Transfer Learning and Sensor Fusion for Cattle Monitoring: An IoT-Driven Approach for Sustainable Livestock Farming

Mariano Ferrero*, Luciano Sebastian Martinez-Rau*†, *Member, IEEE*, Leandro Daniel Vignolo*, José Omar Chelotti*, Julio Ricardo Galli‡§, Sebastian Bader†, *Senior Member, IEEE*, Leonardo Luis Giovanini* and Hugo Leonardo Rufiner*¶

*Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), FICH-UNL/CONICET, 3000 Santa Fe, Argentina

† Department of Computer and Electrical Engineering, Mid Sweden University, Sundsvall, Sweden

‡ Facultad de Ciencias Agrarias, Univ. Nacional de Rosario, Zavalla, Argentina

§ Inst. de Inv. en Cs Agr. de Rosario, IICAR, FCA, UNR-CONICET, S2125 Zavalla, Argentina

¶ Laboratorio de Cibernética, Facultad de Ingeniería, Univ. Nacional de Entre Ríos, Oro Verde 3100, Argentina

mferrero@sinc.unl.edu.ar

***Abstract*—Sustainable agricultural practices, crucial for climate resilience, demand advanced internet of things (IoT) and artificial intelligence (AI) solutions for efficient resource management. This work addresses this challenge within the domain of livestock farming. Precision livestock farming offers significant potential for sustainable agriculture by optimizing resource use and improving animal welfare. Automated monitoring of livestock behaviors, such as feeding, is crucial for supporting animal welfare and improving sustainable milk and meat production. Cattle monitoring systems based on machine learning have been developed over the past decade. Processing sound, movement, or both combined using deep learning (DL) models reached good performance. However, training DL models often requires large labeled datasets, which are challenging to obtain, potentially limiting the model's capability for generalization. Transfer learning (TL) has been identified as a highly effective methodology for diminishing the requirement for labeled training data. This paper explores the application of TL and information fusion techniques to enhance the performance of deep learning models for recognizing masticatory events in cattle using acoustic and inertial sensor data. Our findings demonstrate that TL can significantly improve model performance, contributing to more robust and data-efficient edge-computing monitoring systems for sustainable livestock management.**

***Index Terms*—Transfer learning, deep learning, information fusion, sustainable precision livestock farming, ruminant foraging behavior, internet of things.**

I. Introduction

The global demand for food necessitates more efficient and sustainable agricultural practices. Precision Livestock Farming (PLF) leverages advanced technologies to monitor and manage livestock individually, leading to optimized resource utilization, early disease detection, and enhanced animal welfare [1]. As part of this, Internet of Things (IoT) technologies play a crucial role in monitoring individual livestock behaviors by providing continuous, remotely transmitted data from connected sensors attached to the animal.

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
M. Ferrero, L. Rau, L. D. Vignolo, J. O. Chelotti, J. Galli, S. Bader, L. Giovanini & H. L. Rufiner; "Transfer Learning and Sensor Fusion for Cattle Monitoring: An IoT-Driven Approach for Sustainable Livestock Farming"
2025 IEEE 11th World Forum on Internet of Things (WF-IoT), 2025.

A key aspect of PLF is the automated monitoring of feeding behaviors in ruminants, such as grazing and rumination, which are composed of sequences of masticatory events (Fig. 1). These activities occupy 40%–80% of the daily time budget of free-ranging cattle [2]. A grazing bout consists of a variable sequence of three masticatory events: bites, chews (or grazing-chews), and chew-bites (Fig. 1.c) [3]. Bite events involve the apprehension and severing of herbage, grazing-chew events process previously gathered material, and chew-bites combine both actions in a single event. Rumination bouts comprise multiple consecutive episodes with a pattern that begins by regurgitating previously ingested cud, followed by chew (or rumination-chew) events, and ends with swallowing the cud (Fig. 1.d) [4].

Masticatory events recognition provides actionable data for both resource management and animal welfare. Accurately classifying these masticatory events is essential for reliably estimating actual pasture intake, a key metric for optimizing soil use, avoiding over-grazing, and designing efficient pasture rotation strategies. Furthermore, tracking changes in the type and sequence of these events over time can help characterize and quantify foraging behavior [5]. Variations in the duration and frequency of these feeding behaviors are valuable indicators of animal health, stress states, and can even serve as predictors of estrus and parturition. Therefore, this granular level of monitoring directly supports sustainable production [6].

Deep learning (DL) models, particularly those using convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have emerged as powerful tools to recognize these events by analyzing acoustic and/or inertial data. However, training these complex models typically requires substantial amounts of labeled data, which is often scarce and laborious to acquire in real-world farm environments [7]. Transfer learning (TL) presents a promising solution to data scarcity by leveraging knowledge from a source task, where data is often more abundant, to enhance model performance in a target task with limited data, improving model performance and generalization.

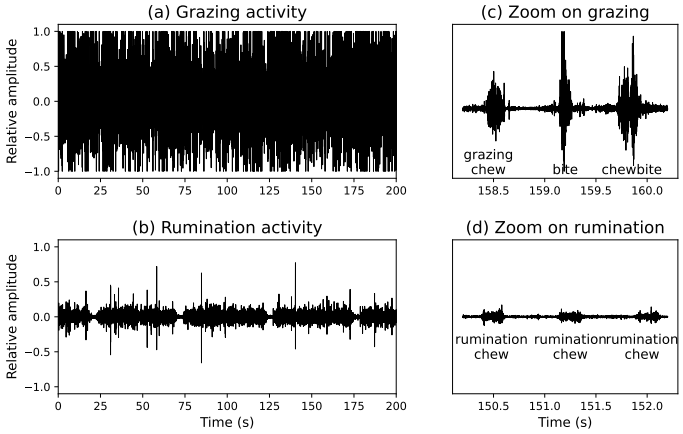

Fig. 1. Waveform of feeding activities and associated masticatory events.

This paper investigates the application of TL for the recognition of masticatory events in free-ranging cattle using data from acoustic and inertial sensors. We demonstrate how pre-trained models can be adapted to new, related tasks, thereby enhancing the accuracy and robustness of feeding behavior monitoring systems. This contributes to the development of more sustainable and efficient IoT-based PLF solutions, aligning with artificial intelligence-driven sustainability goals and smart sensing for resource optimization.

The rest of this paper is organized as follows: Section 2 provides a review of the background and related work in automated livestock monitoring, detailing the evolution from traditional methods to sensor-based systems using machine learning and DL, and establishing the context for applying TL. Section 3 describes our methodology, including the architecture of the base model, the specific TL approach designed to overcome data limitations, and a detailed account of the source and target domain datasets. In Section 4, we present the experiments and results, outlining the training process and evaluating the performance of our proposed model under various conditions. Finally, Section 5 concludes the paper by summarizing our key findings, discussing the implications for PLF, and suggesting directions for future research.

II. Background and related work

Automated monitoring of livestock behaviors, particularly feeding, has seen increasing interest. Traditional methods often rely on direct observation, which is time-consuming and impractical for large herds. IoT-based approaches, utilizing accelerometers, microphones, and other sensors, offer a scalable alternative.

Motion sensors, particularly accelerometers, were widely used to identify ruminant behaviors like grazing and rumination due to their low cost, power consumption, and the minimal computational power required to process their low-frequency signals [8]. However, these sensors face challenges in discriminating specific jaw movement (JM) events [9]. In contrast, acoustic sensors have proven to be a valuable tool, as microphones positioned on the head can capture sounds transmitted through bone to recognize JM events and overall foraging activities precisely [3], [4]. The primary drawbacks of acoustic monitoring are the significant challenges posed by environmental noise and the higher computational and data volume requirements compared to motion-based systems.

Regarding the creation of automatic systems capable of recognizing and classifying masticatory events and activities, machine learning techniques are the most studied [6]. The most commonly used proposals follow a classic scheme of recognition in stages: pre-processing, extraction of characteristics, and classification. However, certain limitations in the analysis of the signals, and the classification of activities are observed [10]. One of the main disadvantages of this approach is the need to manually specify the variables that will serve as input to the models. This introduces a challenge because in this particular problem, there is no consensus on which features to use [6].

DL methods have been successfully applied to classify ingestive events of cattle from sensor data. For instance, [7] explored different architectures using acoustic signals, and later multimodal fusion architectures were proposed to combine acoustic signals with movement signals (accelerometer, gyroscope, and magnetometer) [11]. These models are able to learn hierarchical features directly from raw or minimally processed sensor data, reducing the need for manual feature engineering. Nevertheless, the development of these models is contingent on the availability of substantial quantities of labeled data. In addressing this challenge, TL techniques emerge as a promising solution. As defined by Pan and Yang [12], TL involves transferring knowledge from one or more source tasks to improve learning in a related target task. TL has been widely used in various fields, including sentiment analysis [13], image classification [14], and human activity recognition [15]. The core idea is that features learned by a model on a large, general dataset can benefit a more specific task where labeled data is limited.

The application of TL in animal behavior research is significantly hampered by the absence of foundational models comparable to those available in other domains. Fields like computer vision have been revolutionized by DL models pre-trained on vast datasets such as ImageNet [16], which serve as a powerful, general-purpose starting point for a multitude of specific tasks. In contrast, the domain of acoustic and movement-based behavior detection lacks such resources. In scenarios like this, experimentation with TL techniques involves the selection of source domain training sets and the models to be used.

III. Methodology

A. Base model and transfer learning approach

The primary goal is the accurate recognition of individual masticatory events, which implies the tasks of detection and classification. The selected base model is the architecture presented by Ferrero *et al.* [11], the first end-to-end fusion model capable of working with acoustic and movement signals for masticatory events recognition (Fig. 2). This model has been shown to outperform unimodal DL models [7], [17]. This architecture presents three main blocks: 1) different CNN heads for feature extraction from each sensor signal (acoustic

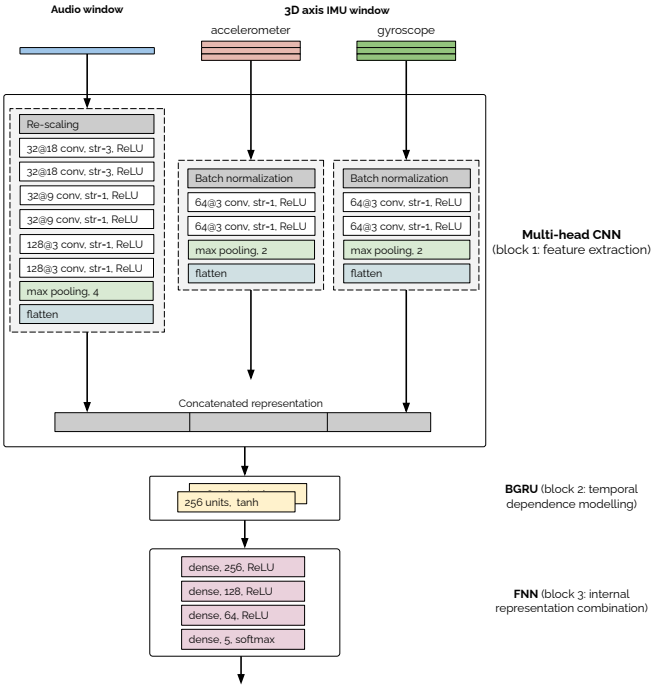

Fig. 2. Multi-head CNN proposed by Ferrero *et al.* [11] to perform information fusion using microphone and IMU signals.

and movements); 2) RNNs composed of bi-directional gated recurrent unit (BGRU) layers to capture temporal dependencies; 3) fully connected neural network (FNN) layers to combine the previously learned representations and predict a final label.

Given the lack of open datasets in the area of interest that include simultaneous audio and movement signals, applying TL directly to the base model was impossible. Consequently, the selected approach using the base model involved training the CNN heads for the acoustic and accelerometer signals independently on separate datasets. Consequently, a new architecture was created for each signal with this goal in mind, utilizing the corresponding CNN head followed by three dense layers for final classification (Fig. 3). These layers used ReLU activation, except for the last one, which employed softmax.

Once the two new architectures were trained in the source domain, the learned CNN weights were transferred to the base model used for classification on the target domain (Fig. 4).

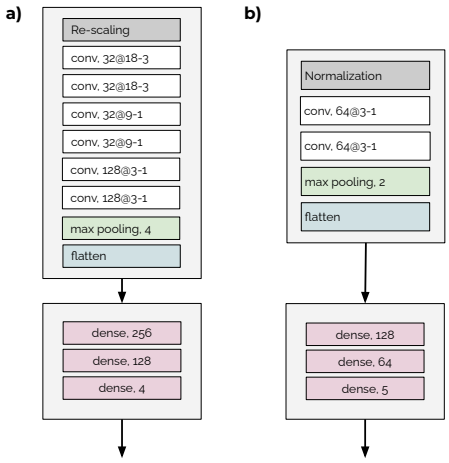

Fig. 3. Architectures extracted from the base model to perform training in the source domain.

B. Datasets

1) Source domains: The source domains selected for training the CNN heads for the acoustic and accelerometer signals were independent. Two publicly available datasets were used to train the CNN designed for processing sound signals, ensuring that the source and target domains have similar characteristics. The first dataset contains grazing sounds of 3 dairy cows collected in Zavalla, Argentina. The dataset comprises 52 audio signals (WAV, 22.05 kHz) featuring 1,617 masticatory events (chew, bite, and chew-bite) labeled by two experts [18]. This dataset was previously employed by Ferrero *et al.* [7] in their reported experimentation, who refined the original labels by eroding the imprecise temporal boundaries based on the signal envelope. Subsequently, the authors resampled the signals to 6 kHz and segmented them into 300 ms windows for analysis, with labels assigned to windows if there was at least a 40% overlap with an event. In this study, we follow the same label refinement and preprocessing procedure.

The second dataset was collected during fieldwork between July 31 and August 19, 2014, at Michigan State University's W. K. Kellogg Biological Station [19]. Audio recordings were made from 5 Holstein cows as part of a voluntary milking sys-

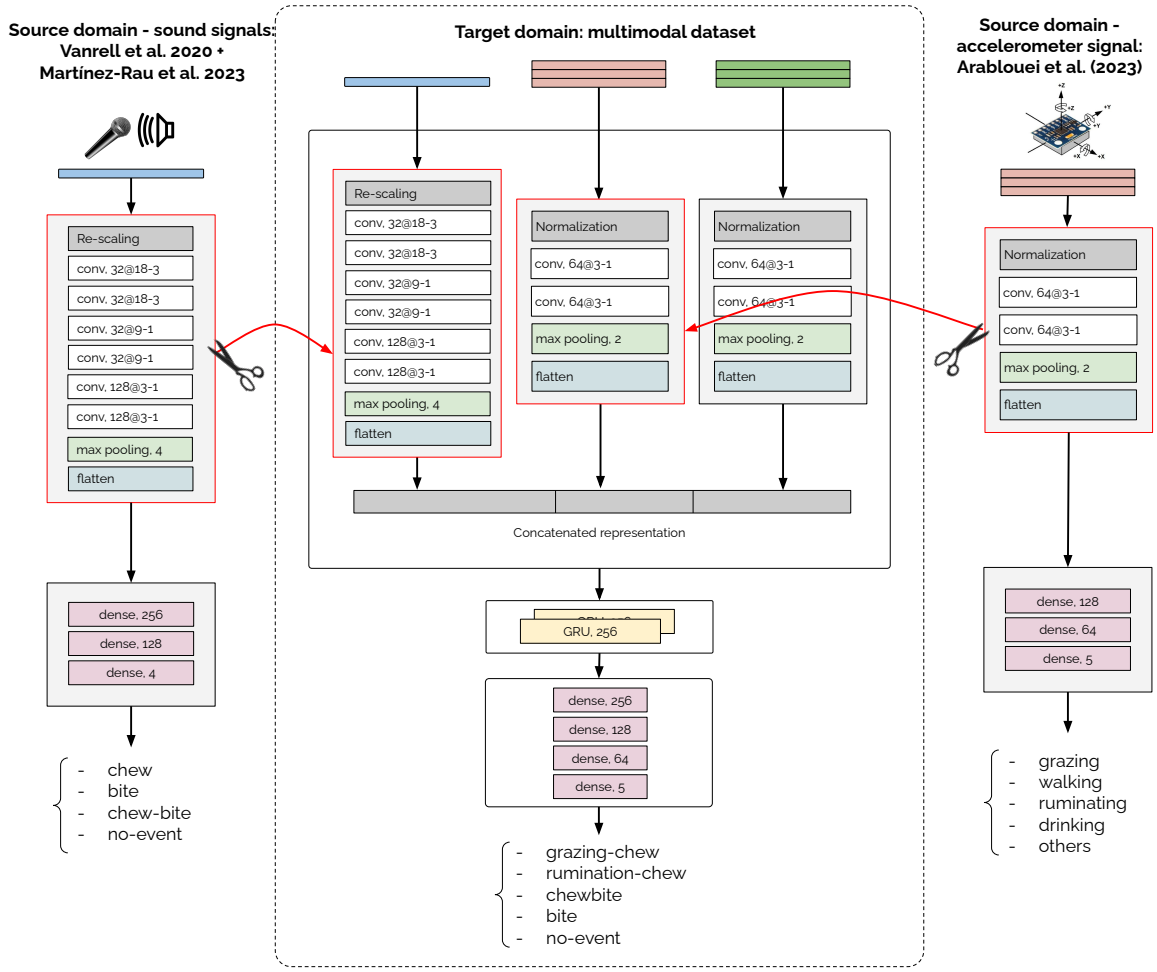

Fig. 4. TL scheme employed, showing an example where all layers belonging to the base sound and motion models (included in the red box) are trained in the source domain, and then the weights are kept fixed during training in the target domain.

tem. Data were collected during five different 24-hour periods using Sony digital recorders with two directional microphones on each animal's forehead (one facing inward, one outward). Ten 5-minute labeled audio signals (WAV format, mono, 16-bit, 44.1 kHz) were utilized, totaling 6,227 events.

The Arm20c dataset presented by Arablouei *et al.* [20] was selected as the source domain for the CNN tasked with motion data processing. It contains accelerometer records from 8 animals under grazing conditions, recorded at a sampling frequency of 50 Hz. The accelerometer was placed on the animal's neck, which is relevant to the target domain data of this work. The data were divided into 5.12-second windows and manually labeled through the observation of video recordings. The 5 classes considered were: grazing, walking, ruminating/resting, drinking, and others. A total of 11,961 windows were recorded.

2) Target domain: The dataset used by Ferrero *et al.* [11] was selected as the target domain in this experimentation. To collect this data, a Moto G6 smartphone, housed in a plastic case, was attached to a collar placed behind the animal's head. An external microphone captured audio signals (AAC format, 44.1 kHz, 128 kbps, mono), while the phone's internal inertial measurement unit (IMU) recorded 3D movement data at 100 Hz over a 5.5-hour experiment. From this, 29 segments

TABLE I
COMPARISON BETWEEN MAIN ASPECTS OF SIGNALS FROM SOURCE AND TARGET DOMAINS.

Aspect	Source	Target
Dataset	[18]–[20]	[11]
Cattle breed	Holstein cows	Holstein cows
Microphone location	Forehead	Forehead
IMU location	Neck	Neck
Accelerometer	✓	✓
Gyroscope	×	✓
Sound	✓	✓
Sound frequency	22.05/44.1 kHz	44.1 kHz
IMU frequency	50 Hz	100 Hz

of specific feeding activities (grazing or rumination), averaging 9 minutes 31 seconds duration, were manually extracted and labeled. Two trained individuals independently used Audacity software to annotate 18,495 masticatory events with one of four labels (bite, grazing-chew, rumination-chew, or chew-bite), achieving a total agreement of 97.63%; discrepancies were resolved through joint discussion. Main characteristics of the source and target domains are summarized in Table I.

C. Base model training methodology

From both motion heads, only the accelerometer one (Fig. 3.a) was trained in the source domain, as the available motion dataset does not include a gyroscope. From each 5.12-second window, 17 non-overlapping 0.3-second sub-windows were extracted, each assigned the original window's label, with a sampling frequency of 50 Hz. The gyroscope architecture was initialized with random weights in the target domain.

The same procedure for the sound signal in the target domain, involving 0.3-second windows at a 6 kHz sampling frequency with 50% overlap, was performed in the source domain to match the acoustic datasets. Each network shown in Fig. 3 was trained using the Adagrad algorithm, sparse categorical cross-entropy cost function for 2000 epochs. The optimal batch size for each CNN was determined through preliminary experiments, being 1,000 samples for the accelerometer CNN and 50 samples for the sound CNN.

IV. EXPERIMENTS AND RESULTS

After training the base models in their respective source domains, the proposed multi-head model was trained. The weights of the CNNs in each "head" were initialized using the values from the respective two pre-trained base models (Fig. 3), transferring learning from one domain to another. However, due to the unavailability of data in the source domain, the CNN layers responsible for processing the gyroscope signal were initialized randomly, following the traditional approach. The frequency of the IMU signals from the target domain was decimated from 100 Hz to 50 Hz to align the number of samples in each window between both domains.

To investigate the impact of retraining the weights in the target domain instead of simply using the weights trained in the source domain, and also to evaluate the benefits of using TL, several experiments were conducted. In each experiment, a specific number of convolutional layers from the base model were retrained (counting backward from the end), while the remaining layers stayed invariant. A curated set of the most promising layer-retraining combinations was selected, which allowed for a practical yet insightful analysis. In the target domain, 24 labeled segments were used for cross-validation, creating 5 folds, each with 4 or 5 segments, always including one rumination segment and the rest from grazing. This data separation remained consistent across all experiments. F1 score, precision, recall, and error rate were calculated by analyzing labels and temporal event boundaries with consistent partitions in each iteration. In contrast to the results presented in [11], a 50 Hz sampling frequency was selected in this article to match the motion signal's source domain characteristics. However, this specific value was incompatible with our selected time window, rendering an overlap strategy unfeasible in our experimental setup and highlighting a key difference in our approach. The aforementioned two conditions prevent the

results from [11] from serving as a baseline for comparison. Despite these facts, the primary focus of this article is to assess whether the use of TL techniques can improve the training of multimodal DL models for JM event classification under limited training data conditions.

The results presented in Table II consistently showed improvements when applying TL compared to using a base model with entirely random weight initialization in the target domain. The best performance across all metrics was achieved when all layers of the acoustic CNN head and none of the layers of the accelerometer CNN head were retrained. This may be due to the fact that the relationship between source and target domains for sound data is weaker compared to the case of accelerometer data, and, on the other hand, more available information allows for better learning of specific features, leading to more accurate classification. Another interpretation of the resulted presented in Table II, is provided by the fact that acoustic source and target datasets varied in cattle breed, and recording equipment, creating a significant "domain gap". This likely explains why retraining all the acoustic layers was optimal, as it allowed the model to adapt to the new data. Conversely, the accelerometer source data taught the model to recognize general movements like walking. These foundational motion features could be generic enough to be useful for detecting the more subtle JM events in target domain, which is why not retraining these layers proved effective.

The use of TL resulted in a lower standard deviation compared with traditional random weights initialization. For instance, when all layers of both CNN heads remained invariant during target domain training, the F1 score's standard deviation dropped from 0.22 to 0.06, a 27% reduction, with similar trends across other metrics. This is likely because reducing the number of trainable parameters limits the model's capacity to learn new, highly specific features from each signal. Additionally, an increase in the standard deviation is presented as the value of the metric increases.

Compared to unimodal DL models or traditional machine learning methods, the improved performance of this multimodal model comes at the cost of increased complexity and a higher number of parameters. This introduces challenges when deploying such models on resource-constrained devices. Ideally, these devices should run all or part of the model locally, reducing the need for remote data transmission, which is often limited in rural areas. One potential solution is to implement the model on the MAX78000 family of microcontrollers (MCUs) (Analog Devices, Wilmington, MA, USA), which features a flexible and customizable CNN hardware accelerator, along with significant memory and processor resources [21]. This solution is also affordable for this application. The specific implementation would depend on the energy budget and model performance trade-offs. For instance, the model could be compressed through deep quantization (1-, 2-, or 4-bits) and deployed on a single MCU, or each data modality and its corresponding CNN, with lighter compression (8-bits), could run on separate MCUs.

TABLE II

AVERAGE AND STANDARD DEVIATION RESULTS PER PARTITION BASED ON THE NUMBER OF CONVOLUTIONAL LAYERS WHOSE WEIGHTS WERE RETRAINED IN THE TARGET DOMAIN: CA FOR ACCELEROMETER AND CS FOR SOUND.

CA	CS	F1 score	Precision	Recall	Error rate
Base		0.45 ± 0.22	0.44 ± 0.20	0.47 ± 0.23	0.89 ± 0.19
0	0	0.48 ± 0.06	0.49 ± 0.05	0.47 ± 0.07	0.81 ± 0.05
1	1	0.48 ± 0.05	0.49 ± 0.04	0.47 ± 0.07	0.84 ± 0.05
1	2	0.46 ± 0.05	0.47 ± 0.03	0.45 ± 0.07	0.83 ± 0.04
1	3	0.47 ± 0.09	0.48 ± 0.09	0.46 ± 0.09	0.83 ± 0.14
2	6	0.48 ± 0.09	0.49 ± 0.10	0.47 ± 0.09	0.84 ± 0.17
0	6	**0.49 ± 0.16**	**0.50 ± 0.14**	**0.49 ± 0.16**	**0.77 ± 0.21**

V. CONCLUSION

This study successfully demonstrated the application and benefits of TL for enhancing the automated recognition of masticatory events in cattle using DL models in multi-sensor data. The findings indicate that by transferring knowledge from

pre-trained models, significant improvements in classification performance can be achieved, especially when labeled data for the target task is limited.

This work has direct implications for advancing PLF, making monitoring systems more accurate and confident. Such improvements are vital for promoting smart and sustainable agricultural practices through better resource management and animal welfare monitoring. The use of artificial intelligence techniques like TL in IoT-enabled sensor networks in agriculture paves the way for smarter, more sustainable food production systems. Deploying the model on low-power edge devices is not merely a technical feasibility but a strategic choice for green and energy-efficient communication. By processing data locally, this approach drastically reduces the energy costs and bandwidth requirements associated with continuous data streaming to the cloud, a critical factor for scalable and sustainable IoT deployments in remote agricultural settings.

In conclusion, applying TL techniques in the context of masticatory event recognition using acoustic and motion signals can lead to more precise classification with less variability. Future work could explore alternative data sources from other related source domains and transfer knowledge from simulated environments, further addressing data scarcity in PLF applications. Another limitation of this work was that the gyroscope CNN head was randomly initialized due to a lack of source data. Exploring different source datasets that also include this sensor might be of interest to evaluate.

ACKNOWLEDGEMENT

This work has been funded by Universidad Nacional del Litoral, CAID 50620190100080LI and 50620190100151LI, Universidad Nacional de Rosario, projects 2013-AGR216, 2016-AGR266 and 80020180300053UR, Agencia Santafesina de Ciencia, Tecnología e Innovación (ASACTEI), project IO-2018–00082, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), project 2017-PUE sinc(i), and Knowledge Foundation, grant number 20180170 (NIIT). The authors would also like to thank the CSIRO staff for sharing the datasets used in their research.

REFERENCES

[1] J. Capper, "The environmental impact of dairy and beef production: Improving productivity offers mitigation opportunities," *Cattle Pract.*, vol. 19, pp. 137–140, 11 2011.

[2] R. J. Kilgour, "In pursuit of "normal": A review of the behaviour of cattle at pasture," *Appl. Anim. Behav. Sci.*, vol. 138, no. 1, pp. 1–11, 2012.

[3] E. D. Ungar, N. Ravid, T. Zada, E. Ben-Moshe, R. Yonatan, H. Baram, and A. Genizi, "The implications of compound chew–bite jaw movements for bite rate in grazing cattle," *Appl. Anim. Behav. Sci.*, vol. 98, no. 3, pp. 183–195, 2006.

[4] J. R. Galli, D. H. Milone, C. A. Cangiano, C. E. Martínez, E. A. Laca, J. O. Chelotti, and H. L. R. and, "Discriminative power of acoustic features for jaw movement classification in cattle and sheep," *Bioacoustics*, vol. 29, no. 5, pp. 602–616, 2020.

[5] A. Andriamandroso, J. Bindelle, B. Mercatoris, and F. Lebeau, "A review on the use of sensors to monitor cattle jaw movements and behavior when grazing," *Biotechnol. Agron. Soc. Environ.*, vol. 20, June 2016.

[6] J. O. Chelotti, L. S. Martinez-Rau, M. Ferrero, L. D. Vignolo, J. R. Galli, A. M. Planisich, H. L. Rufiner, and L. L. Giovanini, "Livestock feeding behaviour: A review on automated systems for ruminant monitoring," *Biosyst. Eng.*, vol. 246, pp. 150–177, 2024.

[7] M. Ferrero, L. D. Vignolo, S. R. Vanrell, L. S. Martinez-Rau, J. O. Chelotti, J. R. Galli, L. L. Giovanini, and H. L. Rufiner, "A full end-to-end deep approach for detecting and classifying jaw movements from acoustic signals in grazing cattle," *Eng. Appl. Artif. Intell.*, vol. 121, p. 106016, 2023.

[8] C. Aquilani, A. Confessore, R. Bozzi, F. Sirtori, and C. Pugliese, "Review: Precision livestock farming technologies in pasture-based livestock systems," *Animal*, vol. 16, no. 1, p. 100429, 2022.

[9] J. O. Chelotti, S. R. Vanrell, L. S. Martinez-Rau, J. R. Galli, S. A. Utsumi, A. M. Planisich, S. A. Almirón, D. H. Milone, L. L. Giovanini, and H. L. Rufiner, "Using segment-based features of jaw movements to recognise foraging activities in grazing cattle," *Biosyst. Eng.*, vol. 229, pp. 69–84, 2023.

[10] J. O. Chelotti, S. R. Vanrell, J. R. Galli, L. L. Giovanini, and H. L. Rufiner, "A pattern recognition approach for detecting and classifying jaw movements in grazing cattle," *Comput. Electron. Agric.*, vol. 145, pp. 83–91, 2018.

[11] M. Ferrero, J. O. Chelotti, L. S. Martinez-Rau, L. Vignolo, M. Pires, J. R. Galli, L. L. Giovanini, and H. L. Rufiner, "A multi-head deep fusion model for recognition of cattle foraging events using sound and movement signals," *Eng. Appl. Artif. Intell.*, vol. 157, p. 111372, 2025.

[12] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[13] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC Med. Imaging*, vol. 22, no. 1, p. 69, 2022.

[14] A. Ray, M. H. Kolekar, R. Balasubramanian, and A. Hafiane, "Transfer learning enhanced vision-based human activity recognition: A decade-long analysis," *Int. J. Inf. Manag. Data Insights*, vol. 3, no. 1, p. 100142, 2023.

[15] M. Shaha and M. Pawar, "Transfer learning for image classification," in *2018 Second Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2018, pp. 656–660.

[16] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[17] 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, 2023.

[18] S. R. Vanrell, J. O. Chelotti, L. A. Bugnon, H. L. Rufiner, D. H. Milone, E. A. Laca, and J. R. Galli, "Audio recordings dataset of grazing jaw movements in dairy cattle," *Data Brief*, vol. 30, p. 105623, Jun. 2020.

[19] L. S. Martinez-Rau, J. O. Chelotti, M. Ferrero, S. A. Utsumi, A. M. Planisich, L. D. Vignolo, L. L. Giovanini, H. L. Rufiner, and J. R. Galli, "Daylong acoustic recordings of grazing and rumination activities in dairy cows," *Sci. Data*, vol. 10, no. 1, p. 782, Nov. 2023.

[20] R. Arablouei, L. Wang, C. Phillips, L. Currie, J. Yates, and G. Bishop-Hurley, "In-situ animal behavior classification using knowledge distillation and fixed-point quantization," *Smart Agric. Technol.*, vol. 4, p. 100159, 2023.

[21] Analog Devices, "Ultra-low power artificial intelligence (AI) MCUs," accessed: 2025-06-12. [Online]. Available: https://www.analog.com/en/product-category/ultralow-power-artificial-intelligence-ai-mcus.html