# Goat Vocalizations (VOCAPRA)

Giulia Cuttone
*Department of Computer Science*
*University of Milan*
Via G.Celoria 18, Milano, 20133, Italy
giulia.cuttone1@studenti.unimi.it

***Abstract*—VOCAPRA is a multidisciplinary project aimed at improving dairy goat management through the analysis of goat vocalizations. This paper focuses on classifying goat vocalizations using machine learning techniques. We utilized the VOCAPRA_all dataset, extracting a wide variety of temporal and spectral features. K-means clustering was applied to visualize the feature space, followed by the training and evaluation of two classical classification models—Support Vector Machine (SVM) and Random Forest—alongside a Neural Network model. Our results indicate that all models effectively classify vocalizations, with Neural Network outperforming the others in terms of accuracy.**

***Index Terms*—Goat vocalizations, audio pattern recognition, machine learning, feature extraction.**

## I. Introduction

The monitoring and understanding of animal vocalizations have gained significant attention in recent years, particularly in the context of livestock management. Animal vocalizations can provide important insights into the health, welfare, and behavioral states of farm animals, providing a non-invasive tool for monitoring their well-being. In the context of precision farming, the analysis of vocalizations can enhance herd management by enabling continuous observation of livestock without physical intervention [1]. This study focuses on the classification of goat vocalizations, aiming to help farmers optimize herd management through automated systems that interpret animal sounds, thus improving the overall welfare of livestock.

The importance of this research lies in its potential to advance animal welfare while enhancing productivity. Goats, like many other livestock species, produce a variety of vocal signals that may indicate stress, hunger, illness, or other physiological states [2]. Automated systems capable of analyzing these signals could reduce the need for manual observation, enabling earlier intervention and improving the health outcomes of the animals [3]. In dairy goat farming, for instance, early detection of distress signals may prevent minor health issues from escalating, potentially reducing mortality rates and increasing milk production efficiency [4].

Potential applications of this research include developing systems that can automatically detect vocal signals associated with specific physiological or emotional states, such as pain or anxiety, allowing for timely and targeted interventions. Furthermore, such systems could reduce labor costs and optimize farming practices by integrating into broader precision livestock farming solutions, enhancing the overall sustainability of the farming operation [5].

While considerable work has been done on analyzing animal vocalizations, particularly in species such as cows, pigs, and birds, research on goat vocalization classification is limited [6]. Some studies have focused on using machine learning techniques to classify animal calls or detect distress signals, utilizing features like Mel-frequency cepstral coefficients (MFCCs) and spectral features [7]. However, much of this work remains exploratory, with few studies applying these methods specifically to goats [8].

This study seeks to address this gap by focusing on goat vocalization classification using advanced machine learning techniques. The process begins with the extraction of key acoustic features from the VOCAPRA_all dataset, which serves as the foundation for the analysis. Following feature extraction, k-means clustering is employed to visualize the feature space, providing an initial understanding of the data distribution. Subsequently, two classical classification models—Support Vector Machine (SVM) and Random Forest— and a Neural Network model are trained and evaluated on the extracted features to classify the vocalizations. The performance of these models is then compared to determine their effectiveness in accurately classifying different goat vocalizations, offering insights into the most suitable approach for future automated systems.

## II. Methodology

The methodology consists of three key steps:

1) **Feature extraction:** Temporal and spectral audio features are extracted from the audio recordings to represent the characteristics of goat vocalizations.
2) **Post-processing:** The extracted features are normalized and visualized using k-means clustering to explore the feature space.
3) **Classification:** Support Vector Machine (SVM), Random Forest and Neural Network models are trained and evaluated on the processed features, with their performance compared to assess classification accuracy.


Fig. 1. Block diagram of the goat vocalization algorithm.

### A. *Feature extraction*

The dataset comprises 4,141 *.wav* audio files, already with a sample rate of 16 kHz (16,000 Hz), commonly used in animal vocalizations processing. This sampling rate provides an optimal balance between audio quality and file size, making it well-suited for the task at hand.

Given that the audio files are consistently short and of equal length (2 seconds), pre-processing steps such as gain normalization or windowing are unnecessary.

The extraction of relevant acoustic features for the classification of goat vocalizations focuses on both temporal and spectral characteristics, including:

- **MFCCs (Mel-frequency cepstral coefficients):** 40 coefficients are extracted, representing the short-term power spectrum of sound.
- **Root Mean Square:** Measures the power or energy of the signal.
- **Spectral Centroid:** Indicates where the center of mass of the spectrum is located, often associated with the perceived brightness.
- **Spectral Bandwidth:** Describes the width of the spectrum.
- **Spectral Rolloff:** Describes the frequency below which a certain percentage of the total spectral energy lies.
- **Zero Crossing Rate:** Measures the rate at which the signal changes sign, useful for distinguishing between noisy and harmonic signals.

Example: Fig. 2 presents the spectrogram features corresponding to the "Calori" class.

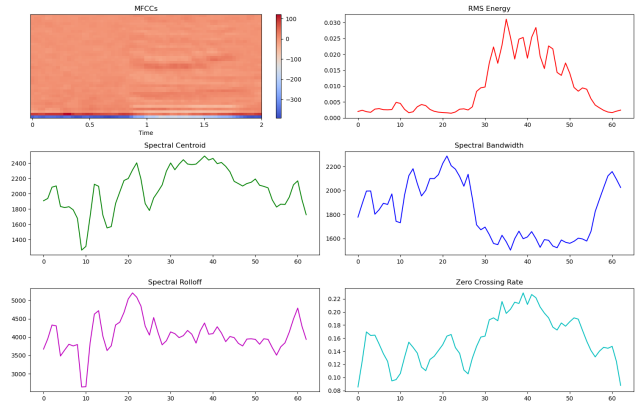


Fig. 2. Temporal and spectral audio features extraction: MFCCs, RMS energy, Spectral centroid, Spectral bandwidth, Spectral rolloff, Zero crossing rate.

The extracted features, along with their corresponding classes (derived from the audio file names), are stored in a DataFrame (Table I).

This DataFrame is then exported as a CSV file named *'Vocapra_dataset.csv'*.

TABLE I
DATAFRAME WITH EXTRACTED FEATURES

| MFCCs_1 | MFCCs_2 | ... | Zero_Crossing_Rate | Class |
|---|---|---|---|---|
| -289.605103 | 60.685940 | ... | 0.164094 | Calori |
| -235.537918 | 72.891380 | ... | 0.143725 | Calori |
| -293.703735 | 58.511467 | ... | 0.177525 | Calori |
| -254.596527 | 66.995712 | ... | 0.144051 | Calori |
| -272.720184 | 60.091812 | ... | 0.170379 | Calori |

### B. *Post-processing*

The post-processing phase includes feature normalization, class restructuring, and the assignment of 'Positive' and 'Negative' labels. For machine learning tasks such as speech recognition or audio classification, Z-score normalization is commonly employed, particularly for features exhibiting high variance (e.g., MFCCs, Spectral Bandwidth).

#### Z-score Normalization

Z-score normalization is a standard method used to standardize features, ensuring that each coefficient has a mean of 0 and a standard deviation of 1. This transformation enhances feature comparability and aids in faster model convergence. The formula for Z-score normalization is as follows:

$$Z = \frac{x - \mu}{\sigma} \tag{1}$$

Where:

- $x$ is the individual feature value
- $\mu$ is the mean of the feature
- $\sigma$ is the standard deviation of the feature

By standardizing features, this technique ensures that all input data is on a similar scale, which is particularly beneficial for optimizing model performance and convergence speed during training.

#### Class Label Assignment

The classes were categorized as follows:

- Calori (calori artificiali, calori naturali)
- Distribuzione Cibo (distribuzione fieno, distribuzione concentrato, distribuzione unifeed)
- Fenomeni legati al parto (doglie del parto, fase esplusiva, parto difficile, aborto)
- Ferita-Morte (ferita, morte capra)
- Isolamento sociale
- Presenza contemporanea di madri e capretti
- Separazione madre capretto
- Visita di estranei

An additional column, 'Emotional_state', was added to the dataframe to classify the dataset's categories into two primary labels: 'Positive' and 'Negative'. These labels are assigned based on the nature of the events or behaviors observed. The 'Negative' label corresponds to events associated with adverse

or unfavorable conditions, while the 'Positive' label represents more favorable or neutral scenarios (see Fig. 5).

The new DataFrame is then exported as a CSV file named 'Vocapra_postprocessing' (Table II).

TABLE II
POST-PROCESSED DATAFRAME

| MFCCs_1 | MFCCs_2 | ... | Zero_Crossing_Rate | Class | Emotional_state |
|---|---|---|---|---|---|
| -0.204031 | -1.781272 | ... | 0.658545 | Calori | Positive |
| 0.513337 | -1.140731 | ... | 0.170917 | Calori | Positive |
| -0.258413 | -1.895388 | ... | 0.980105 | Calori | Positive |
| 0.260466 | -1.450135 | ... | 0.178710 | Calori | Positive |
| 0.019999 | -1.812451 | ... | 0.809027 | Calori | Positive |

### C. *Clustering of goat vocalizations*

Once the features were extracted and organized into a DataFrame, K-Means clustering was employed to analyze and visualize the feature spaces. (Fig. 3) shows the results of a binary clustering, where the data is divided into two main clusters. Building upon this, (Fig. 4) illustrates the second level of classification, where each of the initial clusters is further subdivided, resulting in a total of eight distinct clusters. The first-level clustering (Level 1) achieved an accuracy of 74%, while the second-level clustering (Level 2) reached an accuracy of 39%, indicating a more complex classification in the latter stage.

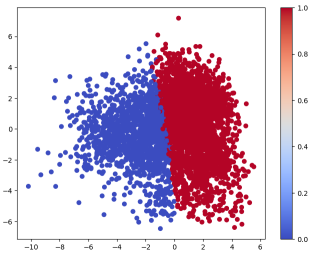

Fig. 3. Level 1: 2 clusters

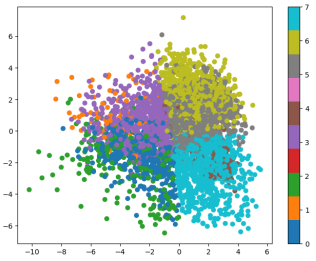

Fig. 4. Level 2: 8 clusters

### D. *Classification*

A hierarchical classifier was implemented to perform the classification task (Fig. 5). The model operates in two levels:

- The first level differentiates between positive and negative states,
- The second level refines this by classifying instances into specific subcategories based on the predictions from the first level.

This approach allows for a more structured and efficient classification process, where broad distinctions are made at the initial level, followed by more granular classifications at the second level.

The post-processing dataset was split into 80% training (3312 rows) and 20% test (829 rows). To ensure balanced class distribution across the splits, 5-fold cross-validation was applied during model training. At the first classification level, a Grid Search was performed, targeting a limited set of hyperparameters to ensure a thorough exploration of key values.

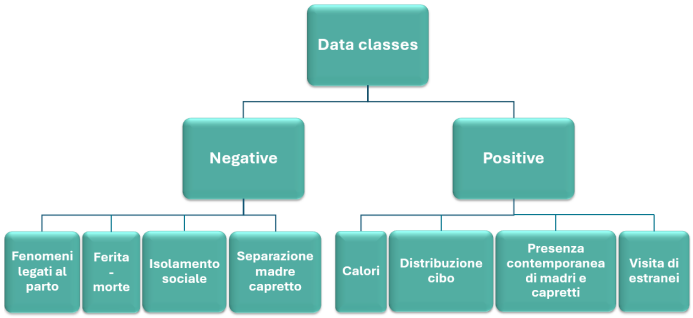


Fig. 5. Hierarchical organization of the classes: the first stage classifies emotions into positive and negative states, while the second stage refines this into the final set of classes.

At the second classification level, where both the number and complexity of hyperparameters increased, a Randomized Search was used. This approach allowed for a broader search across the hyperparameter space, reducing computation time compared to an exhaustive search.
Finally, the model with the optimal hyperparameters was used to make predictions, ensuring the most effective configuration for each classification stage.

## III. EXPERIMENTS AND RESULTS

The classification experiments were conducted using two traditional machine learning models: Support Vector Machine (SVM) and Random Forest; and a Neural Network model. Each model was trained and evaluated on the extracted acoustic features, applying the hierarchical classification approach.

To assess the performance of these models, the following metrics were employed:

- **Accuracy:** The proportion of correct predictions out of the total predictions.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

- **Pecision:** The proportion of true positive predictions out of all positive predictions (focuses on minimizing false positives).

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

- **Recall:** The proportion of true positive predictions out of all actual positive cases (focuses on minimizing false negatives).

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

- **F1-score:** The harmonic mean of precision and recall, balancing both.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

### A. *Support Vector Machine (SVM)*

param_grid = [
{'kernel': ['linear'], 'C': np.logspace(-3, 3, 7)},
{'kernel': ['rbf'], 'C': np.logspace(-3, 3, 7), 'gamma': np.logspace(-3, 3, 7)} ]

- Level 1:
  - **Best Parameters:**
    `C=10.0, gamma=0.01, kernel='rbf'`
  - **Accuracy:** 95%
    (TN = 185, FP = 25, FN = 13, TP = 606)

TABLE III
CLASSIFICATION REPORT LEVEL 1 (SVM)

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Negative | 0.93 | 0.88 | 0.91 | 210 |
| Positive | 0.96 | 0.98 | 0.97 | 619 |
| **Accuracy** | | | 0.95 | 829 |
| **Macro avg** | 0.95 | 0.93 | 0.94 | 829 |
| **Weighted avg** | 0.95 | 0.95 | 0.95 | 829 |

- Level 2:
  - **Best Parameters (Positive):**
    `C=10.0, gamma=0.01, kernel='rbf'`
  - **Best Parameters (Negative):**
    `C=1000.0, gamma=0.01, kernel='rbf'`
  - **Accuracy:** 85%

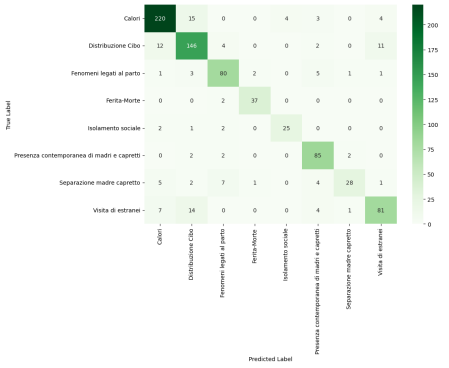


Fig. 6. Level 2 Confusion Matrix (SVM)

TABLE IV
CLASSIFICATION REPORT LEVEL 2 (SVM)

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Calori | 0.89 | 0.89 | 0.89 | 246 |
| Distribuzione Cibo | 0.80 | 0.83 | 0.82 | 175 |
| Fenomeni legati al parto | 0.82 | 0.86 | 0.84 | 93 |
| Ferita-Morte | 0.93 | 0.95 | 0.94 | 39 |
| Isolamento sociale | 0.86 | 0.83 | 0.85 | 30 |
| Presenza contemporanea di madri e capretti | 0.83 | 0.93 | 0.88 | 91 |
| Separazione madre capretto | 0.88 | 0.58 | 0.70 | 48 |
| Visita di estranei | 0.83 | 0.76 | 0.79 | 107 |
| **Accuracy** | | | 0.85 | 829 |
| **Macro avg** | 0.85 | 0.83 | 0.84 | 829 |
| **Weighted avg** | 0.85 | 0.85 | 0.84 | 829 |

### B. *Random Forest (RF)*

- Level 1:
  param_grid = { 'n_estimators': [100, 200, 300], 'max_depth': [20, 30, 50, None] }
  - **Best Parameters:**
    `n_estimators=300, max_depth=30`
  - **Accuracy:** 91%
    (TN = 142, FP = 68, FN = 3, TP = 616)

TABLE V
CLASSIFICATION REPORT LEVEL 1 (RF)

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Negative | 0.98 | 0.68 | 0.80 | 210 |
| Positive | 0.90 | 1.00 | 0.95 | 619 |
| **Accuracy** | | | 0.91 | 829 |
| **Macro avg** | 0.94 | 0.84 | 0.87 | 829 |
| **Weighted avg** | 0.92 | 0.91 | 0.91 | 829 |

- Level 2:
  param_dist = { 'n_estimators': [300, 500, 800], 'max_depth': [20, 30, 50, None], 'min_samples_split': [2, 5, 10] }
  - **Best Parameters (Positive):**
    `n_estimators=500, max_depth=30, min_samples_split=5`
  - **Best Parameters (Negative):**
    `n_estimators=800, max_depth=None, min_samples_split=2`
  - **Accuracy:** 75%


Fig. 7. Level 2 Confusion Matrix (RF)

TABLE VI
CLASSIFICATION REPORT LEVEL 2 (RF)

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Calori | 0.76 | 0.87 | 0.81 | 246 |
| Distribuzione Cibo | 0.62 | 0.78 | 0.69 | 175 |
| Fenomeni legati al parto | 0.80 | 0.68 | 0.73 | 93 |
| Ferita-Morte | 1.00 | 0.95 | 0.97 | 39 |
| Isolamento sociale | 1.00 | 0.53 | 0.70 | 30 |
| Presenza contemporanea di madri e capretti | 0.75 | 0.95 | 0.83 | 91 |
| Separazione madre capretto | 0.85 | 0.23 | 0.36 | 48 |
| Visita di estranei | 0.82 | 0.51 | 0.63 | 107 |
| **Accuracy** | | | 0.75 | 829 |
| **Macro avg** | 0.82 | 0.69 | 0.72 | 829 |
| **Weighted avg** | 0.77 | 0.75 | 0.73 | 829 |

### C. *Neural Network*

- Level 1:
  The model consists of three fully connected layers designed for binary classification. The first and third layers each contain 128 neurons, while the second layer increases the number of neurons to enhance the model's capacity. The output layer consists of a single neuron with a sigmoid activation function, effectively distinguishing between the positive and negative classes.

  param_grid_level1 = { 'batch_size': [16, 32, 64], 'epochs': [120, 200, 250] }

  - **Best Parameters:**

    `epochs=250, batch_size= 16`

  - **Accuracy:** 97%
    (TN = 192, FP = 18, FN = 9, TP = 610)

TABLE VII
CLASSIFICATION REPORT LEVEL 1 (NN)

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Negative | 0.96 | 0.91 | 0.93 | 210 |
| Positive | 0.97 | 0.99 | 0.98 | 619 |
| **Accuracy** | | | 0.97 | 829 |
| **Macro avg** | 0.96 | 0.95 | 0.96 | 829 |
| **Weighted avg** | 0.97 | 0.97 | 0.97 | 829 |

- Level 2:
  This model is flexible and deeper, designed for multi-class classification. It is structured with multiple layers, progressively reducing the number of neurons and incorporating layer-specific dropout, which helps prevent overfitting by limiting the number of parameters at each stage. The output layer consists of four neurons, corresponding to the four subclasses, and uses a softmax activation function to enable multi-class prediction.

  param_grid_level2 = { 'batch_size': [8, 16, 32], 'epochs': [120, 200, 250] }

  - **Best Parameters (Positive):**

    `epochs=200, batch_size=16`

  - **Best Parameters (Negative):**

    `epochs=200, batch_size=32`

  - **Accuracy:** 88%


Fig. 8. Level 2 Confusion Matrix (NN)

TABLE VIII
CLASSIFICATION REPORT LEVEL 2 (NN)

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Calori | 0.93 | 0.95 | 0.94 | 246 |
| Distribuzione Cibo | 0.84 | 0.85 | 0.84 | 175 |
| Fenomeni legati al parto | 0.81 | 0.87 | 0.84 | 93 |
| Ferita-Morte | 0.97 | 0.97 | 0.97 | 39 |
| Isolamento sociale | 1.00 | 0.83 | 0.91 | 30 |
| Presenza contemporanea di madri e capretti | 0.84 | 0.95 | 0.89 | 91 |
| Separazione madre capretto | 0.86 | 0.67 | 0.75 | 48 |
| Visita di estranei | 0.86 | 0.79 | 0.83 | 107 |
| **Accuracy** | | | 0.88 | 829 |
| **Macro avg** | 0.89 | 0.86 | 0.87 | 829 |
| **Weighted avg** | 0.88 | 0.88 | 0.88 | 829 |

## IV. CONCLUSIONS

This study explored the classification of goat vocalizations using machine learning techniques, specifically Support Vector Machine (SVM), Random Forest, and Neural Network classifiers. Each model demonstrated high accuracy, indicating strong performance in classifying vocalizations. However, the Neural Network model excelled with superior accuracy and balanced performance, particularly effective in handling complex, multi-class distinctions

The successful classification of goat vocalizations suggests that automated systems can improve animal welfare by facilitating early detection of distress or health issues. The study opens avenues for further exploration, including the incorporation of additional features and advanced machine learning techniques to enhance model performance.

## ACKNOWLEDGMENT

I would like to thank the VOCAPRA project team, including the staff at the University of Milan (Unimi) and the partnering goat farms, for providing the dataset and valuable insights that made this research possible. Their contributions were instrumental to the success of this study. For more information about the VOCAPRA team, please visit https://vocapra.lim.di.unimi.it/staff.php.

## REFERENCES

[1] Cymbaluk, N. F., & Schaefer, A. L. (2012). Animal health and welfare monitoring technologies in precision farming. Livestock Science.
[2] Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. Journal of Zoology.
[3] Špinka, M. (2012). How animals perceive and respond to distress signals. Animal Cognition.
[4] Douglas, C., Bateson, M., Walsh, C., Bédué, A., & Edwards, S. A. (2019). Environmental enrichment reduces indicators of boredom in caged farmed animals. Biology Letters.
[5] Berckmans, D. (2014). Precision livestock farming technologies for welfare management in intensive livestock systems. Rev. Sci. Tech. Off. Int. Epiz.
[6] Schön, P. C., Puppe, B., Manteuffel, G., & Tuchscherer, A. (2020). Classification of vocalizations using machine learning in livestock production. Computers and Electronics in Agriculture.
[7] Chung, Y. K., Oh, S. J., & Lee, J. K. (2013). Animal sound classification using a machine learning technique: A review. The Journal of Animal Welfare Science.
[8] Briefer, E. F., & McElligott, A. G. (2011). Indicators of age, body size and sex in goat kid calls revealed using the source-filter theory. Applied Animal Behaviour Science.