Detecting the Number of Bite Prehension of Ggrazing Cows in an Extensive System Using an Audio Recording Method

Roberta Avanzato[1], Marcella Avondo[2], Francesco Beritelli[1], Francesco Di Franco[1] and Serena Tumino[2]

[1]*Department of Electrical, Electronic and Computer Engineering University of Catania, Catania, Italy*
[2]*Department of Agricultural, Food and Environmental Science University of Catania, Catania, Italy*

Abstract
In the context of cattle farming, understanding the feeding behavior of animals is essential to ensure their welfare and maximize productivity. However, monitoring and interpret- ing the acoustic signals associated with grazing, particularly the sound events related to grass intake, pose a significant challenge. This study proposes an innovative method based on 1D convolutional neural networks to automatically classify such sound events during grazing. The approach was developed using a balanced dataset composed of 322 prehension samples and 1000 non-prehension samples, extracted from audio recordings of grazing cattle in real conditions. The results obtained show a high accuracy of 100% during the testing and validation phases of the model. However, there is concern about overfitting of the model due to the limited size of the dataset used. Consequently, future expansion of the dataset is suggested by collecting a larger and more diverse number of audio recordings to improve the generalization and robustness of the model in real cattle farming contexts.

Keywords
Precision Livestock Farming, Prehension detection, Audio Signal Analysis, Automatic Classification, 1D Convolutional Neural Networks

1. Introduction

Recent advances in automated monitoring systems have opened up new possibilities for precision breeding, including the radio signal power [1, 2] and computer vision to recognize cow behavior and location within the barn [3, 4, 5].

The feeding behaviour of grazing animals is an aspect whose knowledge represents an added value in understanding the methods of direct use of the herbage by the animals and the consequent implications in terms of pasture response and production (milk, growth) and qualitative characteristics of products (nutritional, nutraceutical and sensorial aspects). Parameters such as the herbage intake, the number of prehensions, the rumination activity are characterized by objective detection difficulties, taking into account the complete freedom of behaviour of grazing animals, especially in extensive conditions. The traditional method consists of direct observations of the herbage prehension behaviour but also, with greater executive difficulty, of the number of bites made per minute, the time dedicated to eating and the time dedicated to ruminating.

To overcome the executive difficulties that characterize this method, several techniques for the automatic detection of some grazing behavioural parameters have been developed.

Among the first automatic detection systems were those based on the 24h-recording of chewing movements by pressure sensors nose bands [6, 7].

In more recent years the use of accelerometers is probably the most adopted precision farming practice, being able to detect changes in the position of the neck, head and mouth [8, 9].

This allows detecting activities such as prehension, chewing, rumination, searching, lying, but also urination and defecation if accelerometers are attached to various points along the back of the spine or in the tail as illustrated by the studies of Marsden and Shorten et al. [10, 11].

However, [12] the accelerometers cannot easily identify the individual herbage bite prehensions made by the animals in part explainable by undesirable signals during recording sessions due to head movements not related to grazing activity.

In [13], the authors, demonstrated that acoustic sensors attached to the hind leg can differentiate seven behaviors of cattle (Grazing, Breathing, Walking, Lying down, Defecating, Vocalizing, Other) with an accuracy of 96.2%. Acoustic technology offers non-invasive alternatives to monitor the welfare and behavior of cows, including estimating breathing during sleep and quantifying the duration of defecation events.

Determining the feeding behaviors of dairy cows is crucial for assessing their productivity and health status. Various research contributes to progress in livestock management, providing a basis for the development of more precise and reliable decision support tools.

Audio systems represent an additional mean available for the detection of grazing behavior. The recording of sounds, suitably codified through direct observations or video recordings, can be an effective mean of characterizing eating behavior, as the activities of bite prehension, chewing, rumination can be recognized on the basis of the sound frequencies recorded over the day at pasture [14, 15, 16, 17, 18, 19, 20].

The studies conducted by Chelotti et al. presented an analysis system called Real-Time Chews-Boluses Recognition Algorithm (CBRTA) [14], which operates completely automatically in real-time to detect and classify grazing livestock ingestion events, capable of detecting ingestion events with a success rate of 97.4%, while achieving up to 84.0% success in their classification as exclusive chews, boluses, or composite chews-boluses. Additionally, they proposed an algorithm called Jaw Movement Food Activity Recognizer (JMFAR) [19], based on the calculation and analysis of temporal, statistical, and spectral features of jaw movement sounds for the detection of rumination and grazing periods.

Milone et al.'s work [21] demonstrated that the analysis of ingestion events such as chewing and biting allows monitoring and characterizing the grazing behavior of cows and constructing automated methods to decode livestock ingestion sounds; specifically, three types of ingestion events

ICYRIME 2023: 8th International Conference of Yearly Reports on Informatics, Mathematics, and Engineering. Naples, July 28-31, 2023
roberta.avanzato@unict.it (R. Avanzato); mavondo@unict.it (M. Avondo); francesco.beritelli@unict.it (F. Beritelli); francescodifranco@unict.it (F. D. Franco); serena.tumino@unict.it (S. Tumino)

(bites, chews, and chews-boluses) were successfully recognized by cows grazing on tall (24.5 ± 3.8 cm) or short (11.6 ± 1.9 cm) alfalfa grass or grass or fescue hay. Additionally, Vanrell et al. [20] presented a study that calculates amplitude, duration, zero crossings, and envelope symmetry for each raw audio segment corresponding to a grazing cow's jaw movement; the results demonstrated that these chewing events are also useful for constructing automated methods for classifying cow ingestion jaw movements.

In the research conducted by Li et al. [22], deep learning models were evaluated to classify ingestion behaviors (bites, chews, and bite-chews) of dairy cows based on forage characteristics. It was found that ingestion sounds are louder and more prolonged for tall forages. However, although deep learning was effective in classification, further improvements are needed to differentiate behaviors based on forage characteristics.

In this work, the authors propose a study aimed at the identification of a suitable method to classify and quantify the sounds detected during grazing in an extensive system, representing the key aspect for its use in real farm conditions. The objective of the research is to quantify the number of herbage prehensions made by 20 h grazing cows through the development of an audio classifier.

In Subsection 1.1, we will focus on the development and implementation of the proposed methodology for audio event classification in the context of cattle farming. In particular, we present a detailed analysis of the application context and the provided data, as well as describe the development process of the 1D Convolutional Neural Network (CNN) and the creation of the dataset for model training and testing.

In Section 2, we will describe the results obtained following the training and testing phase of the neural network.

Finally, there is the concluding section that gathers the obtained results, study limitations, and future work.

1.1. 1D Convolutional Neural Netwrok

The choice to use a 1D convolutional neural network (CNN 1D), the same as used in the study in [23], for the classification of audio events in the context of cattle farming was driven by the following key factors:

1. **Audio Segment Size for 'Prehension' Class**: Preliminary analysis revealed that audio segments corresponding to the prehension event have a maximum duration of about 350 milliseconds. This specific segment size made the CNN 1D an ideal option for processing and classification, allowing for accurate analysis over short time intervals.
2. **Flexibility in Adding Future Classes**: A crucial aspect of choosing a CNN 1D was its flexibility and scalability. Given the potential need to add new classes of audio events in the future, the convolutional network offers the ability to easily adapt to new data and categories without requiring significant restructuring of the model.
3. **CNN's Notable Performance in Raw Audio Classification**: Convolutional networks are known for their effectiveness in classifying unstructured data, such as images and raw audio. In particular, 1D CNNs have demonstrated excellent performance in classifying audio signals, thanks to their ability to extract significant features from temporal data and effectively handle intra-class variation present in audio data.

1.2. Data Analysis

The data analysis phase included evaluating the quality, variability, and distribution of the provided audio samples. A detailed analysis was conducted to identify the presence of background noise, variation in signal quality among different samples, and consistency in labeling. This thorough examination allowed for establishing a solid foundation for creating a balanced and representative dataset, necessary for the effective training of the network [24, 25, 26, 27, 28].

The approach chosen for audio classification in the context of cattle farming not only addresses the immediate needs of the project but also lays the groundwork for a scalable and adaptable platform for future research and development needs.

1.3. Python Script for the Datasets Creating

To facilitate the training process, a dedicated Python script was developed for creating a structured dataset. This script efficiently organized audio samples in preparation for subsequent training and validation phases. The dataset creation phase was crucial to ensure the effectiveness and accuracy of the audio classification network. It required careful planning and execution, particularly in terms of data preparation and verification.

Initially, the Python script developed for this phase began by reading the CSV file containing the audio data and their corresponding labels. This file served as the backbone of the dataset, providing essential information needed for the subsequent classification process.

A fundamental aspect was verifying the semantic correctness of the timestamps associated with each audio segment. This step was essential to ensure that each label precisely matched the desired audio segment.

To maximize process efficiency and minimize errors, an automated procedure was introduced to pre-examine each label. This verification mechanism was tasked with identifying and flagging any inconsistencies or errors in the labels, such as invalid timestamps or incorrect formats.

In case an error is detected, the procedure specifically reports the row in the CSV file that requires revision. This approach aims to facilitate prompt corrective action, reducing the risk of introducing incorrect data into the final dataset.

Once the dataset has been verified and cleaned, the script proceeds to calculate the maximum and average lengths of the audio segments. This step is crucial for understanding the variation in audio segment sizes and for setting appropriate parameters during the network training phase.

Knowledge of the maximum and average lengths of the audio segments plays a key role in configuring the convolutional network. It influences crucial aspects such as input size, layer structure, and output management, ensuring that the network is optimized to effectively handle the variety of audio segments in the dataset.

1.4. Dataset's Creation

For the creation of the audio classification dataset focused on distinguishing between prehension and non-prehension,

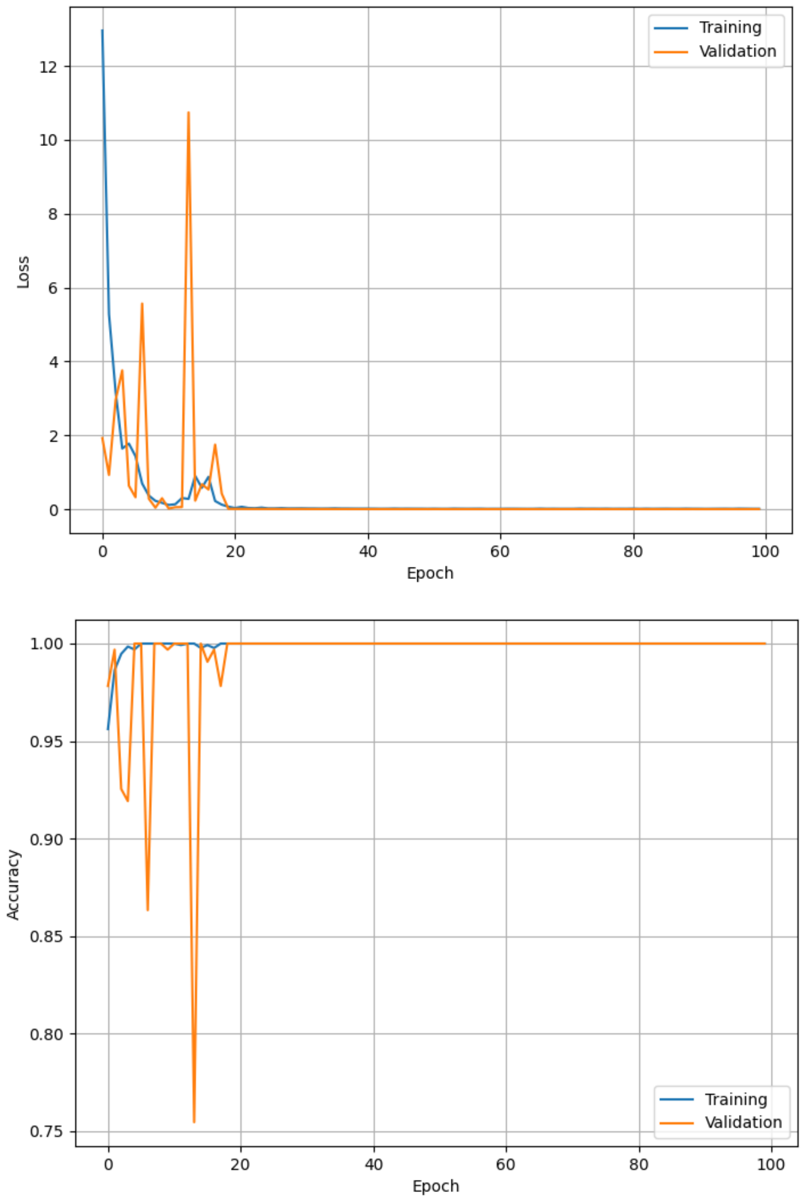

Figure 1: Accuracy and loss trends during training and validation.

specific procedures were followed to ensure the balance and quality of the dataset.

Given the limited nature of available samples, a total of 322 samples were obtained for the prehension class. Considering the abundance of data for the non-prehension class, it was chosen to limit this category to 1000 segments, randomly selected to avoid bias in the dataset.

Due to the discrepancy in the number of samples between the two classes, a strategy was adopted to balance the dataset. This is essential to prevent overfitting or bias towards the more represented class. Therefore, techniques such as downsampling of the most represented class (non-prehension) were employed.

The dataset, in turn, was divided into training, validation, and test sets. This division allows for effective model training, validation to prevent overfitting, and testing its performance on unseen data.

These stages represent a systematic and balanced approach to creating a dataset for an audio classification project, ensuring that the data is representative, balanced, and of high quality.

2. Experimental results

2.1. Network training

For the training phase of the machine learning model, we follow the following approach, keeping in mind the specific context of the project, where we faced the challenge of an imbalanced dataset and a limited number of samples.

Training is conducted for 100 epochs. It was noticed that already in few epochs (Fig. 1), the network had achieved high accuracy. In fact, it is observed that the network reaches 100% accuracy in just 20 epochs, indicating a possible overfitting due to the limited number of samples.

Despite the high accuracy, it is important to emphasize

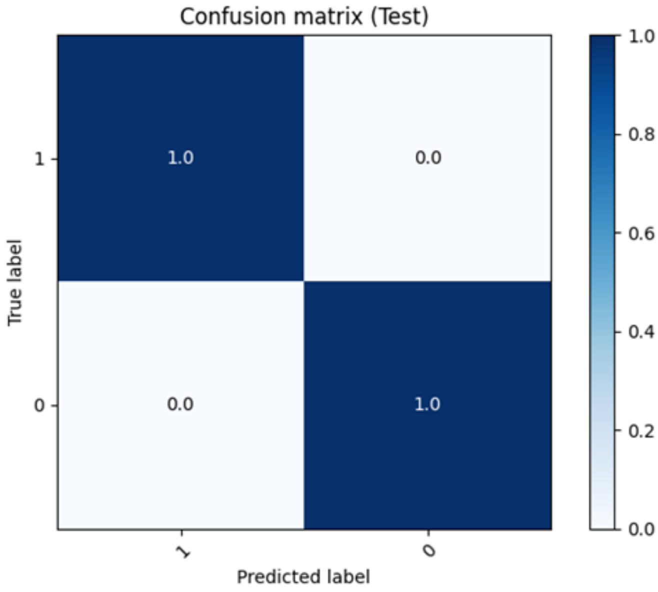

Figure 2: Confusion matrix for the testing dataset.

that this result may not be indicative of the actual performance of the model on unseen data, due to the limited training dataset.

2.2. Testing and Validation

The final phase included thorough testing and validation of the network, using confusion matrices (see Fig. 2) to assess performance. Despite achieving 100% accuracy, this result suggests the possibility of overfitting.

The neural network model demonstrated high accuracy, reaching 100% in both testing and validation phases. However, this exceptionally high result raises concerns regarding the phenomenon of overfitting, where a model overly adapts to the training data, thereby losing the ability to generalize to new data.

3. Conclusions

Despite the promising performance of the classifier, the limited availability of data makes it difficult to determine with certainty whether the network is actually capable of effectively functioning in real-world scenarios or if it is suffering from overfitting. Therefore, acquiring a larger and more diverse dataset is recommended for further testing and to improve the model's generalization. This step is crucial to confirm the classifier's reliability in the real context of cattle farming.

Despite the limitations, the results obtained provide a preliminary indication of the problem's feasibility. To further improve the model, expanding the dataset, utilizing data augmentation techniques, or exploring more complex models and regularization techniques to reduce overfitting are considered.

It is essential to increase the size and variety of the dataset to ensure a more robust evaluation of the model.

Testing various neural network architectures and training parameters to find the optimal configuration is suggested [29].

In conclusion, the training phase has provided significant insights into the feasibility of the project, despite the limitations imposed by the dataset.

Future developments include a significant increase in the dataset size and also the classes to be identified.

Acknowledgments

This study was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

[1] R. Avanzato, F. Beritelli, An innovative technique for identification of missing persons in natural disaster based on drone-femtocell systems, Sensors 19 (2019) 4547.

[2] R. Avanzato, F. Beritelli, A smart uav-femtocell data sensing system for post-earthquake localization of people, IEEE Access 8 (2020) 30262–30270.

[3] R. Avanzato, F. Beritelli, V. F. Puglisi, Dairy cow behavior recognition using computer vision techniques and cnn networks, in: 2022 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS), IEEE, 2022, pp. 122–128.

[4] G. De Magistris, R. Caprari, G. Castro, S. Russo, L. Iocchi, D. Nardi, C. Napoli, Vision-based holistic scene understanding for context-aware human-robot interaction 13196 LNAI (2022) 310 – 325. doi:10.1007/978-3-031-08421-8_21.

[5] S. Pepe, S. Tedeschi, N. Brandizzi, S. Russo, L. Iocchi, C. Napoli, Human attention assessment using a machine learning approach with gan-based data augmentation technique trained using a custom dataset, OBM Neurobiology 6 (2022). doi:10.21926/obm.neurobiol.2204139.

[6] P. Penning, A technique to record automatically some aspects of grazing and ruminating behaviour in sheep, Grass and Forage Science 38 (1983) 89–96.

[7] S. M. Rutter, Graze: a program to analyze recordings of the jaw movements of ruminants, Behavior Research Methods, Instruments, & Computers 32 (2000) 86–92.

[8] J. Werner, C. Umstatter, L. Leso, E. Kennedy, A. Geoghegan, L. Shalloo, M. Schick, B. O'brien, Evaluation and application potential of an accelerometer-based collar device for measuring grazing behavior of dairy cows, Animal 13 (2019) 2070–2079.

[9] G. Pereira, B. Heins, B. O'Brien, A. McDonagh, L. Lidauer, F. Kickinger, Validation of an ear tag–based accelerometer system for detecting grazing behavior of dairy cows, Journal of dairy science 103 (2020) 3529–3544.

[10] K. Marsden, L. Lush, J. Holmberg, I. Harris, M. Whelan, S. Webb, A. King, R. Wilson, D. Jones, A. Charteris, et al., Quantifying the frequency and volume of urine deposition by grazing sheep using tri-axial accelerometers, Animal 15 (2021) 100234.

[11] P. Shorten, B. Welten, Assessment of a non-invasive accelerometer for detecting cattle urination and defecation events, Smart Agricultural Technology 2 (2022) 100031.

[12] V. Giovanetti, M. Decandia, G. Molle, M. Acciaro, M. Mameli, A. Cabiddu, R. Cossu, M. Serra, C. Manca, S. Rassu, et al., Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer, Livestock Science 196 (2017) 42–48.

[13] P. Shorten, Acoustic sensors for detecting cow behaviour, Smart Agricultural Technology 3 (2023) 100071.

[14] J. O. Chelotti, S. R. Vanrell, D. H. Milone, S. A. Utsumi, J. R. Galli, H. L. Rufiner, L. L. Giovanini, A real-time algorithm for acoustic monitoring of ingestive behavior of grazing cattle, Computers and Electronics in Agriculture 127 (2016) 64–75.

[15] A. Alfarano, G. De Magistris, L. Mongelli, S. Russo, J. Starczewski, C. Napoli, A novel convmixer transformer based architecture for violent behavior detection 14126 LNAI (2023) 3 – 16. doi:10.1007/978-3-031-42508-0_1.

[16] J. R. Galli, C. A. Cangiano, M. Pece, M. Larripa, D. H. Milone, S. Utsumi, E. Laca, Monitoring and assessment of ingestive chewing sounds for prediction of herbage intake rate in grazing cattle, Animal 12 (2018) 973–982.

[17] V. Ponzi, S. Russo, V. Bianco, C. Napoli, A. Wajda, Psychoeducative social robots for an healthier lifestyle using artificial intelligence: a case-study, volume 3118, 2021, pp. 26 – 33.

[18] Laca, WallisDeVries, Acoustic measurement of intake and grazing behaviour of cattle, Grass and Forage Science 55 (2000) 97–104.

[19] 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, H. L. Rufiner, Using segment-based features of jaw movements to recognise foraging activities in grazing cattle, Biosystems Engineering 229 (2023) 69–84.

[20] S. R. Vanrell, J. O. Chelotti, L. A. Bugnon, H. L. Rufiner, D. H. Milone, E. A. Laca, J. R. Galli, Audio recordings dataset of grazing jaw movements in dairy cattle, Data in brief 30 (2020) 105623.

[21] D. H. Milone, J. R. Galli, C. A. Cangiano, H. L. Rufiner, E. A. Laca, Automatic recognition of ingestive sounds of cattle based on hidden markov models, Computers and electronics in agriculture 87 (2012) 51–55.

[22] G. Li, Y. Xiong, Q. Du, Z. Shi, R. S. Gates, Classifying ingestive behavior of dairy cows via automatic sound recognition, Sensors 21 (2021) 5231.

[23] W. Dai, C. Dai, S. Qu, J. Li, S. Das, Very deep convolutional neural networks for raw waveforms, in: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, 2017, pp. 421–425.

[24] G. De Magistris, M. Romano, J. Starczewski, C. Napoli, A novel dwt-based encoder for human pose estimation, volume 3360, 2022, pp. 33 – 40.

[25] G. Capizzi, G. L. Sciuto, C. Napoli, M. Woźniak, G. Susi, A spiking neural network-based long-term prediction system for biogas production, Neural Networks 129 (2020) 271–279.

[26] G. D. Magistris, C. Rametta, G. Capizzi, C. Napoli, Fpga implementation of a parallel dds for wide-band applications, volume 3092, 2021, pp. 12 – 16.

[27] G. Capizzi, G. Lo Sciuto, M. Woźniak, R. Damaševicius, A clustering based system for automated oil spill detection by satellite remote sensing, in: Artificial Intelligence and Soft Computing: 15th International Conference, ICAISC 2016, Zakopane, Poland, June 12-16, 2016, Proceedings, Part II 15, Springer, 2016, pp. 613–623.

[28] N. Brandizzi, S. Russo, G. Galati, C. Napoli, Addressing vehicle sharing through behavioral analysis: A solution to user clustering using recency-frequency-monetary and vehicle relocation based on neighborhood splits, Information (Switzerland) 13 (2022). doi:10.3390/info13110511.

[29] G. Capizzi, G. Lo Sciuto, C. Napoli, R. Shikler, M. Woźniak, Optimizing the organic solar cell manufacturing process by means of afm measurements and neural networks, Energies 11 (2018) 1221.