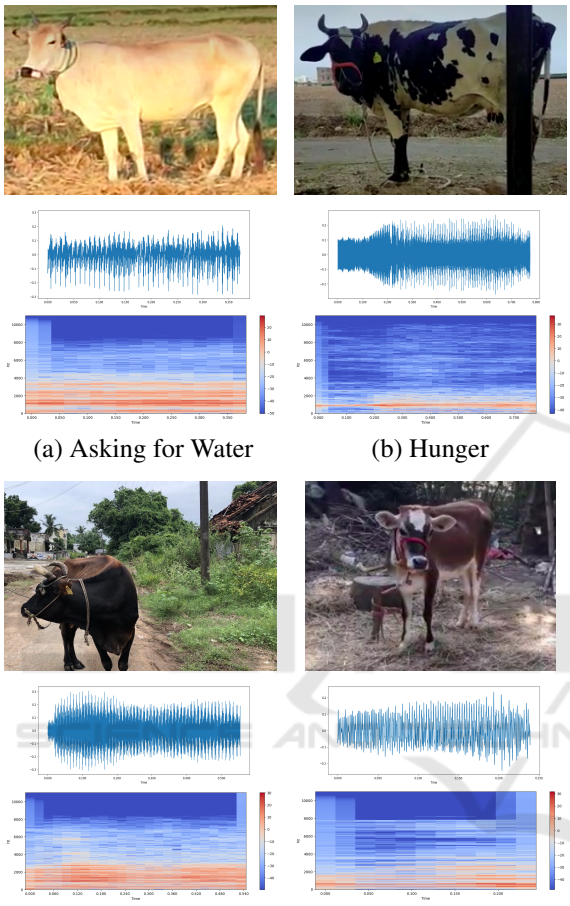
3 LIVESTOCK SOUNDS AND PROPOSED WORK

3.1 High Level Features Extraction

For audio samples in the dataset discussed in Sec. 4.2.3, multiple feature vectors are extracted for analysis, which includes (a) global features from OpenSMILE (Eyben et al., 2010) and (b) mean MFCC features as the low-level features. OpenSMILE features include a 6552-dimensional vector made up of statistics for a number of low-level de-

Table 1: Dataset Across Breeds.

Breed	Collected files	Extracted chunks
Exotic	34	75
Indigenous	26	45



(a) Asking for Water (b) Hunger

(c) Wants to go for Grazing (d) Calling Mother

Figure 1: Example Dataset.

scriptors (LLDs), including mean, standard deviation, skewness, kurtosis, extremes, linear regressions, etc. It supports low-level audio descriptors like loudness, Mel-frequency cepstral coefficients, perceptual linear predictive cepstral coefficients, linear predictive coefficients, fundamental frequency, formant frequencies, etc. Separately, the mean MFCC features are used to prepare a 40-dimensional feature-vector to train the model. MFCC features extraction included several steps, viz., Pre-emphasis, Frame Segmentation, Fast Fourier Transform (FFT), Mel Filerbank, Logarithmic Transformation, Discrete Cosine Transform (DCT), Delta and Delta-Delta Features. Pre-emphasis is a filtering technique that emphasizes higher frequencies. Its objective is to balance the spectrum of spoken sounds, which has a sharp roll-off in the high-frequency range. The transfer function below provides the pre-emphasis filter.

$$H(z) = 1 - bz^{-1} \tag{1}$$

where the value of b controls the slope of the filter and is usually between 0.4 and 1.0 (Picone, 1993). To limit spectral leakage, audio signal is divided into frames and window function is applied. DFT is used to transform each windowed frame into a magnitude spectrum.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{-j2\pi nk}{N}}; 0 \leq k \leq N-1 \tag{2}$$

where N is the number of points used to compute the DFT. The Fourier transformed signal is run through the Mel-filter bank, a collection of band-pass filters, to compute the Mel spectrum. A Mel is a unit of measurement based on the perceived frequency by human ears. Mel can be approximated by physical frequency as follows:

$$f_{Mel} = 2595 \log_{10}(1 + \frac{f}{700}) \tag{3}$$

where f denotes the physical frequency in Hz, and f_{Mel} denotes the perceived frequency (Deller Jr, 1993). The magnitude spectrum is multiplied by each of the triangular Mel weighting filters to get the Mel spectrum of the magnitude spectrum X(k).

$$s(m) = \sum_{k=0}^{N-1} [|X(k)|^2 H_m(k)]; 0 \leq m \leq M-1 \tag{4}$$

where M is total number of triangular Mel weighting filters (Zheng et al., 2001), (Ganchev et al., 2005). $H_m(k)$ is the weight given to the k^{th} energy spectrum bin contributing to the m^{th} output band. To obtain MFCC coefficients, DCT is applied and first few coefficients are retained. To capture temporal changes in MFCCs, delta and delta-delta coefficients are computed.

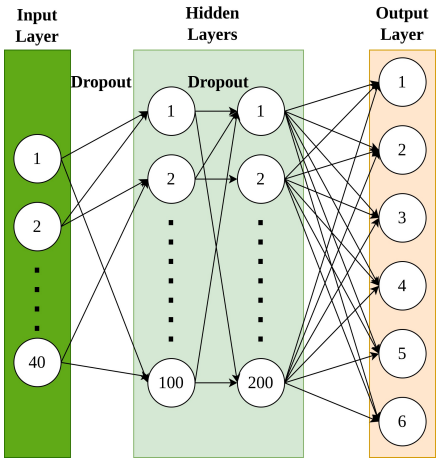
MFCC is calculated as (Picone, 1993),

$$c(n) = \sum_{m=0}^{M-1} \log_{10}(s(m)) cos(\frac{\pi n(m-0.5)}{M}) \tag{5}$$

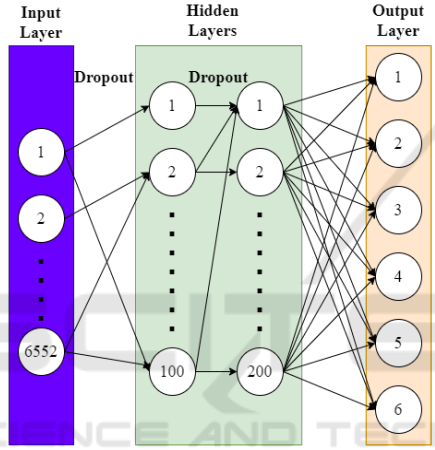
where $c(n)$ are the cepstral coefficients, $n = 0, 1, 2, ...,$ $C-1$ and C is the number of MFCC features.

3.2 Classification Model Architecture

To classify the features extracted from audio, a neural network (ANN) model is proposed with six classes, which constitute the output shape. The ANN is created with three dense layers, as shown in Fig. 2.



(a) With MFCC Features



(b) With OpenSMILE Features

Figure 2: Classification Model Architecture.

ReLU (Rectified Linear Unit) (Hara et al., 2015) and Softmax Activation Functions (Kouretas and Paliouras, 2019) are used with the dropout layer at a rate of 0.5. ReLU outputs the input for positive values and zero for negative values, aiding in model convergence. Softmax Activation Function used in the output layer for multi-class classification. Converts raw scores into probabilities, making it easier to interpret results. To compile the model, Sparse Categorical Cross Entropy loss function (Dousti Mousavi et al., 2023) is defined and Adam optimizer (Zhang, 2018) is used. Sparse Categorical Cross Entropy loss function is suitable for multi-class classification problems. Particularly useful when the target classes are mutually exclusive. An adaptive optimisation algorithm efficiently adjusts learning rates for each parameter, improving convergence speed. Augmentation methods discussed in Sec. 4.2.2 that involve noise injection, pitch shift, time stretch, time shift and all of these together are used to develop five datasets. Five models are created, one for each of the datasets with MFCC features. Similarly, five models are created with OpenSMILE features. A split of 60 : 20 : 20 is used for training, validation, and testing data from the combination of original and augmented data. This translates to a split of 208 : 70 : 70 audio chunks for each augmentation approach and a split of 522 : 174 : 174 audio chunks for the overall dataset as discussed in Sec. 4.2.3. Each model is trained for 100 epochs.

Table 2: Dataset Across Intents.

Intent	Collected files	Extracted chunk
Asking for water	14	31
Calling Mother	11	15
Hunger	10	20
Mating	14	19
Wants to go for grazing	6	30
Wants to go home	5	5

4 EXPERIMENTS AND RESULTS

4.1 Data Collection

A total of 60 audio or video files across breeds of interest were collected as shown in Tables 1 and 2. Data is collected from different places associated with the usual environment and movement context of the cattle, such as cattle shed, grazing field, and veterinary hospital. The associated utterances and corresponding intents were recorded such as: wants to go for grazing, asking for water, calling for hunger, wants mating, calves calling their mother, wants to go home as shown in Fig. 1 and Table 2. Fig. 1 shows example utterances in waveform in the time-domain followed by spectrograms in the frequency-domain. Collected audio and video files of cattle utterances are converted into wav files. Table 1 gives the utterance data across 34 animals for exotic (HF, Jersey) and indigenous (Sindhi, Sahiwal, and Country (crossbred)) breeds where average cattle age varies from 0.5 year to 7 years.

4.2 Data Preprocessing and Cleaning Techniques

4.2.1 Data Annotation and Extracting Audio Chunks

To annotate the data, the utterance start and end times are recorded for every small utterance captured in the entire audio file. Intents of cows are recorded

as sound description, which are our targeted output classes. The trigger that caused the utterance is also recorded. Annotated utterance start and end times can be used to manually extract small audio chunks from a complete audio file. However, to automate the process, we have used a threshold method to extract utterances. We used silence length as 100 ms and silence threshold as 30 dB. Audios in these small chunks are verified for cattle sounds. 120 small audio chunks are generated after extracting all utterances as shown in Tables 1 and 2. A pictorial view of the extracted utterance for some intents is shown in Fig. 1.

4.2.2 Augmentation

Four different audio augmentation methods (Nanni et al., 2020b) are used, viz., noise injection, pitch shift, time stretch, and time shift for six different classes as wants to go for grazing, asking for water, hunger, calling mother, mating, and wants to go home. In noise injection, a randomly generated value of the noise factor between 0 and 0.01 multiplied by noise is added to the data. In pitch shift, pitch is changed with a randomly generated pitch factor between -1 and 1. In time stretch, the times series is stretched by a randomly generated speed factor between 0.8 and 1.2. In time shift, audio is shifted to left or right with a random second value between -0.5 and 0.5.

4.2.3 Database

Five datasets are prepared after augmentation. The first dataset, having each class with 58 audio utterances, contains original audio chunks as shown in Table 2 and remaining augmented chunks with noise injection. The resultant dataset contains 348 audio utterances with six classes, each with 58 chunks. Similarly, 348 audio utterances are generated for each augmentation approach. The fifth dataset, consisting of 870 utterances, is created by combining datasets from all augmentation approaches. Datasets have mean, median and mode durations of 1.28, 1.27 and 2 seconds respectively. The minimum and maximum file durations are 0.24 and 2.98 seconds, respectively. Datasets have a variance and standard deviation of 0.33 and 0.58 seconds, respectively. One more dataset is created by keeping unseen entities (cattle) aside to test the speaker independent performance of the model. These audios are not used in model training, i.e., utterances by new cattle. The unseen dataset contains five classes as wants to go for grazing, asking for water, hunger, mating, calling mother. Each class has 3 to 4 chunks, which are augmented to make 15 audio chunks per class.

Table 3: Comparison of Accuracy.

Model	Accuracy	
	With MFCC Features	With OpenSmile Features
Pitch Shift	83	89
Noise Injection	86	86
Time Shift	87	89
Time Stretch	89	87
Overall Data	97	97
Unseen Data	87	88

Table 4: Scores for Pitch Shift Model with MFCC Features.

Class	Precision	Recall	F1-Score
Asking for Water	0.60	0.25	0.35
Calling Mother	1	1	1
Hunger	0.56	0.82	0.67
Mating	1	1	1
Wants to go for grazing	0.79	0.92	0.85
Want to go home	1	1	1

Table 5: Scores for Pitch Shift Model with OpenSMILE Features.

Class	Precision	Recall	F1-Score
Asking for Water	0.86	0.5	0.63
Calling Mother	0.92	1	0.96
Hunger	0.67	0.91	0.77
Mating	0.92	1	0.96
Wants to go for grazing	1	0.92	0.96
Want to go home	1	1	1

Table 6: Scores for Noise Injection Model with MFCC Features.

Class	Precision	Recall	F1-Score
Asking for Water	0.71	0.42	0.53
Calling Mother	1	1	1
Hunger	0.63	0.91	0.74
Mating	1	1	1
Wants to go for grazing	0.83	0.83	0.83
Want to go home	1	1	1

4.3 Results

4.3.1 All Datasets

The classification of six classes is achieved with test accuracy of 83%, 86%, 87%, 89% and 97% as shown in Table 3 with kappa coefficients of 0.79, 0.83, 0.85, 0.86 and 0.97 for models with pitch shift, noise in-

Table 7: Scores for Noise Injection Model with OpenSMILE Features.

Class	Precision	Recall	F1-Score
Asking for Water	0.67	0.5	0.57
Calling Mother	1	1	1
Hunger	0.64	0.82	0.72
Mating	0.92	1	0.96
Wants to go for grazing	0.91	0.83	0.87
Want to go home	1	1	1

Table 8: Scores for Time Shift Model with MFCC Features.

Class	Precision	Recall	F1-Score
Asking for Water	1	0.33	0.50
Calling Mother	1	1	1
Hunger	0.61	1	0.76
Mating	1	1	1
Wants to go for grazing	0.85	0.92	0.88
Want to go home	1	1	1

Table 9: Scores for Time Shift Model with OpenSMILE Features.

Class	Precision	Recall	F1-Score
Asking for Water	0.78	0.58	0.67
Calling Mother	1	1	1
Hunger	0.69	1	0.81
Mating	1	1	1
Wants to go for grazing	0.90	0.75	0.82
Want to go home	1	1	1

Table 10: Scores for Time Stretch Model with MFCC Features.

Class	Precision	Recall	F1-Score
Asking for Water	1	0.42	0.59
Calling Mother	1	1	1
Hunger	0.65	1	0.79
Mating	1	1	1
Wants to go for grazing	0.85	0.92	0.88
Want to go home	1	1	1

jection, time shift, time stretch and overall data, respectively, with MFCC features. For OpenSMILE features, the test accuracies achieved are 89%, 86%, 89%, 87% and 97% with kappa coefficients of 0.86, 0.83, 0.86, 0.85 and 0.97 for models with pitch shift, noise injection, time shift, time stretch and overall data, respectively. The models were verified with the ground-truth data after training. The detailed performance metrics for each model with MFCC fea-

Table 11: Scores for Time Stretch Model with OpenSMILE Features.

Class	Precision	Recall	F1-Score
Asking for Water	0.73	0.67	0.70
Calling Mother	1	1	1
Hunger	0.62	0.73	0.67
Mating	1	1	1
Wants to go for grazing	0.91	0.83	0.87
Want to go home	1	1	1

Table 12: Scores for Overall Data Model with MFCC Features.

Class	Precision	Recall	F1-Score
Asking for water	0.96	0.86	0.91
Calling Mother	0.97	1	0.98
Hunger	0.94	1	0.97
Mating	1	0.97	0.98
Wants to go for grazing	0.97	1	0.98
Wants to go home	1	1	1

Table 13: Scores for Overall Data Model with OpenSMILE Features.

Class	Precision	Recall	F1-Score
Asking for Water	0.93	0.97	0.95
Calling Mother	1	0.93	0.96
Hunger	0.91	1	0.95
Mating	1	0.97	0.98
Wants to go for grazing	1	0.97	0.98
Want to go home	1	1	1

Table 14: Scores for Unseen Data with MFCC Features.

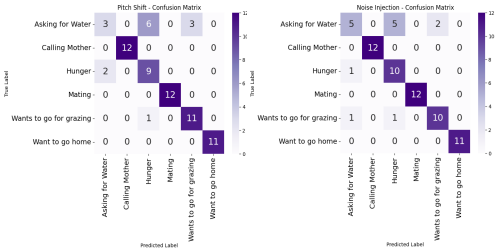
Class	Precision	Recall	F1-Score
Asking for Water	0.87	0.87	0.87
Calling Mother	1	1	1
Hunger	0.91	0.67	0.77
Mating	0.75	0.80	0.77
Wants to go for grazing	0.83	1	0.91

Table 15: Scores for Unseen Data with OpenSMILE Features.

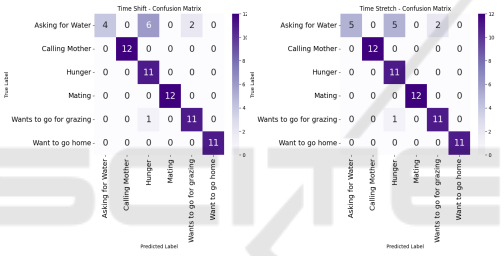
Class	Precision	Recall	F1-Score
Asking for Water	0.75	1	0.86
Calling Mother	1	0.93	0.97
Hunger	1	0.67	0.80
Mating	1	0.80	0.89
Wants to go for grazing	0.79	1	0.88

Table 16: Comparison of prediction accuracies for animal sound classification.

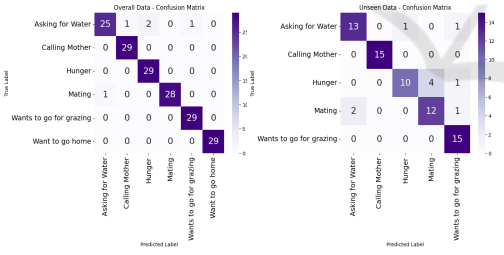
Animal	Classification Target	Approach	Accuracy (%)
Bird (Lucio et al., 2015)	Forty-six species	Handcrafted features with Support Vector Machine (SVM)	88.80
Whale (Nanni et al., 2020a)	Whale identification	CNN	97.80
Bird (Zhao et al., 2017)	Eleven bird species	SVM	96.60
Cow (Jung et al., 2021a)	Four behaviours	MFCC with CNN	81.96
Cow (Sattar, 2022)	Four behaviours	Multiclass Support Vector Machine (MSVM)	84.00
Chicken (Huang et al., 2021)	Eating behavior	Deep Learning	96.00
Cow	Six behaviours	MFCC or OpenSMILE features with ANN	97.00



(a) Pitch Shift (b) Noise Injection

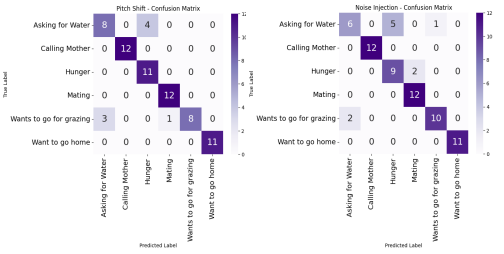


(c) Time Shift (d) Time Stretch

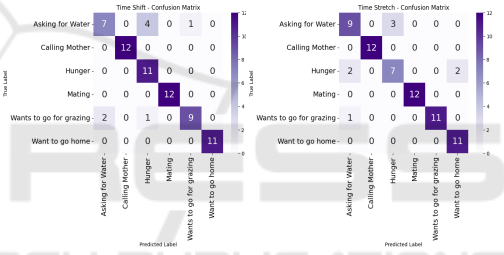


(e) Overall Data (f) Unseen Data

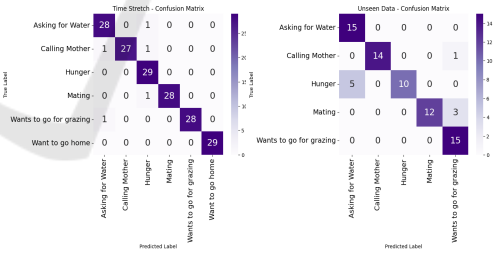
Figure 3: Confusion Metrices with MFCC Features.



(a) Pitch Shift (b) Noise Injection



(c) Time Shift (d) Time Stretch



(e) Overall Data (f) Unseen Data

Figure 4: Confusion Metrices with OpenSMILE Features.

tures and OpenSMILE features can be seen in Tables 4, 6, 8, 10, 12 and 5, 7, 9, 11, 13 respectively. The confusion matrix for each model with MFCC features and OpenSMILE features can be seen in Fig. 3 and Fig. 4 respectively. Kappa coefficient measures inter-rater agreement for categorical items. A value of 1 indicates perfect agreement, 0 indicates agreement equivalent to chance, and negative values suggest less agreement than expected by chance. Precision, also known as positive predictive value, it's the ratio of true positive predictions to the total predicted positives. High precision means fewer false positives. Recall, also called sensitivity or true positive rate, it's the ratio of true positive predictions to the total actual positives. High recall indicates capturing most of the positives. F1-Score is harmonic mean of precision and recall. Balances both metrics and provides a single score. Useful when there's an imbalance between classes. The time-stretch data model excelled in individual augmentation techniques, while the overall dataset model outperformed others. The majority of models demonstrated superior performance in distinguishing between "calling mother", "mating"

and “wants to go home”. These three classes exhibited higher feature variations when compared to the closely related other classes. Collecting data for more number of unique entities (cows) for ”Asking for water” and ”Wants to go home” classes could increase accuracy of classification of these classes. Table 16 summarizes performance comparison of prior arts methods. This table provides a representative reference for the animal speech classification accuracy of the existing deep learning technologies. Our results are added in the last row.

4.3.2 Unseen Dataset

The performance of the models with overall data is checked for subject independent unseen data. The classification of five classes is achieved with test accuracy of 87% and 88% with kappa coefficients of 0.83 and 0.85 for MFCC features and OpenSMILE features, respectively. The detailed performance metrics and confusion matrix for the model with MFCC features for unseen data can be seen in Table 14 and in Fig. 3, respectively. Similarly, the detailed performance metrics and confusion matrix for the model with OpenSMILE features for unseen data can be seen in Table 15 and in Fig. 4, respectively. From Table 3, the accuracy was highest at 97% with both the feature extraction methods for overall data (i.e., when all augmentation methods are used together). We find that the model with OpenSMILE features performed slightly better for subject independent unseen data with 88% accuracy than with MFCC features with 87% accuracy. Results indicate that the proposed methodology can be effective in monitoring cattle behaviour.

5 CONCLUSIONS

The results of this study could be utilised to create sophisticated non-invasive systems for tracking the behaviour of cattle and generating a richer understanding about their individual and collective response in various scenarios. This would be helpful for farmers and ranchers to manage them. Future studies could focus on expanding the horizon of the study to include aspects like handling the effect of background noise, intrusion detection by recognizing non-cattle sounds, and new cattle scenarios.

REFERENCES

Briggs, F., Huang, Y., Raich, R., Eftaxias, K., Lei, Z., Cukierski, W., Hadley, S. F., Hadley, A., Betts, M., Fern, X. Z., et al. (2013). The 9th annual mlsp competition: New methods for acoustic classification of multiple simultaneous bird species in a noisy environment. In *IEEE international workshop on machine learning for signal processing (MLSP)*, pages 1–8.

Deller Jr, J. R. (1993). Discrete-time processing of speech signals. In *Discrete-time processing of speech signals*, pages 908–908.

Dousti Mousavi, N., Aldirawi, H., and Yang, J. (2023). Categorical data analysis for high-dimensional sparse gene expression data. *BioTech*, 12(3):52.

Esposito, M., Valente, G., Plasencia-Calaña, Y., Dumontier, M., Giordano, B. L., and Formisano, E. (2023). Semantically-informed deep neural networks for sound recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5.

Eyben, F., Wöllmer, M., and Schuller, B. (2010). Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1459–1462.

Ganchev, T., Fakotakis, N., and Kokkinakis, G. (2005). Comparative evaluation of various mfcc implementations on the speaker verification task. In *Proceedings of the SPECOM*, volume 1, pages 191–194. Citeseer.

Gong, Y., Yu, J., and Glass, J. (2022). Vocalsound: A dataset for improving human vocal sounds recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 151–155.

Green, A., Clark, C., Favaro, L., Lomax, S., and Reby, D. (2019). Vocal individuality of holstein-friesian cattle is maintained across putatively positive and negative farming contexts. *Scientific Reports*, 9(1):18468.

Hara, K., Saito, D., and Shouno, H. (2015). Analysis of function of rectified linear unit used in deep learning. In *2015 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE.

Herlin, A., Brunberg, E., Hultgren, J., Högberg, N., Rydberg, A., and Skarin, A. (2021). Animal welfare implications of digital tools for monitoring and management of cattle and sheep on pasture. *Animals*, 11(3).

Huang, J., Zhang, T., Cuan, K., and Fang, C. (2021). An intelligent method for detecting poultry eating behaviour based on vocalization signals. *Computers and Electronics in Agriculture*, 180:105884.

Ikeda, Y. and Ishii, Y. (2008). Recognition of two psychological conditions of a single cow by her voice. *Computers and Electronics in Agriculture*, 62(1):67–72. Precision Livestock Farming (PLF).

Jindal, S., Nathwani, K., and Abrol, V. (2021). Classification of infant behavioural traits using acoustic cry: An empirical study. In *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 97–102.

Jung, D.-H., Kim, N. Y., Moon, S. H., Jhin, C., Kim, H.-J., Yang, J.-S., Kim, H. S., Lee, T. S., Lee, J. Y., and Park, S. H. (2021a). Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering. *Animals*, 11(2).

Jung, D.-H., Kim, N. Y., Moon, S. H., Kim, H. S., Lee, T. S., Yang, J.-S., Lee, J. Y., Han, X., and Park, S. H. (2021b). Classification of vocalization recordings of laying hens and cattle using convolutional neural network models. *Journal of Biosystems Engineering*, 46(3):217–224.

Kouretas, I. and Paliouras, V. (2019). Simplified hardware implementation of the softmax activation function. In *2019 8th international conference on modern circuits and systems technologies (MOCAST)*, pages 1–4. IEEE.

Lin, N., Sun, H., and Zhang, X.-P. (2018). Overlapping animal sound classification using sparse representation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2156–2160.

Lucio, D. R., Maldonado, Y., and da Costa, G. (2015). Bird species classification using spectrograms. In *2015 Latin American Computing Conference (CLEI)*, pages 1–11. IEEE.

Nanni, L., Costa, Y. M., Aguiar, R. L., Mangolin, R. B., Brahnam, S., and Silla, C. N. (2020a). Ensemble of convolutional neural networks to improve animal audio classification. *EURASIP Journal on Audio, Speech, and Music Processing*, 2020(1):1–14.

Nanni, L., Maguolo, G., and Paci, M. (2020b). Data augmentation approaches for improving animal audio classification. *Ecological Informatics*, 57:101084.

Orihuela, A. (2021). Management of livestock behavior to improve welfare and production. *Animal*, 15:100290.

Picone, J. W. (1993). Signal modeling techniques in speech recognition. *Proceedings of the IEEE*, 81(9):1215–1247.

Sattar, F. (2022). A context-aware method-based cattle vocal classification for livestock monitoring in smart farm. *Chemistry Proceedings*, 10(1).

Schröter, H., Nöth, E., Maier, A., Cheng, R., Barth, V., and Bergler, C. (2019). Segmentation, classification, and visualization of orca calls using deep learning. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8231–8235.

Sharma, B. and Koundal, D. (2018). Cattle health monitoring system using wireless sensor network: a survey from innovation perspective. *IET Wireless Sensor Systems*, 8(4):143–151.

Song, W., Han, J., and Song, H. (2021). Contrastive embeddind learning method for respiratory sound classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1275–1279.

Thakur, A., Sharma, P., Abrol, V., and Rajan, P. (2019). Conv-codes: audio hashing for bird species classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8241–8245.

Wang, Y., Feng, C., and Anderson, D. V. (2021). A multichannel temporal attention convolutional neural network model for environmental sound classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 930–934.

Wu, D., Han, M., Song, H., Song, L., and Duan, Y. (2023). Monitoring the respiratory behavior of multiple cows based on computer vision and deep learning. *Journal of Dairy Science*, 106(4):2963–2979.

Zhang, Z. (2018). Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, pages 1–2. Ieee.

Zhao, Z., Zhang, S.-h., Xu, Z.-y., Bellisario, K., Dai, N.-h., Omrani, H., and Pijanowski, B. C. (2017). Automated bird acoustic event detection and robust species classification. *Ecological Informatics*, 39:99–108.

Zheng, F., Zhang, G., and Song, Z. (2001). Comparison of different implementations of mfcc. *Journal of Computer science and Technology*, 16:582–589.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS