

# Comparative analysis of LS-SVM and random forest models for sensor-based lameness detection in cattle\*

Konstantinos Dolapsis<sup>1,\*†</sup>, Georgios Tziotzios<sup>1,†</sup>, Antonios Morellos<sup>1,†</sup>, Dimitrios Kateris<sup>1,†</sup> and Dionysis Bochtis<sup>1,†</sup>

<sup>1</sup> Institute for Bio-Economy and Agri-Technology (iBO), Centre for Research and Technology-Hellas (CERTH), 6<sup>th</sup> km Charilaou-Thermi Rd., 57001 Thessaloniki, Greece

## Abstract

Lameness is a significant concern in dairy cattle management, affecting both animal welfare and farm productivity. Despite efforts to mitigate its impact, traditional methods of lameness detection often overlook early signs, leading to delayed intervention and prolonged suffering for affected cows. This challenge underscores the need for more effective and proactive approaches to identifying and managing lameness. This study seeks to create an objective lameness detection methodology using sensor data from cattle limbs. Inertial Measurement Units (IMUs) were attached to cattle in Eastern Macedonia and Thrace, Greece, to record movement data. The collected data were preprocessed to address missing values, and features from both the time and frequency domains were extracted. Key gyroscope and accelerometer features were selected through Neighborhood Components Analysis. These features were then used to train Least-Squares Support Vector Machine (LS-SVM) and Multiclass Random Forest (MRF) models to classify lameness severity into healthy, mild, and severe categories, achieving an overall accuracy of more than 0.90 for both models. MRF has shown a better performance than LS-SVM.

## Keywords

Lameness detection, machine learning, inertial measurement unit, cattle, Multiclass Random Forest, Least Squares Support Vector Machine

## 1. Introduction

Lameness in dairy cattle is a major welfare concern and economic burden, leading to decreased milk production, increased treatment costs, and early culling [1]. It causes pain and distress, further reducing productivity. Early detection is crucial but traditional methods, like visual assessment, are subjective and inconsistent [2]. Automated systems are needed to improve accuracy and enable timely interventions [3].

Techniques such as human observation and pressure-sensitive walkways have historically been used to detect lameness, though these methods are labor-intensive and costly [4]. Advances in sensor-based technologies, like inertial measurement units (IMU), have enabled real-time monitoring of cattle, providing data that can be analyzed to detect lameness earlier [5].

Machine learning (ML) algorithms, particularly support vector machines (SVM) and random forests (RF) are increasingly applied to analyze IMU data for lameness detection [6]. This study compares LS-SVM and MRF models, using features selected through neighborhood components analysis (NCA) to enhance classification accuracy [7].

---

\* Short Paper Proceedings, Volume I of the 11<sup>th</sup> International Conference on Information and Communication Technologies in Agriculture, Food & Environment (HAICTA 2024), Karlovasi, Samos, Greece, 17-20 October 2024.

\* Corresponding author.

† These authors contributed equally.

✉ k.dolapsis@certh.gr (K. Dolapsis); g.tziotzios@certh.gr (G. Tziotzios); a.morellos@certh.gr (A. Morellos); d.kateris@certh.gr (D. Kateris); d.bochtis@certh.gr (D. Bochtis)

ORCID 0009-0000-8752-3493 (K. Dolapsis); 0000-0001-8290-3171 (G. Tziotzios); 0000-0002-8583-6215 (A. Morellos); 0000-0002-5731-9472 (D. Kateris); 0000-0002-7058-5986 (D. Bochtis)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## 2. Materials and Methods

### 2.1. Data Sampling

Data collection was carried out using the Blue Trident IMU device from (Figure 1). Four devices were placed on the limbs of each animal. The device features three types of sensors: an accelerometer, a gyroscope, and a magnetometer, all capturing data across three axes. In this study, only the data from the accelerometer and gyroscope were used, with a sampling rate of 500 Hz. Consequently, the dataset consists of six columns for sensor readings, one for timestamps, and another for the animal's class or state at the time of measurement, which was determined by expert visual observation and used as the ground truth. Each animal was observed freely for 10 minutes while the IMU sensors were active.



**Figure 1:** The Vicor Blue Trident IMU sensors (left) and their attachment to the cow's legs (right)

### 2.2. Feature Extraction and Selection

Feature extraction is widely used in the analysis of locomotion monitoring data. This approach involves deriving new features from the raw data to improve the interpretation of the captured information and to enable more precise analysis compared to using raw data alone. The extracted features can be categorized into several groups, such as statistical (e.g., mean, standard deviation), time domain (e.g., minimum and maximum values, signal energy), and frequency domain (e.g., dominant frequency, spectral area) [8], [9]. For this study, 13 unique features relevant to both lameness detection and general livestock activity were identified from the literature and extracted for each sensor axis (see Table 1).

**Table 1**

Time and frequency domain features that were extracted from the IMU sensor data

Feature Name	Formula
Mean	$Mean = \frac{\sum_{n=1}^N x(n)}{N}$
Standard Deviation	$STD = \sqrt{\frac{\sum_{n=1}^N (x(n) - mean)^2}{N - 1}}$
Root Mean Square	$RMS = \sqrt{\frac{\sum_{n=1}^N (x(n))^2}{N}}$
Peak Value	$Peak\ Value = \max x(n) $
Skewness	$Skewness = \frac{\sum_{n=1}^N (x(n) - Mean)^3}{(N - 1) STD^3}$
Kurtosis	$Kurtosis = \frac{\sum_{n=1}^N (x(n) - Mean)^4}{(N - 1) STD^4}$
Min	The lowest value in the window
Max	The highest value in the window
Zero Crossing	The N number of times the values change from positive to negative sign and vice versa

Neighborhood Components Analysis (NCA) was employed to identify the most relevant features for lameness detection. NCA is a feature selection method that seeks to maximize the accuracy of classification by learning a linear transformation of the input data. It selects features that improve the performance of the model by emphasizing those that contribute most to class separation. In this study, a threshold of 0.5 was applied to select the top-performing features for further analysis.

### 2.3. Machine Learning Algorithms

In this study two machine learning algorithms have been employed for the detection of lameness in the dairy cattle; the Multiclass Random Forest (MRF) and the Least-Squares Support Vector Machine (LS-SVM).

A Multiclass Random Forest model extends the traditional Random Forest algorithm to address classification tasks involving more than two classes. This model builds multiple decision trees, each trained on a random subset of the dataset. Each tree independently classifies the input into one of the available classes, and the final prediction is made based on the majority vote from all the trees involved [10]. Multiclass Random Forest models are versatile and can be used with both small and large datasets. However, they are particularly effective for large datasets due to their scalability, as they are designed to handle high-dimensional data and large amounts of information efficiently. [11].

The Least-Squares Support Vector Machine (LS-SVM) is a modified version of the traditional SVM, designed to solve classification tasks more efficiently by transforming the problem into a set of linear equations [12]. LS-SVM is particularly effective for smaller datasets and binary classification [12]. However, it may not capture nonlinear patterns as effectively unless an appropriate kernel is used.

### 2.4. Evaluation Metrics

For the evaluation of classification algorithms, the methods used were accuracy, precision and recall, as presented in Equations (1), (2) and (3).

where TP, TN, FP and FN represent the True Positive, True Negative, False Positive and False Negative samples in the confusion matrix, respectively.

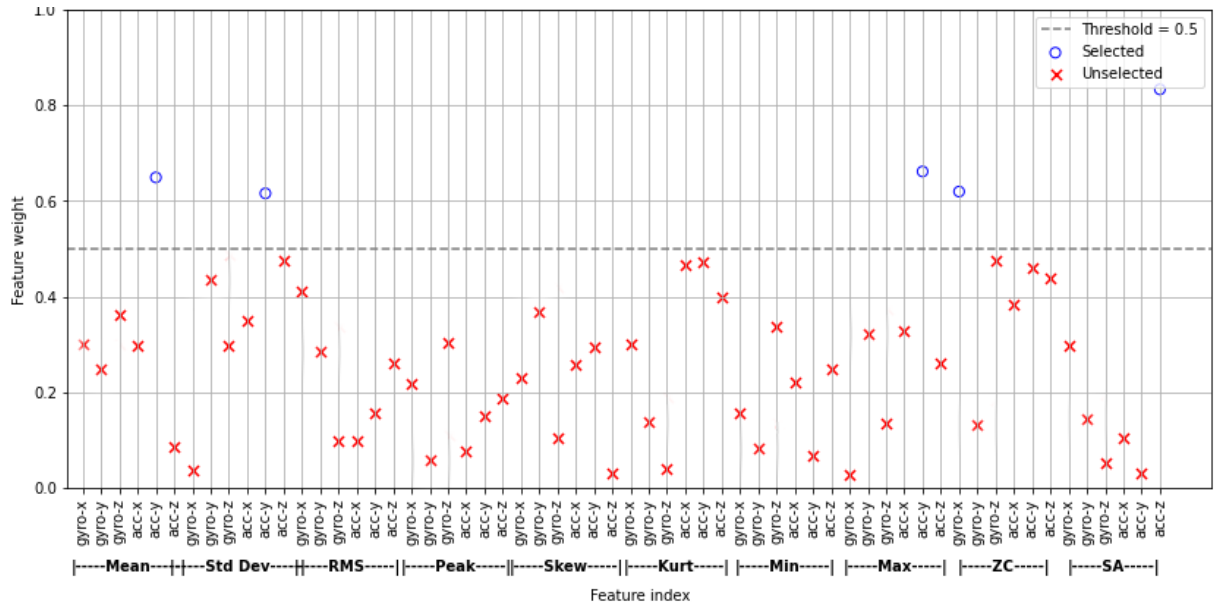
The accuracy estimation method is one of the most common, but it suffers due to its sensitivity to imbalanced data. Another issue is that two classification algorithms may have the same accuracy but different performance in terms of the correctness of the decisions they make [13]. For this reason, it is recommended not to be the sole method chosen for evaluating such models.

## 3. Results and Discussion

### 3.1. Features Selection

The features selection as it was mentioned, was based on the NCA algorithm, as is shown in Figure 2.

The selected variables, as they have been decided by NCA algorithm were the mean, the standard deviation and the maximum value across the y axis of the accelerometer, the zero crossing across the x axis of the gyroscope and the signal area, across the z axis of the accelerometer. The mean, standard deviation, and max in the y-axis of the accelerometer measure changes in vertical limb movement, providing insights into stride length and height. The zero crossing in the x-axis of the gyroscope may reflect shifts in limb orientation and balance, essential for detecting irregular gait patterns. The signal area in the z-axis of the accelerometer tracks overall limb movement energy, which is critical for identifying reduced or exaggerated movement, both of which can indicate lameness. These features together offer a comprehensive view of movement abnormalities.



**Figure 2:** Selected variables for the training of the machine learning models, as they have been decided by the neighborhood components analysis. ZC corresponds to the zero crossing and SA to the signal area features

### 3.2. Hyperparameter Tuning

Hyperparameter selection is a crucial step in the machine learning pipeline, affecting model performance, training time, and generalization. Techniques like grid search are commonly used for hyperparameter tuning. The grid search method systematically tests combinations of hyperparameters, using cross-validation to ensure results aren't dependent on data splits. The best hyperparameter set is determined based on performance metrics from a validation set. The list of the possible hyperparameters for each ML algorithm and their values range is given in Table 2.

The selected hyperparameters after the grid search method were the following: for the LS-SVM model, a C of 10, a Gamma of 0.05 and an Epsilon of 0.01 and for the MRF model a number of trees of 100, a Gini criterion, a maximum depth of 10 and a minimum samples to split a node of 2.

**Table 2**

Hyperparameter tuning ranges used in grid search optimization for the LS-SVM and MPL ML models

Model	Hyperparameter	Short Description	Possible Values
LS-SVM	C	trade-off between model complexity and training error.	0.5, 1, 10, 20, 50
	Gamma (sigma)	width of the Gaussian kernel	0.01, 0.05, 0.1, 0.5
	Epsilon	tolerance for errors	0.1, 0.05, 0.01, 0.001
MRF	Number of Trees	number of decision trees in the model	50, 100, 200, 500
	Criterion	function to measure quality of a split	Gini, Entropy
	Maximum Depth	longest path from the root to a leaf node	None, 10,20, 50
	Minimum Samples to Split a Node	the minimum number of data samples required for a node to be split into child nodes	2, 5, 10, 20

### 3.3. Model's Performance

The performance of the classification models was evaluated using a confusion matrix, which provides insights into the model's accuracy and ability to distinguish between the different lameness

status of the cattle. The data split followed a 70-30% scheme, for training and test, respectively, with the test set containing 168 values per feature for each class. The following confusion matrices of Table 3 and performance metrics summarize the results.

**Table 3**

Confusion Matrix for the LS-SVM and MRF algorithm on the detection of cattle lameness. In the table H denotes the healthy condition and ML and SL the mild and severe lameness, respectively

		LS-SVM				MRF					
		Actual			Recall	Actual				Recall	
		H	ML	SL		H	ML	SL			
Predicted	H	154	9	5	0.92	Predicted	H	162	4	2	0.96
	ML	7	146	15	0.87		ML	5	161	12	0.90
	SL	0	3	165	0.98		SL	0	1	167	0.99
Precision		0.96	0.92	0.89		Precision		0.97	0.97	0.92	
Overall Accuracy		0.92				Overall Accuracy		0.95			

The results from both confusion matrices demonstrate a strong overall ability of the models to detect varying levels of lameness in cattle, with notable differences in performance between the LS-SVM and Random Forest classifiers. Comparatively, Multiclass Random Forest outperforms LS-SVM in both precision and recall across each lameness category. For precision, Random Forest achieved slightly higher values across all categories (0.97, 0.97, and 0.92 for healthy, mild, and severe lameness, respectively) compared to LS-SVM (0.96, 0.92, and 0.89). In terms of recall, Random Forest also displays superior results (0.96, 0.90, and 0.99) in comparison to LS-SVM (0.90, 0.87, and 0.98), underscoring its enhanced effectiveness in identifying actual cases across all categories, especially in the "healthy" and "mild lameness" classes.

Overall, Random Forest demonstrates a more balanced and reliable classification performance in this scenario, effectively capturing a greater number of true instances with fewer false alarms. This improvement can likely be attributed to Random Forest's ability to capture complex, non-linear relationships within the data, a characteristic that is often limited in LS-SVM models due to their reliance on a fixed kernel function [14,15].

## 4. Conclusions

- The use of Inertial Measurement Units (IMUs) combined with machine learning models (MRF and LS-SVM) was effective in detecting lameness in cattle, achieving over 90% accuracy.
- MRF outperformed LS-SVM across all categories. This is likely due to MRF's ability to handle complex, non-linear relationships.
- Neighborhood Components Analysis (NCA) successfully identified key gyroscope and accelerometer features, enhancing model performance.
- Overall, sensor-based detection offers an objective and efficient alternative to traditional visual assessments, with the potential for early intervention and improved animal welfare.

## Acknowledgements

This research was funded by the Action «Rural Development Programme of Greece 2014-2020» under the call Measure 16 “Co-operation”, Sub-Measure 16.1 – 16.2» that is co-funded by the European Regional Development Fund and Region of Eastern Macedonia and Thrace, project

«COWLAM - Lameness identification system at milk- producing cow units in the Region of Eastern Macedonia and Thrace» (Project code: M16ΣYN2-00031).

## Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

## References

- [1] S. Lucey, G. J. Rowlands, and A. M. Russell, "The association between lameness and fertility in dairy cows," *Vet Rec*, vol. 118, no. 23, pp. 628–631, 1986.
- [2] H. R. Whay, D. C. J. Main, L. E. Green, and A. J. F. Webster, "Assessment of the welfare of dairy cattle using animal-based measurements: direct observations and investigation of farm records," *Veterinary record*, vol. 153, no. 7, pp. 197–202, 2003.
- [3] J. Wang, H. Zhang, J. Ji, K. Zhao, and G. Liu, "Development of a wireless measurement system for classifying cow behavior using accelerometer data and location data," *Appl Eng Agric*, vol. 35, no. 2, pp. 135–147, 2019.
- [4] F. C. Flower and D. M. Weary, "Effect of hoof pathologies on subjective assessments of dairy cow gait," *J Dairy Sci*, vol. 89, no. 1, pp. 139–146, 2006.
- [5] M. Alsaad and W. Büscher, "Detection of hoof lesions using digital infrared thermography in dairy cows," *J Dairy Sci*, vol. 95, no. 2, pp. 735–742, 2012.
- [6] M. Taneja, J. Byabazaire, N. Jalodia, A. Davy, C. Olariu, and P. Malone, "Machine learning based fog computing assisted data-driven approach for early lameness detection in dairy cattle," *Comput Electron Agric*, vol. 171, p. 105286, 2020.
- [7] N. Kavya *et al.*, "Feature selection using neighborhood component analysis with support vector machine for classification of breast mammograms," in *International Conference on Communication, Computing and Electronics Systems: Proceedings of ICCCES 2019*, Springer, 2020, pp. 253–260.
- [8] D. Kateris, I. Gravalos, T. Gialamas, P. Xyradakis, and D. Moshou, "A new approach to fault diagnosis in agricultural tractor mechanical gearbox," In *Proceedings of the 6<sup>th</sup> International Conference on Trends in Agricultural Engineering*, 7 - 9 September 2016, Prague, Czech Republic, pp. 290-299.
- [9] J. Kaler, J. Mitsch, J. A. Vázquez-Diosdado, N. Bollard, T. Dottorini, and K. A. Ellis, "Automated detection of lameness in sheep using machine learning approaches: Novel insights into behavioural differences among lame and non-lame sheep," *R Soc Open Sci*, vol. 7, no. 1, p. 190824, 2020.
- [10] A. Prinzie and D. Van den Poel, "Random forests for multiclass classification: Random multinomial logit," *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1721–1732, 2008.
- [11] A. Tripathi, T. Goswami, S. K. Trivedi, and R. D. Sharma, "A multi class random forest (MCRF) model for classification of small plant peptides," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, pp. 100029, 2021.
- [12] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process Lett*, vol. 9, pp. 293–300, 1999.
- [13] V. García, R. A. Mollineda, and J. S. Sánchez, "Theoretical analysis of a performance measure for imbalanced data," in *2010 20th International Conference on Pattern Recognition*, IEEE, 2010, pp. 617–620.
- [14] M. Singla, D. Ghosh, and K. K. Shukla, "A survey of robust optimization based machine learning with special reference to support vector machines," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 7, pp. 1359–1385, 2020.
- [15] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009.