Deep Learning for Lameness Level Detection in Dairy Cows

Shahid Ismail

1. Universidad del Atlántico Medio, Spain.
2. College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Rawalpindi, Pakistan.

Moises Diaz*

Instituto Universitario para el Desarrollo Tecnológico y la Innovación en Comunicaciones (iDeTIC), Universidad de Las Palmas de Gran Canaria, Spain.

Miguel Angel Ferrer

Instituto Universitario para el Desarrollo Tecnológico y la Innovación en Comunicaciones (iDeTIC), Universidad de Las Palmas de Gran Canaria, Spain.

Abstract

Lameness detection using raw sensor data is a very challenging task, as the data are devoid of specific information regarding predictors such as gait distribution, weight among legs, etc. We have addressed this challenge using a deep learning technique, named LLP-Cow (Lameness level predictor for Cow), which is an application of artificial intelligence (AI). For objective comparison, LLP-Cow is validated using CowScreeningDB, an unbalanced public dataset composed of sensor data. This dataset is recorded during the normal life of dairy cows. Hence, LLP-Cow models the normal behaviour of cows and consists of feature extraction, application-specific deep network and a voting system. The technique presented is able to model the behaviour of a cow for both binary and multiclass classification. The precision and specificity reported by our technique stand at 0.94 and 0.98 for multiclass and 0.91 and 0.90 for binary protocols for the best case scenario. Moreover, F1 measure, Matthews correlation coefficient and Kappa are 0.94, 0.91, and 0.91, respectively. The technique introduced provides a margin for human intervention through the use of a voting system at the classification stage. The technique presented is therefore an implemented AI system for cow lameness detection that offers room for exploration in terms of real time implementation.

Keywords: Artificial intelligence lameness detection, deep learning application, voting system application of artificial intelligence.

*Corresponding author
Email addresses: 1. shahid.ismail@atlanticomedio.es and 2. shahid.ismail@ceme.nust.edu.pk (Shahid Ismail), moises.diaz@ulpgc.es (Moises Diaz), miguelangel.ferrer@ulpgc.es (Miguel Angel Ferrer)

1. INTRODUCTION

Livestock is farmed for industrial production of dairy animal products such as milk, meat, yogurt, etc. [51]. Although livestock farming increases output of the products mentioned above, it does bring new challenges in terms of costs related to in-house keeping, feed input, and dairy diseases. Veterinarian clinicians generally monitor dairy health, productivity, and related costs. They use visual information from gait and posture to detect any defects in the cows under observation. This classification process is time consuming and labour intensive. The vet also requires vast experience for enhanced classification. However, with the rise of artificial intelligence, dairy-related processes are also being automated [14]. An AI-based system for dairy is generally composed of a layered structure, consisting of a sensor-based hardware layer at the bottom with multiple layers of software abstractions on top. These systems, however, generally suffer from being perceived as "black boxes" and they have robustness issues with respect to the input, undermining trust in them. A recent trend in the studies to address this has therefore been to focus on reliability, safety, and trustworthiness, while maintaining human societal norms such as human values, ethical principles, and legal requirements. Both technical and societal aspects are being covered by using the concept of human-in-the-loop systems. These new systems, also known as human-centred AI systems, are being explored in ethically sensitive applications such as agriculture. Especially important is Agriculture 5.0, which blends technological power with human societal requirements [22].

During automation, sensors of both invasive (attached to the body of a cow) and non-invasive types are used to gather data which can be used to assess overall cow health. AI-based systems, a major focus of the dairy industry, are being used to detect cow behaviour [31, 4, 9] and for behavioural change detection (BCD) [37], teat end condition [46, 47], diseases [17], heat stress [11, 55], lactation prediction [42], oestrus detection [50, 33, 48], etc. BCD can be attributed to the disease, feed intake, etc., but it is generally physiological in nature. BCD encompasses both healthy and sick cow behaviours. In one of the recent studies for BCD, Nagy et al. [37] used deep learning (DL) features with a fast region-based convolutional neural network (R-CNN) [18]. In a study on behaviour, Fuentes et al. [15] utilized deep learning to monitor individual cow behaviour. Bloch et al. [7] used transfer-based learning for feeding behaviour. Similarly, in their study on behaviour, Wu et al. [52] used both machine and deep learning for respiratory behaviour. It can be concluded from the discussion presented above that deep learning is the main approach used in AI-based dairy systems for BCD. Among BCD, lameness is the most important behavioural change from a financial perspective. Lameness is basically an abnormal animal gait/stance and it reflects a spectrum of physiological behaviours such as claw horn disruption lesions [35, 19], skin lesions [28], and non-foot lameness [35].

Lameness is usually assessed by trained vets using visual inspection and predictors such as pedometer

Table 1: Literature review. The table shows that supervision-based deep learning systems are the focus of recent studies. In the table, Bn, CNN, R-CNN, SWT, and YOLO stand for binary, convolutional neural network, region-based CNN, squeezed wavelet transform, and you only look once, respectively.

Study	Signal	Score	Features	Technique	Focus of study
Nagy et al. [37], 2023	Video recording	1-5	Deep	RCNN	Body condition prediction
Sadeghi et al. [41], 2022	Images	-	Deep	Tiny YOLO3	Endometritis diagnosis
Wang et al. [49], 2022	Thermal images	-	Deep	YOLO5	Mastitis detection
Afridi et al. [1], 2022	Images	-	Deep	CNN	Udder traits
Balsso et al. [5], 2022	Acceleration signal	-	Deep	CNN	Cow behaviour
Arazo et. al. [3], 2022	Colour and depth videos	1-5	Deep	SlowFast and application-specific	Lameness
Leach et. al. [29], 2022	Video recording	0-3	Statistical	Mask-RCNN	Lameness
Karoui et. al. [27], 2021	3D time sequence	Bn	Deep	Lenet [30]	Lameness
Jarchi et. al. [24], 2021	Accelerometer input	0-2	Deep	Hierarchical deep learning and SWT	Lameness
Gao et. al. [16], 2021	Textual input	-	Deep	Knowledge graph and transfer learning	Disease diagnosis

activity ([36, 21]), foot disorder [10], neck activity [44] etc. To reduce dependency on vets, researchers have started employing machine learning methods. Using these methods, predictors such as gait symmetry [54], back posture [43, 38], accelerometer activity [40], etc. are sampled and machine learning methods are employed to assess lameness. However, a recent trend in lameness studies is the use of deep learning. In the study reported by Jarchi et al. [24], the authors used a deep neural network (DNN) based classifier (long short-term memory (LSTM)) complemented by wavelets. In the study, both forward transform (synchrosqueezed wavelet transform (SSWT)) and reverse transform (inverse SSWT (ISSWT)) were used. The approach used was hierarchical in nature as it employed LSTM to generate instantaneous frequency which was then utilized for lameness detection. The lameness levels utilized were between 0-1 (healthy) and 2 (lame) using accelerator data. In a similar study, 3D time kinematics sequence data along with rotational matrices were used by Karoui et al. [27]. The DNN (convolutional neural network (CNN)) used during the research was LeNet [30] and the dataset was balanced between lame and healthy. However, authors did use data augmentation to generate additional signals for training/testing the deep network used (LeNet). Despite LeNet being a shallow network (60 k), Karoui et al. [27] achieved 80% accuracy for binary classification of cows. Another study involving binary classification was conducted by Arazo et al. [3]. The researchers used SlowFast [13] along with an application-specific network to predict lameness. SlowFast is fundamentally a video-based technique which uses two temporal rates and these rates are generally related. Hence, the authors utilized the red, blue, green (RGB) channels of video to segment the video inform

of binary segments which either show the absence or presence of a cow. From segmented areas, features such as spine shape and leg distances were selected to train the classifier for lameness. Multi-cattle lameness detection is another study which is based on video processing. It was conducted by Leach et al. [29] and the unique feature of the study was detection of multiple lame cows in real time. Lameness detection was initiated by analysing the back posture and head and neck movement of cows. A modified version of a mask region-based convolutional neural network (Mask R-CNN) [20] tracking algorithm was used to estimate the seven feature points in a video sequence. The detected features were tracked by applying a SORT (simple, online, and real-time tracking) algorithm [6]. Finally, a categorical boosting (CatBoost) gradient boosting algorithm [12] was applied for lameness classification of cows.

It is clear from the literature review presented that lameness is a major problem in the dairy industry and lameness detection has evolved from the classical approach of vet-based assessment to the state-of-the-art methodology of deep learning. Deep learning-based methods can be further divided into methods using images/video as an input and sensor-based methods. Azaro et al. [3] and Leach et al. [29] studied lameness while using video as an input. In contrast to these studies, Karouri et al. [27] and Jarchi et al. [24] made use of sensor-based (3D time sequence, accelerometer) input. Subsequently, the sensor-based data is converted to an image using signal-to-image-based techniques such as SSWT. The summarized view of the studies given in Table 1 shows that lameness is a focus in the dairy-related industry. However, other applications such as disease diagnosis [16, 41, 49], body condition [37], cow behaviour [5], and cow udder traits [1] can also be found. Notably, none of the studies mentioned above or highlighted in Table 1 show lameness detection on a public dataset. The importance of a study on a public dataset stems from the objectivity that it could bring for lameness detection. To the best of the authors' knowledge, not a single study has been conducted on a public dataset which is composed of raw data from sensors (accelerometer, pedometer, etc.). Generally, data from sensors are converted to images and images are then passed to a convolutional neural network [27]. Human intervention is not possible in such systems, which are generally opaque in nature. To fulfil the gaps mentioned in the research, the authors have conducted a study on public data while utilizing the deep learning system named LLP-Cow. LLP-Cow is unique as it is using CowScreeningDB [23], which is a sensor-based public dataset. Features extracted from the raw signals are fed to an application-specific network, introducing a complete system for lameness detection. Human intervention is also possible at classification stage in this system. Based on our study, the following are the main contributions.

- The study introduces LLP-Cow, a deep learning system for lameness detection on a public dataset for objective comparison.
- The proposed technique is composed of an application-specific network, a small yet unique set of features and a voting system for binary as well as multiclass classification. The feature set is capable

of yielding very high classification measures of subject-specific nature, offering physical interpretation of the results as well.

- LLP-Cow can be semi-automatic, allowing for human intervention at the voting stage, i.e. the final stage before the classification. Nevertheless, LLP-Cow works fully automatically as well.
- An in-depth ablation study is presented to demonstrate the impact of the depth of the network, subsignals, and resolution of features.

The paper is organized as follows. Firstly, in the Material and Methods section (Section 2), we provide details on the dataset used to validate the technique introduced. The details of features and classifier are given in Section 2.3, followed by the Results (Section 3). The Discussion section (Section 4) presents the details on study impact, its comparative analysis with the state-of-the-art, and future implications. The conclusions drawn are then given in the final section (Section 5).

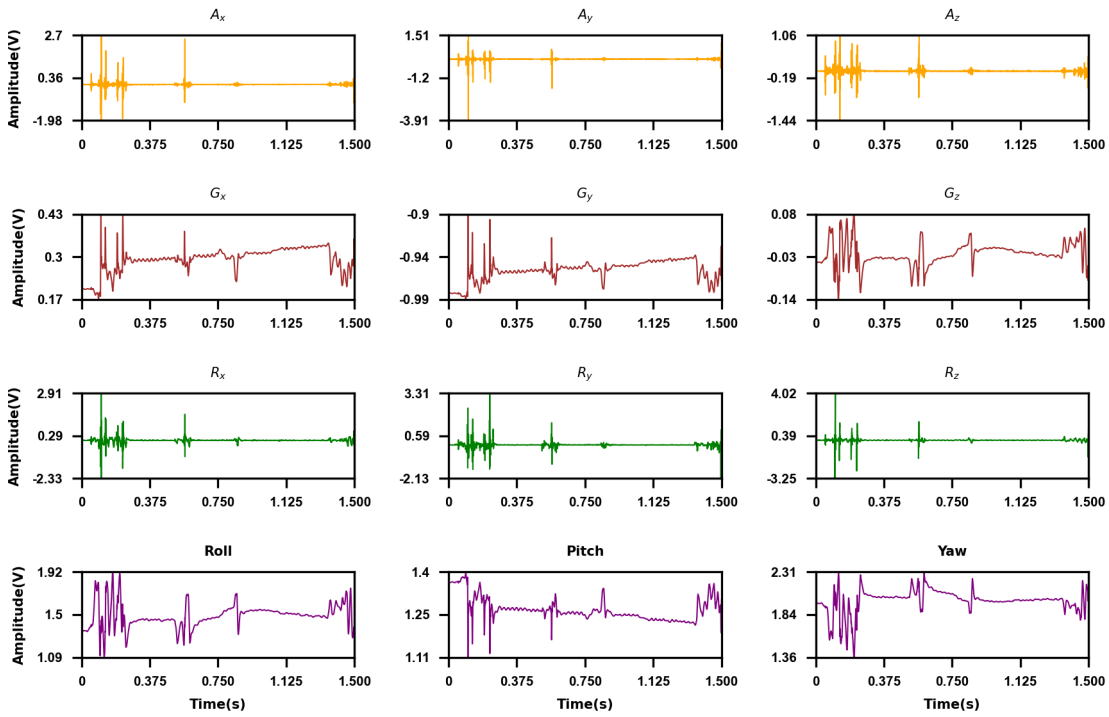


Figure 1: Details of an example signal used in LLP-Cow [23]. There are four types of subsignals namely A, G, R, and angular positions, where A, G, R, and angular positions represent acceleration, subsignals from magnetometer i.e. gravity, angular rotational and angular positions (roll, pitch and yaw), respectively. X_a represents the subsignal along an axis, with 'a' being x, y or z. Hence, R_x would represent rotation along x-axis.

2. MATERIAL and METHODS

The material and methods section is divided into material used, data generation, classification system, and classification protocols. The subsections included in the classification system are feature extraction,

feature detail, and classification system.

2.1. Material: CowScreeningDB

The material used is a public dataset introduced as CowScreeningDB in Shahid et al. [23]. This dataset is sampled from 43 cows and each signal is composed of 12 subsignals. These subsignals are sampled using accelerometer, magnetometer, and gyroscope across all three axes, namely the x, y, and z axes. Figure 1 shows an example signal which contains all the subsignals. The dataset mentioned is from healthy as well as lame cows. Lame cows are further divided into four categories, with a lameness score of 1 reflecting a moderate level and a lameness score of 4 indicating severe lameness. Hence, lameness levels for cows range from 1-5 with level 1 reflecting a completely healthy cow.

2.2. Data generation

As mentioned in the material section (Section 2.1), the data is sourced from 43 cows. However, there are 11,518 files of 90 s (1.5 m) each. These files can be appended in series or in parallel, during which all the files can be considered independent of each other, resulting in two protocols, as shown in Figure 2 (a).

2.2.1. Serial data generation

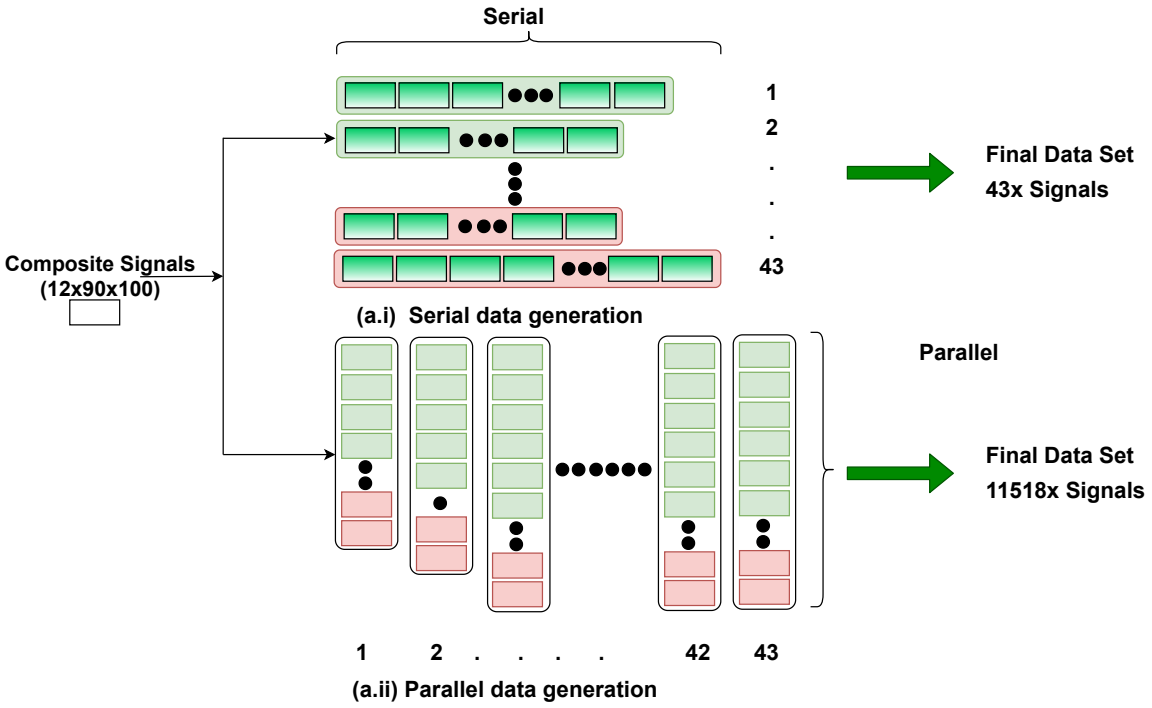
During serial data generation (SDG), all the files from a cow are considered as a composite signal divided into multiple files. Each file is composed of 12 subsignals as shown in Figure 1. Therefore, similar subsignals from each file are appended in a series, resulting in a total of 43 signals for the entire dataset. This data generation is useful in a global representation of lameness.

2.2.2. Parallel data generation

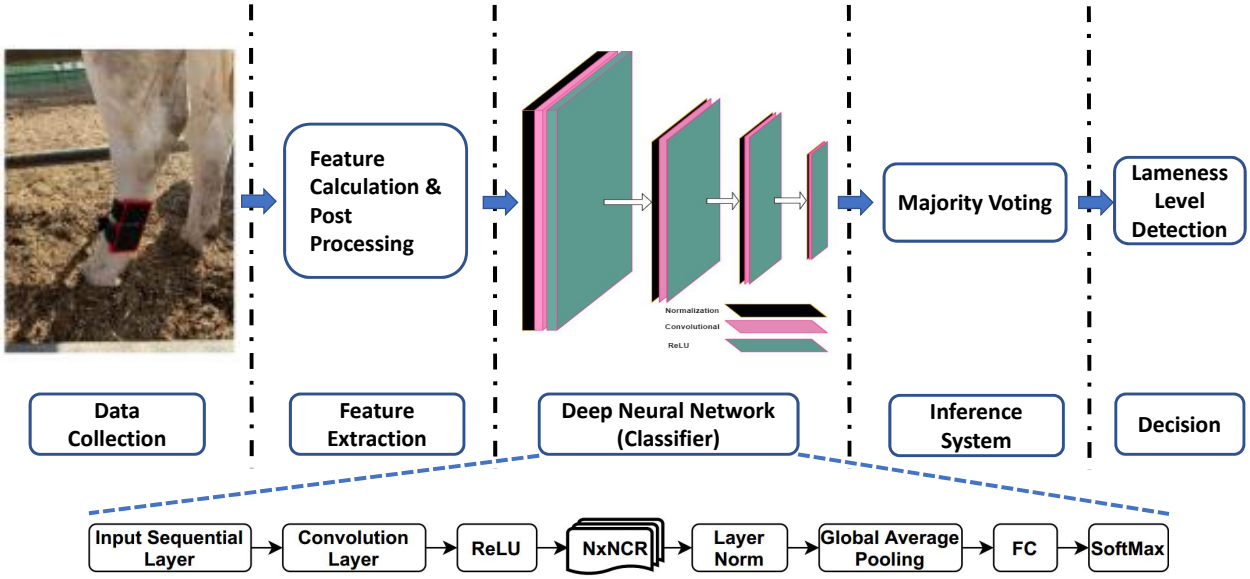
During parallel data generation (PDG), each file is considered independent of the other files, resulting in 11,518 files. Therefore, it is a requirement to use a voting system for final classification as multiple labels are generated during inference. During a PDG-based study, files from each cow are present in training as well as in testing. This data division is therefore stratified in nature and reflects the subject-specific nature. The focus of this type of data generation is to detect the lameness level of a cow from a single file (90 s of data) using local information.

2.3. Classification system

A block diagram for the classification system is shown in Figure 2 (b). It starts with data collection, which covers both data sampling and data generation. Both data-related aspects are already presented in the methods (Section 2.1) and data generation (Section 2.2) sections, respectively. The remaining classification system is divided into feature extraction, feature details, the classifier utilized and the voting-based inference. It should be noted that the classifier, along with the voting system, constitute the expert system presented in this study. The details on feature extraction are presented below.



(a) Appending of raw data. Data can be appended in serial or in parallel, as shown in Figures (a.i) and (a.ii), respectively. Depending on the methodology used, different numbers of signals result.



(b) Classification system: Major modules are feature extraction, the deep classifier, and a voting system, where, NxNCR reflects N layers of normalization, convolutional, and ReLU layers.

Figure 2: Figure 2 (a) shows the appending of the data for different classification protocols. The generated data is then used in the classification system given in Figure 2 (b).

2.3.1. Features extraction

As mentioned in the dataset section (Section 2.1), the data utilized during this study is multi-sensor in nature as it is composed of four different types of sensor data. This sensor data refers to different modalities, which are acceleration, gravity, angular position, and angular velocity. Features are extracted from each of these sensor's data and the extracted features are combined together as a matrix. This matrix is given as an input to the classifier, with dimensions of 21×12×900, as presented below in Table 2. Here, 21, 12, and 900 are the number of features, N_{tf} , the number of sensors, N_s and the feature length, respectively. Moreover, N_f reflects the feature number for a certain type of feature. For example, most numerous (11) features are extracted from data, i.e. window and Hjorth-based features. Features extracted during the feature extraction are data features, histogram representation of files, instantaneous frequency, correlation, spectrum, singular spectrum analysis and Hjorth parameters. Many features have been proposed over the years for classification problems related to dairy (livestock identification [2]) and non-dairy systems (heat transfer [39], defect detection [32], etc.). However, we have selected the above-mentioned features according to the computational cost related to their calculation, the ability of a feature to offer a physical interpretation, and the nature of the properties they capture. Details of the methodology for calculating the features, their mathematical description, and the physical interpretation of the features are given in following section (Section 2.3.2).

Table 2: Features utilized in the study, where, Hist. $Freq_{Inst}$, Corr. and $Featue_{Dim}$ represent histogram, instantaneous frequency, correlation based features, and feature dimensionality, respectively.

	Data		*Hist.*	$Freq_{Inst}$	*Corr*	*Spectral*		*SSA*	N_{tf}	N_s	$Feature_{Dim}$
	Window	Hjorth				Raw	Mean				
N_f	8	3	1	1	1	1	1	5	21	12	21×12×900

2.3.2. Features utilized during study

The features are described below:

- Data features: Two types of features, namely window and Hjorth features, are combined together under data features. Window-based features are presented first.
  - **Windowed-based features:** Various features, namely sub-sampling, fractional Brownian motion, and statistical features, are categorized as window-based features. For subsampling, files of length L are decimated using decimation factor D_F (10). In order to calculate the fractional Brownian motion-based (fBm) features, the signal is assumed to be a continuous time Gaussian process and two Hurst parameters based on discrete second derivative and linear regression of variance are used during the study. The statistical features considered are inverse coefficient of

variation (ratio of mean to standard deviation), moving mean, moving median, moving standard deviation, and the signal range.

- **Hjorth features:** Activity, mobility, and complexity are the parameters which are collectively considered Hjorth parameters and these are based on variance of signal (activity) and its derivative. They are interrelated as shown in Equation 1 given below

$$Activity(y(t)) = var(y(t)) \tag{1a}$$

$$Mobility(y(t)) = \sqrt{\frac{var(y'(t))}{activity}} \tag{1b}$$

$$Complexity(y(t)) = \frac{Mobility(y'(t))}{Mobility(y(t))} \tag{1c}$$

Where, $y'(t)$ and $var(y(t))$ are the derivative with the respect to time and the variance of the signal, $y(t)$, respectively.

- Histogram features are based on a normalized histogram. Figure 3 shows the histogram-based feature distribution for healthy as well as lame cows. Figure 3 shows that histogram-based features are capable of distinguishing between healthy and lame cows.

- Instantaneous frequency: For a signal, $s(t)$ and its analytical representation $z(t)$, the instantaneous frequency is given by Equation 2 [8].

$$s(t) = a(t)\cos\phi(t)$$

$$z(t) = s(t) + jH|s(t)| = a(t)e^{j\phi(t)} \tag{2a}$$

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}[argz(t)] \tag{2b}$$

In the above Equation 2, $f(t)$ represents the instantaneous frequency at time t and this is given by the derivative of angle of $z(t)$, the analytical representaion of $s(t)$.

- Correlation-based Eigen value features: For correlation features, first, an unbiased correlation signal is generated and this is then converted to a Toeplitz matrix. Then, Pearson's correlation is performed on the Toeplitz matrix and Eigen values are calculated from the result, which are then converted to the real values. The final values are sorted in descending order representing the features. The mathematical manipulations involved in the calculation are outlined in Appendix A.1.

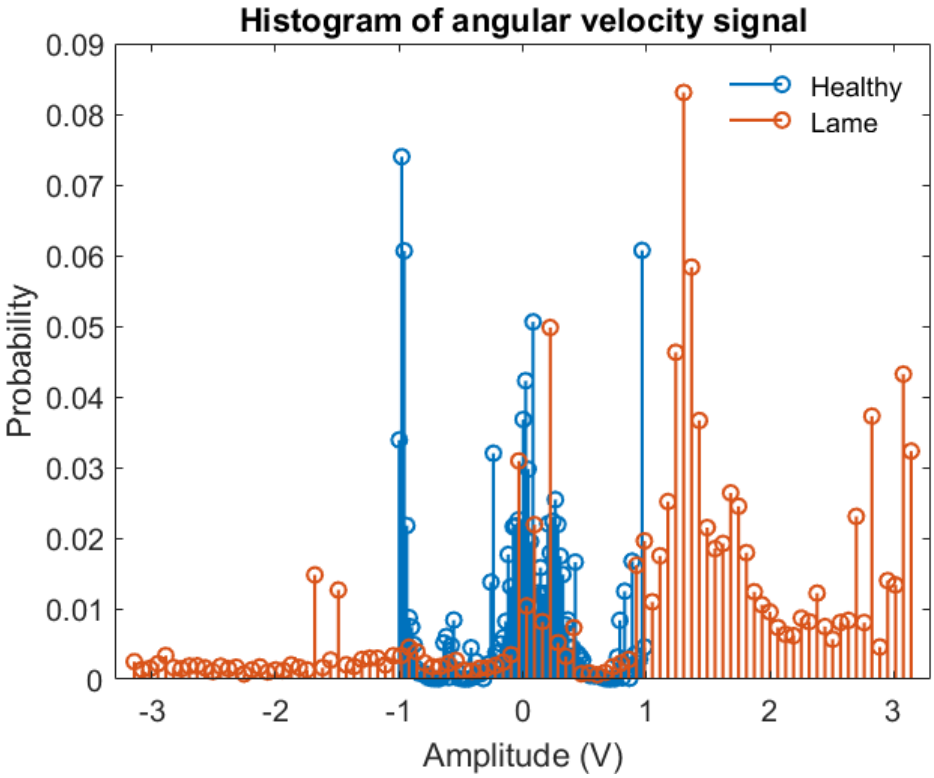


Figure 3: Histogram-based features. The figure shows that histograms have different morphologies, both in terms of distribution and contents. Hence they can be used to distinguish between a healthy cow and a lame one.

- Spectral Chirplet Z Transform-based features: Features based on Chirplet Z Transform (CZT) are calculated [34] using Equation 3. CZT is similar to Fourier transform, but gives more detailed information regarding the spectra of a signal.

$$X(Z_k) = \sum_{n=0}^{N-1} x[n] z_k^{-n} \qquad n, k = 0, 1, ...N-1 \tag{3a}$$

$$z^{-n} = re^{j\omega} = 1.e^{\frac{j\pi k}{N}} \tag{3b}$$

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{\frac{-j\pi nk}{N}} \tag{3c}$$

In Equations 3a and 3c, $x[n]$, $X(k)$ and $X(Z_k)$ represent signal, Fourier Transform, and Chirplet Transform, respectively. From the Equations 3a and 3c, it can be inferred that CZT and FFT are calculated over unite circle ($e^{\frac{j\pi k}{N}}$) and arbitrary curve(z_k^{-n}), respectively. The calculation over curve gives more control of resolution during calculation of CZT.

- Singular spectrum analysis-based features: These features are calculated using singular spectrum analysis (SSA), which is based on embedding, singular value decomposition (SVD), grouping, and diagonal averaging [45]. In the present study, features are based on signals from an SVD stage. Hence, noise components are retained for further processing without using the final stage of reconstruction, i.e. grouping and diagonal averaging.

Table 3 presents details about features in terms of sub-features, their dimension, feature space, physical interpretation of the feature, and calculation methodology. Table 3 shows that features are calculated from multiple domains including time, frequency, time-frequency, statistical, and Hilbert features. Calculation of the features in various domains enables the system to analyse the signal from different perspectives due to the enriched information these perspectives provide. Once features have been calculated, post-processing is applied, which is composed of removal of $\pm\infty$ and NaN values, converting the features to z-scores and decimation by factor of 10, respectively.

2.3.3. Classification system

The classification system used during the study is a deep neural network composed of multiple layers (Figure 2 (b)). The major layers are input sequential layer, multiple layers of NCR (normalization, convolution, and rectified linear unit (ReLU)) followed by global average pooling and fully connected layer. Finally, softmax is utilized for classification.

Table 3: The table presents details about features, their physical interpretation and methodology used for their calculation. Also presented are feature dimensions and the spaces these features represent, where, N/A refers to not available in the context of features having no sub-features.

Name	**Subfeature**	**Dimension**	**Space**	**Physical Interpretation**	**Calculation methodology**
Windowed	Subsampling	1	Time	Decimated version of original data	Non overlapped local windowed data
	Statistics	7	Time	Measure of first order statistics	
	Hurst	3	Time	Measure of long term memory	
Histogram	N/A	1	Statistical	Measure of central tendency	Global
Instantaneous frequency	N/A	1	Hilbert	Measure of non-stationarity	Global
Correlation	N/A	1	Time	Vector measurement of similarity	Global
Chirplet Z	N/A	2	Frequency	Frequency component in range ([1-50] Hz)	Global
SSA	N/A	5	Time-frequency	Frequency variation of the signal w.r.t time	Global

Input layer. The input is a sequential layer with dimensioning compatible with the feature dimension as given in Equation 4.

$$Input_{Size} = N_s \cdot N_f \tag{4}$$

In the above Equation 4, N_s and N_f are the number of sensors and number of the feature, respectively. For the present study, $Input_{Size}$ is 252 (12x21), which is calculated by multiplying N_s and N_f together.

Normalization, convolution, and ReLU layers. The goal of normalization is to transform the features so that they are on a similar scale. This improves model performance and training stability. The convolution layer is one of the major layers used in the classification problem. Stacks of filters of various sizes (5, 7, and 9) are used during convolutional layers. Small filters are useful for extracting detail, while larger filters are useful for detecting the size of an object. The number of filters at various convolutional layers varies between 1 and 15, with 15 filters used at the top convolutional layer. Finally, an ReLU layer is used. This is an activation layer and it sets all negative values to zero while transferring all positive values to the output unchanged.

Global average pooling, fully connected, and softmax layers. Global average pooling and fully connected perform the averaging and linear operations, respectively. Softmax is a classification layer which converts

the input to the probabilities using Equation 5 given below

$$f(x_i) = \frac{e_i^x}{\sum_{j=1}^{K} e_j^x} \tag{5}$$

Equation 5 is generally used to convert the output of a neural network in the form of a real number to probabilities for classification. By using the exponential transformation (e_i^x, e_j^x), negative as well as positive numbers are converted to probability values between 0 and 1 and $f(x_i)$ with maximum value reflects the classified class. The total learnable parameters for the network are approximately 1.8 M, with the majority of contributions coming from convolution layers which use numbers of filters of different sizes as weights and biases. Another contributing factor for bias is from the normalization layer. Details of the deep neural network utilized during our study are given in Table A.

2.4. Classification protocols

Based on the data generation presented in Section 2.2, two classification protocols are designed. Details of the protocols are presented below, starting with the serial protocol.

2.4.1. Classification based on serial data

Classification based on serial data (CSD) is a machine learning-based classification that uses SDG (Figure 2 (a)). This protocol is further divided into two sub protocols namely, CSD-Binary (CSD-Bn) and CSD-Multiclass (CSD-MC). The results of CSD-Bn have already been reported in Shahid et al. [23]. CSD-MC is similar to CSD-Bn, but all lameness levels are classified during classification.

2.4.2. Classification based on parallel data

Classification based on parallel data (CPD) is the protocol used for deep learning and it uses PDG (Figure 2 (b)). This protocol reflects binary (CPD-Bn) and multiclass(CPD-MC) lameness detection and is based on stratified sampling.

3. RESULTS

Presented below are the results reported for all types of protocols, as presented in Section 2.4. We have utilized the classical metrics of precision, sensitivity, specificity, and accuracy for evaluation. Moreover, the results are also reported using advanced measures such as F1 measure, Matthews correlation coefficient (MCC), Cohen's Kappa, area under the curve (AUC), and confusion matrices. The definition of the metrics is given in Table 4. The results are reported in tabular form in Table 5 using 10-fold cross validation with data split of 70/30% for training and test portions, respectively. MATLAB 2022a was used to generate the afore-mentioned results using Dell Inspiron 15, 7700 (Ci7, 7^{th} generation) computer which was equipped

Table 4: Definition of evaluation metrics used. All metrics are defined in terms of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), respectively.

Evaluation Metric	Definition
Precision	$\frac{TP}{(TP+FP)}$
Sensitivity	$\frac{TP}{(TP+FN)}$
Specificity	$\frac{TN}{(TN+FP)}$
Overall Accuracy (OA)	$\frac{TP+TN}{TP+TN+FP+FN}$
F1 Measure	$\frac{2 \cdot TP}{2 \cdot TP+FP+FN}$
Expected Accuracy (EA)	$\frac{(TP+FP)(TP+FN)+(FN+TN)(FP+TN)}{(TP+TN+FP+FN)^2}$
MCC	$\frac{(TP \cdot TN)-(FP \cdot FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
Kappa	$\frac{OA-EA}{1-EA}$

with 32 GB system RAM and 4 GB GPU RAM. The number of epoch, learning rate (β), and mini-batch for the protocols were maintained at 100, 0.0005 and 50, respectively.

It is clear from Table 5 that a major focus of the study is on the use of deep learning, which outperforms the machine learning approach. There are two types of results mentioned in the table using top-1 and top-3 classification methodologies. Top-1 uses the classical approach, where the label with highest probability is considered final. During top-3, if the correct label is present in the top three labels, classification is considered correct. In our study, if a signal from a healthy cow is detected as healthy or lameness 1 then the signal is classified as healthy. Similarly, if cow at health level 2 is detected as a cow at health levels 1, 2, or 3, it is still considered as cow at health level 2. This top-3 approach helps to avoid serious errors during classification and it also enhances the overall classification rate reported by the methodology presented. In Table 5, X and X3 refer to top-1 and top-3 approaches respectively. For example, CPD and CPD3 are classifications based on parallel data using top-1 and top-3 methodologies. From Table 5, it is evident that top-3-based results surpass top-1 by a wide margin. For deep learning-based multiclass, the best results are reported for CPD3, which is a classification based on parallel data while using top-3 methodology. All classical/advanced measures are very high, which is due to stratified sampling.

For machine learning, the best results stem from CSD-Bn, as reported in Shahid et al. [23]. However, when

Table 5: Results on CowScreeningDB ([23]). The difference between the measures reported during X and X3 is the use of top-3 methodology for classification during X3, in contrast to X where the label reported by the classifier is final. Here, X is the protocol, for example CPD, which stands for classification based on parallel data, where, DL and ML refer to deep learning and machine learning, respectively.

Domain	Protocol	Precision	Sensitivity	Specificity	Accuracy	F Measure	MCC	Kappa
DL	CPD	0.90	0.86	0.97	0.89	0.88	0.85	0.85
DL	CPD3	0.94	0.93	0.98	0.93	0.94	0.91	0.91
DL	CPD-Bn	0.91	0.90	0.90	0.91	0.91	0.81	0.81
ML [23]	CSD-Bn	0.77	0.77	0.79	0.77	0.76	0.74	0.73
ML	CSD-MC	0.27	0.23	0.81	0.29	0.24	0.05	0.05
ML	CSD-MC3	0.50	0.50	0.88	0.55	0.49	0.40	0.39

the same features were fed to the classifier for multiclass, the classification measures become low. Use of

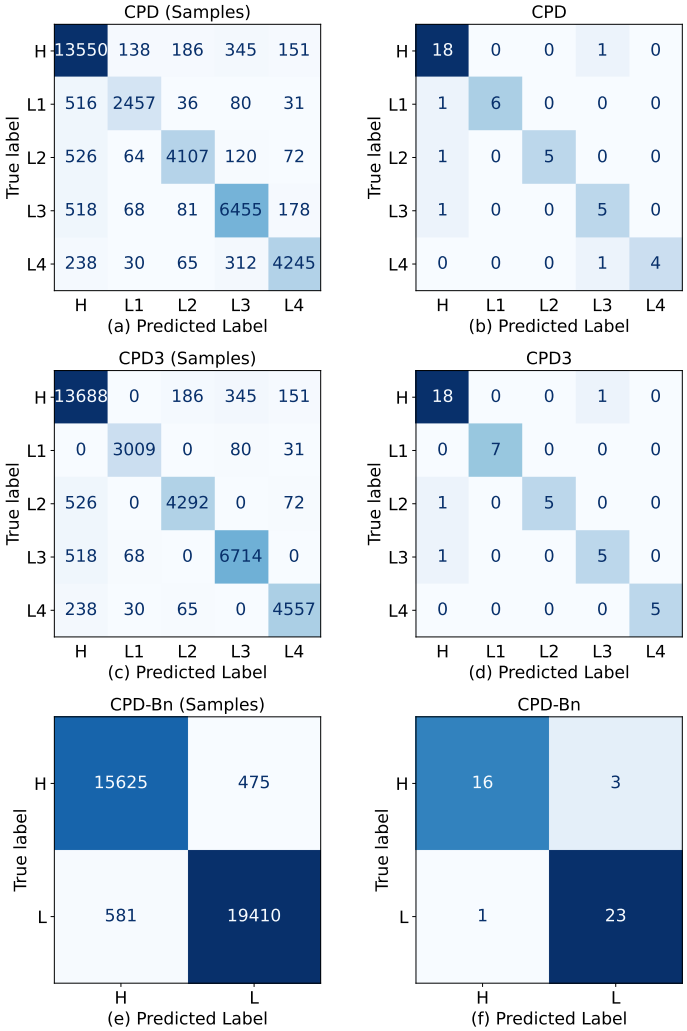


Figure 4: Confusion matrices for all deep learning protocols. Results in the form of confusion matrices are presented first for samples, followed by the signals for CPD-MC (multiclass) and CPD-Bn (binary) classifications.

top-3 brings a considerable enhancement to all the measures whether classical or advance. The comparison of machine and deep learning in Table 5 also demonstrates that the use of deep learning enhances system efficacy in the lameness detection from sensor data. Moreover, we observe that the precision and specificity for binary classification (CPD-Bn) are slightly lower when compared to multiclass classifications (CPD, CPD3). In multiclass, the variance present in all individual categories was utilized. But for binary class, variance for the lame class was increased due to merging of lameness levels 2-5. This imbalance in variance resulted in the slightly lower results in case of CPD-Bn in comparison to CPD-MC (multiclass) protocols, which may be associated with the imbalance in the dataset (files from healthy cows (4,787) and files from lame cows (6,731) [23]). This imbalance highlighted in the dataset may also be contributing towards the generalization ability of the technique proposed.

In the figures, Figure 4 and Figure 5, results are presented in form of confusion matrices and AUCs for deep learning and machine learning, respectively. In Figure 4, first the results of top-1 are presented for samples followed by the results for signals. Hence, Figures 4 (a,b), Figures 4 (c,d), and Figures 4 (e,f) represent classifications for CPD, CPD3 and CPD-Bn, respectively. In Figures 4 (a,c,e), actual results for 10-fold are presented for samples which are then averaged together in Figures 4 (b,d,f), respectively. Figure 5 uses the data given in Figure 4 for AUC. Here, 10-fold data is used to smooth the ROC curves which improves the confidence level in the reported AUCs. Moreover, ROC before serious error removal and ROC after serious error removal refer to top-1 and top-3 classification approaches, respectively. The figures show that top-3 enhances the AUC by a significant margin, not only for CPD, but also for CSD.

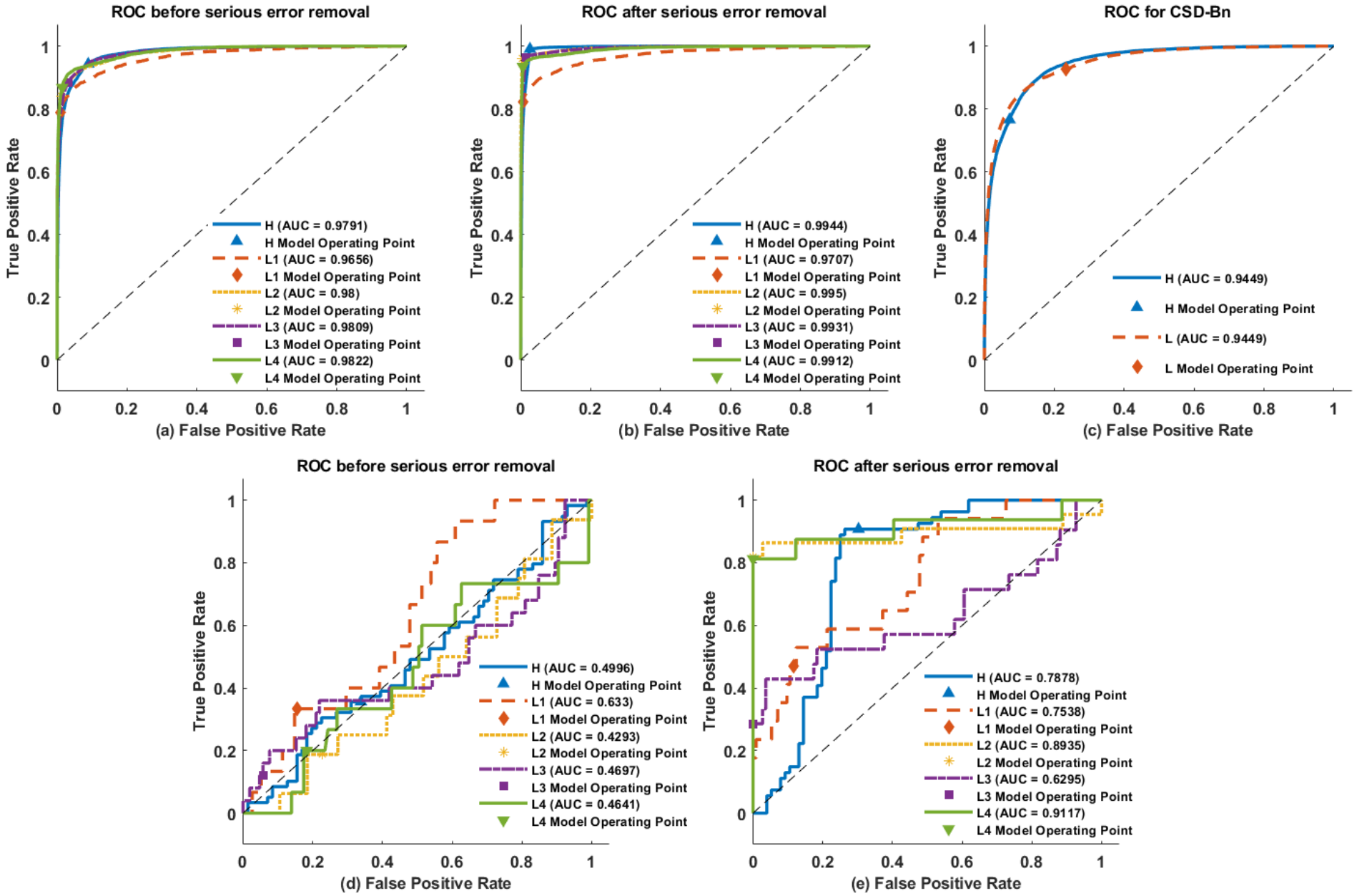


Figure 5: : Results of the classification for all protocols. ROC curves in form of false positive rate versus true positive rate are plotted for CPD (a,b), CPD-Bn (c), and CSD (d,e), respectively. Figures (b,e) present ROC after serious error removal, which refers to top-3 methodology for CPD3 and CSD3, respectively.

Results of ablation for all subsignals are presented in Figure 6 (a). It is clear from the figure that the best results on average are reported by the R_z component i.e. the rotation component along the z-axis. The results reported by the remaining subsignals are the same on average. The worst results are reported by A_z and yaw components. For features (Figure 6 (b)), window-based features report the best classification. However, the classification accuracy of Hjorth features is presented separately and it is similar to other features except instantaneous frequency-based feature. Although, histogram, instantaneous and correlation features have same dimensionality of 1, the correlation feature reports better classification. The quantitative evaluation of subsignals and features is presented in Figure 7. Figure 7 (a,d) presents pie charts for all subsignals and grouped/individual features, respectively. Figure 7 (a)) shows that the average contribution by all subsignals, namely, acceleration, magnetometer subsignals, angular position, and angular rotation is the same. Similar trends are observed for the individual features, where feature accuracy varies between 12.5% and 14.57% for all features except windowed features, which are most numerous (Figure 7 (e,f)). Moreover, this discrepancy is further increased when looking at the grouped performance, where contribution by data features is increased to 30%. All the ablation studies mentioned are based on the CPD-Bn protocol using 10-fold cross validation for averaging purposes.

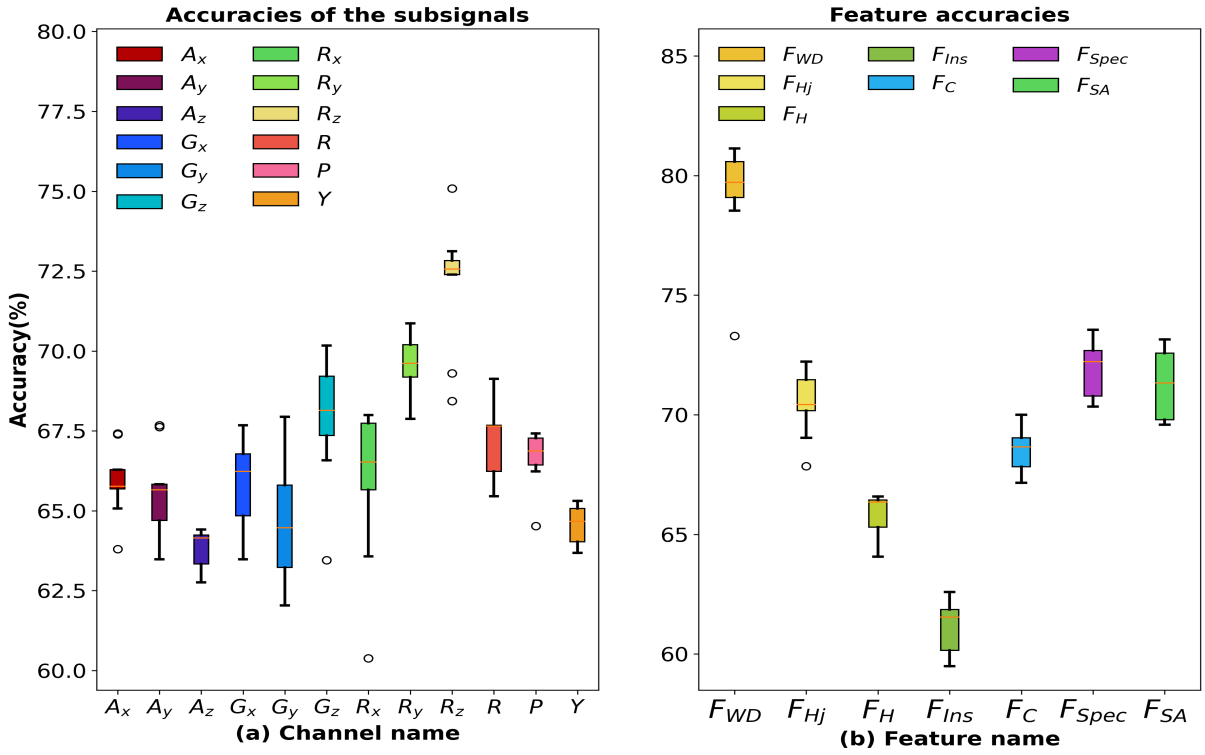


Figure 6: Boxplots of accuracies for all subsignals and features. (a) A, G, and R represent acceleration, magnetometer, and rotation for x, y, and z axis (angular velocity), respectively. R, P, and Y are roll, pitch, and yaw (angular position), respectively. (b) Accuracy of feature sets, namely, window (F_{WD}), Hjorth (F_{HJ}), histogram (F_H), instantaneous frequency (F_{Ins}), correlation (F_C), spectral (F_{Spec}), and feature based on SSA analysis (F_{SA}) are presented.

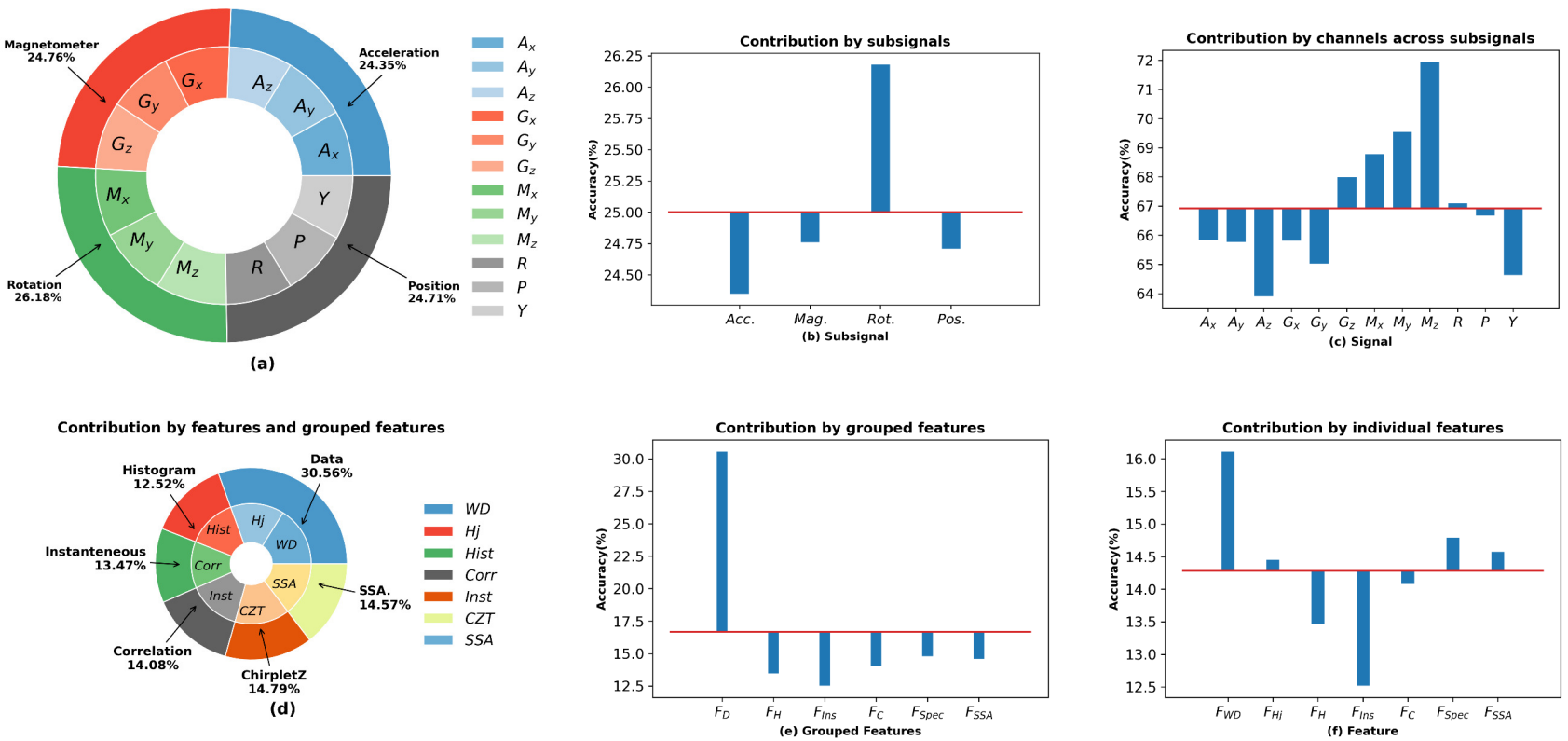


Figure 7: Pie charts of subsignals, channels across subsignals, grouped features, individual features, and their individual contributions. (a,d) Pie charts for subsignals and grouped features/features, respectively. (b-c) Accuracy reported by different subsignals and channels in the subsignals. (e-f) Figures showing the contributions by different features, whether grouped or individual. The red line here reflects the average accuracy in Figures (b,c,e,f).

4. DISCUSSION

This discussion section is divided into four major subsections, namely: objective analysis of inertial data for lameness detection, comparison with the state-of-the-art, complexity analysis, and potential for HCAI-based systems and real time implementation.

4.1. Objective analysis of inertial data for lameness detection

For objective analysis of the inertial data, the efficiency of the sampling system is measured using four parameters, namely λ_1, λ_2, $\mathcal{P}$, and $\mathcal{M}$, as given in Equations 6 and 7. In the Equation 6, l_i, δ_j, N_l, $\mathcal{L}$, and $\mathcal{L}_\Delta$ represent individual file length, time difference between files, number of files for a specific cow, resultant length, and resultant length based on time difference between files. Here, l_i, δ_j, $\mathcal{L}$, and $\mathcal{L}_\Delta$ are measured in seconds and N_l is a constant. λ_1, λ_2, and $\mathcal{M}$ are based on $\mathcal{L}$, $\mathcal{L}_\Delta$, and N_l, as given in Equation 7. However, for $\mathcal{P}$, the samples which fulfill the condition ($\delta_j > \delta_{thr}$) are considered in N_{pl} and δ_{thr} was set to 95, empirically. From the definitions, it can be inferred that λ_1 and λ_2 are simply ratios of lengths related to the files of a specific cow. $\mathcal{P}$ is a ratio based on individual δ_j fulfilling the condition mentioned above. Similarly, $\mathcal{M}$ is the average of δ_j for a cow under observation. Table 6 details the calculation of parameters mentioned. However, all measures are converted to percentage by multiplication by 100. Although the calculations shown are for a selected set of cows, similar calculation can be performed for other cows as well. From Table 6, it can been seen that when λ_1 and λ_2 are close to each other, then $\mathcal{P}$ is equal to zero and $\mathcal{M}$ is 90 ± 2. For other cases, however, a greater difference between λ_1 and λ_2 leads to higher values of $\mathcal{M}$. A greater difference between lambdas (λ_1 and λ_2) points to gaps in the sampling, which is the reason for higher values in $\mathcal{M}$.

$$\mathcal{L} = \sum_{i=0}^{i=N_l-1} l_i; \qquad \mathcal{L}_\Delta = \sum_{j=0}^{j=N_l-2} \delta_j \tag{6}$$

$$\lambda_1 = \frac{\mathcal{L}}{\mathcal{L}+\mathcal{L}_\Delta}; \qquad \lambda_2 = \frac{\mathcal{L}_\Delta}{\mathcal{L}+\mathcal{L}_\Delta}; \qquad \mathcal{P} = \frac{N_{pl}}{N_l - 1}; \qquad \mathcal{M} = \frac{\mathcal{L}_\Delta}{N_l - 1} \tag{7}$$

4.2. Comparison with the state-of-the-art

Table 7 compares the proposed technique with the state-of-the-art using dataset, feature statistics, the technique introduced, depth of analysis, and evaluation metrics used to quantify the reported results. All the studies are generally video-based studies, except Jarchi et al. [24] which is a predictor-based study for lameness detection. Video-based studies come at a huge cost in terms of video sampling, storage, and post-processing. Moreover, additional computational cost is associated with the classifier used. All the classifiers mentioned have greater depth than the system proposed, except for the classifiers used in Karoui et al. [27]. However, they have classified the cow as lame or healthy which represents the binary classification. Better

Table 6: Efficiency of the sampling system is presented for a selected number of cows. Here, λ_1, λ_2, $\mathcal{P}$, and $\mathcal{M}$ are used to calculate efficiency using Equations 6 and 7.

Name	**Cow Type**	$\mathcal{L}$	$\mathcal{L}_\Delta$	$\mathcal{L}+\mathcal{L}_\Delta$	λ_1	λ_2	$\mathcal{P}$	$\mathcal{M}$
749	1	24390	24314	48704	50.08	49.92	0.00	90.05
5160	1	5940	19091	25031	23.73	76.27	67.69	293.71
808	2	24390	24250	48640	50.14	49.86	0.00	89.81
1032	2	1170	17637	18807	6.22	93.78	91.67	1469.75
904	3	46080	46016	92096	50.03	49.97	0.00	90.05
1184	4	46620	46436	93056	50.10	49.90	0.00	89.82
1212	4	49860	110462	160322	31.10	68.90	0.18	199.39
1299	5	20970	61521	82491	25.42	74.58	66.38	265.18
5181	5	25830	25753	51583	50.07	49.93	0.00	90.05

Table 7: Comparison of LLP-Cow with the state-of-the-art. From the table, it is apparent that the present study is the only one to provide information at both binary (Bn) and multiclass (MC) levels and is application-specific (AS). N_C, $N_{sig.}$, T_{ss}, LS, N_F, Dim. and Par. represent the total number of cows, number of signals, total number of subsignals, lameness score, number of features, dimension, and parameters, respectively. Depth is measured in megabytes (MB) and parameters are given in million (M) except for the study by Karoui et al. [27], for which they are given in thousands (k).

	Data				**Features**		**Technique**				**Analysis**	
Study	$\mathbf{N}_C$	$\mathbf{N}_{sig.}$	$\mathbf{T}_{ss}$	**LS**	N_F	**Type**	**Dim.**	**Depth**	**Par.**	**Size**	**Level**	**Evaluation**
Kang et. al. [26], 2020	100	100	3	1-3	1	1920×1080	VGG16	16	138	515	MC	0.96
Wu et. al. [53], 2020	50	750	3	0-1	-	704×576	DarkNet-53	53	41.6	155	Bn	0.96
Jiang et. al. [25], 2020	1080	1080	3	1-4	-	1920×1080	DenseNet-201	201	20	77	MC	0.98
Jarchi et. al. [24], 2021	23	23	12	0-1	16	-	AS.	12	2.45	-	Bn	1.00
Karoui et. al. [27], 2021	-	24k	-	0-1	8	32×192	LeNet	8	60k	-	Bn	0.91
Arazo et. al [3], 2022	116	5164	3	1-5	2304	340×256	ResNet-50	50	25.6	98	Bn	0.77
Leach et. al. [29], 2022	250	25	3	0-3	2	2704×1520	ResNet-101	101	44.6	167	MC	1.00
Proposed	**43**	**11518**	**12**	**1-5**	**21**	**252×900**	**AS.**	**29**	**1.8**	**1.83**	**Bn, MC**	**0.91, 0.93**

classification is reported by Leach et al. [29] and Jarchi et al. [24] . However, the system put forth by Jarchi et al. [24] is a binary classifier, as mentioned above. The authors in the study ([29]) report 100% classification, but their system has a very high computational cost and is essentially an offline system. This is in contrast to the system proposed here, which has great potential for being an online system due to its sensor base (raw data) and the low computational cost associated with it.

The level of analysis is generally binary in nature, except for Leach et al. [29], Kang et al. [26], and Jiang et al. [25]. The protocols used during the assessment of the proposed technique are unique, as the results we report are not only binary but also multiclass. The results presented in Table 7 show that the results reported are competitive with the state-of-the-art for a very challenging dataset which is sampled during daily life and only uses the raw data. Moreover, the previous section (Section 4.1) also showed that additional complexity is introduced due to the sampling system and the unbalanced nature of the dataset. The results are reported using multiple metrics of AUC, Kappa, F1 measure, and classical measures.

4.3. Complexity analysis

For complexity analysis, we have analysed our technique using feature resolution and depth of the deep network utilized. Figure 8 presents the impact of feature resolution and network depth. For reference, a 29-layer network for data features was employed. This reported accuracy of 78.66% for binary classification, i.e. CPD-Bn, in 2.18 hours. The resolution of the features was kept at 900 time stamps for this reference. Figure 8(a) shows the impact of depth by removing the top-most layers, composed of convolution, ReLU, and normalization operations. When these layers are removed, accuracy decreases to 70.13%, but the time for reporting the results is also reduced to approximately 1.4 hours. Similarly, for a depth of 23 layers, accuracy decreases to 66.45% in 1.1 hours. Hence, accuracy is directly proportional to the depth of classifier and it is reduced to 56.15% for 14 layers.

The impact of decreasing the resolution of features is demonstrated in Figure 8 (b). Features at 450 time stamps representing the decimation factor of 2 report an accuracy of 68.13%. Similarly, decimation factors of 3, 4, 5, and 6 represent resolutions of 300, 224, 180, and 150 time stamps, respectively. It can be seen from the results reported that there is a direct relationship between accuracy and resolution of the features. However, both labels are detected until the feature resolution of 180 timestamps, i.e. resolution of features is reduced by factor of 5. Healthy label detection starts to decrease after this and is lowest at a decimation factor of 50, after which the classifier starts to become unstable.

4.4. Potential for the real time implementation and future perspectives

It can be inferred from the block diagram given in Figure 2 (b) and discussion detailed in Section 3 that the system proposed offers potential for an intervention-based system. The voting system at the out-

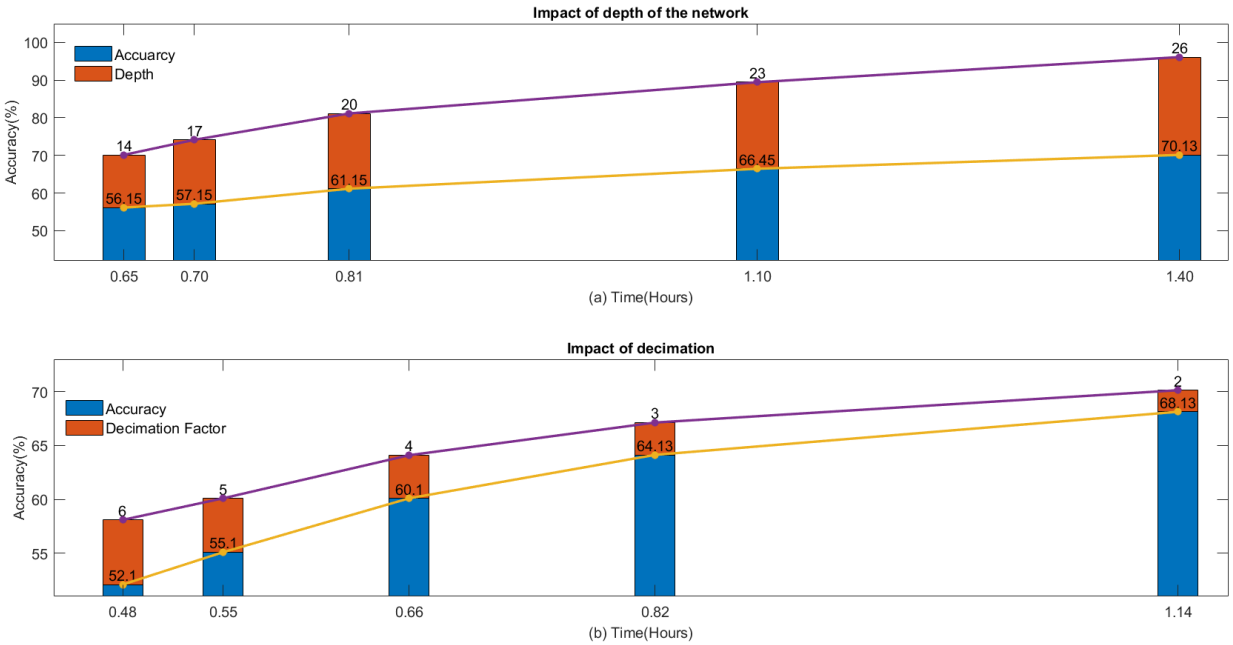


Figure 8: The figures show the impact of the depth of network (number of layers) and the resolution of features, i.e., time stamps that each feature contains, on classification. The reference used consists of 29 layers composed of 1.8 million parameters, which reported an accuracy of 78.66% for binary classification.

put stage provides labels which could be used for human intervention, during which a vet could change labels using their knowledge for top-1 verses top-3 methodology. The provision for human intervention offers clinical significance as well. The system introduced does not require signal-to-image conversion, so saving both computational cost and human-machine effort. The features introduced are robust and they also have physical interpretation. The total system space occupied for various trained networks, including feature-based networks, binary-class and multi-class, vary between 123 kB and 6.83 MB. Moreover, the time taken to train a full 29-layer network for training (multi-class and binary-class networks) is approximately 6 hours. The small size of the network and comparatively short time to train the network offer great potential for real-time implementation for lameness detection in dairy cows. However, the system's computational resources (Ci7, 32 GB RAM and 4 GB GPU RAM) were fully utilized during training.

To improve the generalization ability of the technique proposed, improve binary and multiclass classifications, and reduce the resources utilized, a perspective study could explore the use of networks with parallel structures and additional features with different resolutions. Future studies could also explore the implementation of a framework similar to Agriculture 5.0, blending human technical prowess, knowledge, and experience [22].

5. CONCLUSION

In this study we have introduced LLP-Cow, a technique based on an application-specific deep network and majority voting for detection of the lameness level of a cow. The study presents not only the details regarding sampling efficiency of collected data, feature extraction, deep network, and voting system, but it also presents the implication of each module in both pictorial and tabular form. The most unique aspect of the study is the application of deep learning for lameness detection on a public dataset. Other aspects include the introduction of a very small set of features having physical interpretation, protocols for assessment and a voting-based system. The study can be regarded as a foundational study for future work, setting a benchmark for lameness using multi-sensor data.

6. Acknowledgments

This research was partly supported by PID2023-146620OB-I00, funded by MICIU/AEI10.13039/501100011033 and the European Union's FEDER programme, and partly by the CajaCanaria and la Caixa (2023DIG05).

References

[1] Afridi, H., Ullah, M., Nordbø, Ø., Cheikh, F. A., Larsgard, A. G., 2022. Optimized deep-learning-based method for cattle udder traits classification. Mathematics 10 (17), 3097.

[2] Ahmad, M., Abbas, S., Fatima, A., Ghazal, T. M., Alharbi, M., Khan, M. A., Elmitwally, N. S., 2023. Ai-driven livestock identification and insurance management system. Egyptian Informatics Journal 24 (3), 100390.

[3] Arazo, E., Aly, R., McGuinness, K., 2022. Segmentation enhanced lameness detection in dairy cows from rgb and depth video. arXiv preprint arXiv:2206.04449.

[4] Balasso, P., Marchesini, G., Ughelini, N., Serva, L., Andrighetto, I., 2021. Machine learning to detect posture and behavior in dairy cows: Information from an accelerometer on the animal's left flank. Animals 11 (10), 2972.

[5] Balasso, P., Taccioli, C., Serva, L., Magrin, L., Andrighetto, I., Marchesini, G., 2022. Deep learning performance in predicting dairy cows' behaviour from a tri-axial accelerometer data.

[6] Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B., 2016. Simple online and realtime tracking. In: 2016 IEEE international conference on image processing (ICIP). IEEE, pp. 3464–3468.

[7] Bloch, V., Frondelius, L., Arcidiacono, C., Mancino, M., Pastell, M., 2023. Development and analysis of a cnn-and transfer-learning-based classification model for automated dairy cow feeding behavior recognition from accelerometer data. Sensors 23 (5), 2611.

[8] Boashash, B., 1992. Estimating and interpreting the instantaneous frequency of a signal. i. fundamentals. Proceedings of the IEEE 80 (4), 520–538.

[9] Borchers, M., Chang, Y., Proudfoot, K., Wadsworth, B., Stone, A., Bewley, J., 2017. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. Journal of dairy science 100 (7), 5664–5674.

[10] Bruijnis, M., Hogeveen, H., Stassen, E., 2010. Assessing economic consequences of foot disorders in dairy cattle using a dynamic stochastic simulation model. Journal of dairy science 93 (6), 2419–2432.

[11] Chung, H., Vu, H., Kim, Y., Choi, C. Y., 2023. Subcutaneous temperature monitoring through ear tag for heat stress detection in dairy cows. Biosystems Engineering 235, 202–214.

[12] Dorogush, A. V., Ershov, V., Gulin, A., 2018. Catboost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

[13] Feichtenhofer, C., Fan, H., Malik, J., He, K., 2019. Slowfast networks for video recognition. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6202–6211.

[14] Ferrero, M., Vignolo, L. D., Vanrell, S. R., Martinez-Rau, L. S., Chelotti, J. O., Galli, J. R., Giovanini, L. L., Rufiner, H. L., 2023. A full end-to-end deep approach for detecting and classifying jaw movements from acoustic signals in grazing cattle. Engineering Applications of Artificial Intelligence 121, 106016.

[15] Fuentes, A., Han, S., Nasir, M. F., Park, J., Yoon, S., Park, D. S., 2023. Multiview monitoring of individual cattle behavior based on action recognition in closed barns using deep learning. Animals 13 (12), 2020.

[16] Gao, M., Wang, H., Shen, W., Su, Z., Liu, H., Yin, Y., Zhang, Y., Zhang, Y., 2021. Disease diagnosis of dairy cow by deep learning based on knowledge graph and transfer learning. International Journal Bioautomation 25 (1), 87.

[17] Genemo, M., 2023. Detecting high-risk area for lumpy skin disease in cattle using deep learning feature. Advances in Artificial Intelligence Research 3 (1), 27–35.

[18] Girshick, R., 2015. Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448.

[19] Griffiths, B. E., Mahen, P. J., Hall, R., Kakatsidis, N., Britten, N., Long, K., Robinson, L., Tatham, H., Jenkin, R., Oikonomou, G., 2020. A prospective cohort study on the development of claw horn disruption lesions in dairy cattle; furthering our understanding of the role of the digital cushion. Frontiers in Veterinary Science 7, 440.

[20] He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969.

[21] Higginson, J. H., Millman, S. T., Leslie, K. E., Kelton, D. F., 2010. Validation of a new pedometry system for use in behavioural research and lameness detection in dairy cattle. In: Proc. First North Am. Conf. Precision Dairy Management, Toronto, ON, Canada. Progressive Dairy Operators, Elora, ON, Canada.

[22] Holzinger, A., Fister Jr, I., Fister, I., Kaul, H.-P., Asseng, S., 2024. Human-centered ai in smart farming: Towards agriculture 5.0. IEEE Access.

[23] Ismail, S., Diaz, M., Carmona-Duarte, C., Vilar, J. M., Ferrer, M. A., 2024. Cowscreeningdb: A public benchmark database for lameness detection in dairy cows. Computers and Electronics in Agriculture 216, 108500.

[24] Jarchi, D., Kaler, J., Sanei, S., 2021. Lameness detection in cows using hierarchical deep learning and synchrosqueezed wavelet transform. IEEE Sensors Journal 21 (7), 9349–9358.

[25] Jiang, B., Yin, X., Song, H., 2020. Single-stream long-term optical flow convolution network for action recognition of lameness dairy cow. Computers and Electronics in Agriculture 175, 105536.

[26] Kang, X., Zhang, X., Liu, G., 2020. Accurate detection of lameness in dairy cattle with computer vision: A new and individualized detection strategy based on the analysis of the supporting phase. Journal of dairy science 103 (11), 10628–10638.

[27] Karoui, Y., Jacques, A. A. B., Diallo, A. B., Shepley, E., Vasseur, E., 2021. A deep learning framework for improving lameness identification in dairy cattle. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. pp. 15811–15812.

[28] Kielland, C., Ruud, L., Zanella, A., Østerås, O., 2009. Prevalence and risk factors for skin lesions on legs of dairy cattle housed in freestalls in norway. Journal of Dairy Science 92 (11), 5487–5496.

[29] Leach, M., Barney, S., Dlay, S., Crowe, A., Kyriazakis, I., 2023. Deep learning pose estimation for multi-cattle lameness detection.

[30] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86 (11), 2278–2324.

[31] Li, Y., Shu, H., Bindelle, J., Xu, B., Zhang, W., Jin, Z., Guo, L., Wang, W., 2022. Classification and analysis of multiple

cattle unitary behaviors and movements based on machine learning methods. Animals 12 (9), 1060.

[32] Liu, H., Jia, X., Su, C., Yang, H., Li, C., 2023. Tire appearance defect detection method via combining hog and lbp features. Frontiers in Physics 10, 1099261.

[33] Lodkaew, T., Pasupa, K., Loo, C. K., 2023. Cowxnet: An automated cow estrus detection system. Expert Systems with Applications 211, 118550.

[34] Mann, S., Haykin, S., 1992. Adaptive" chirplet" transform: an adaptive generalization of the wavelet transform. Optical Engineering 31 (6), 1243–1256.

[35] Mason, C., 2007. Preventing Lameness in Dairy Cows; Hoof Lesions; Their Identification, Treatment, Management and Prevention.

[36] Mazrier, H., Tal, S., Aizinbud, E., Bargai, U., 2006. A field investigation of the use of the pedometer for the early detection of lameness in cattle. The Canadian Veterinary Journal 47 (9), 883.

[37] Nagy, S. Á., Kilim, O., Csabai, I., Gábor, G., Solymosi, N., 2023. Impact evaluation of score classes and annotation regions in deep learning-based dairy cow body condition prediction. Animals 13 (2), 194.

[38] Piette, D., Norton, T., Exadaktylos, V., Berckmans, D., 2020. Individualised automated lameness detection in dairy cows and the impact of historical window length on algorithm performance. animal 14 (2), 409–417.

[39] Raza, R., Naz, R., Murtaza, S., Abdelsalam, S. I., 2024. Novel nanostructural features of heat and mass transfer of radiative carreau nanoliquid above an extendable rotating disk. International Journal of Modern Physics B, 2450407.

[40] Riaboff, L., Poggi, S., Madouasse, A., Couvreur, S., Aubin, S., Bédère, N., Goumand, E., Chauvin, A., Plantier, G., 2020. Development of a methodological framework for a robust prediction of the main behaviours of dairy cows using a combination of machine learning algorithms on accelerometer data. Computers and Electronics in Agriculture 169, 105179.

[41] Sadeghi, H., Braun, H.-S., Panti, B., Opsomer, G., Bogado Pascottini, O., 2022. Validation of a deep learning-based image analysis system to diagnose subclinical endometritis in dairy cows. Plos one 17 (1), e0263409.

[42] Sharma, A. K., Sharma, R., Kasana, H., 2007. Prediction of first lactation 305-day milk yield in karan fries dairy cattle using ann modeling. Applied Soft Computing 7 (3), 1112–1120.

[43] Van Hertem, T., Bahr, C., Tello, A. S., Viazzi, S., Steensels, M., Romanini, C., Lokhorst, C., Maltz, E., Halachmi, I., Berckmans, D., 2016. Lameness detection in dairy cattle: single predictor v. multivariate analysis of image-based posture processing and behaviour and performance sensing. Animal 10 (9), 1525–1532.

[44] Van Hertem, T., Maltz, E., Antler, A., Romanini, C., Viazzi, S., Bahr, C., Schlageter-Tello, A., Lokhorst, C., Berckmans, D., Halachmi, I., 2013. Lameness detection based on multivariate continuous sensing of milk yield, rumination, and neck activity. Journal of dairy science 96 (7), 4286–4298.

[45] Vautard, R., Yiou, P., Ghil, M., 1992. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. Physica D: Nonlinear Phenomena 58 (1-4), 95–126.

[46] Vutukuri, K. S., 2023. Dairy cows teat-end condition classification using deep learning.

[47] Vutukuri, K. S., 2023. Second dairy cows teat-end condition classification using deep learning.

[48] Wang, R., Bai, Q., Gao, R., Li, Q., Zhao, C., Li, S., Zhang, H., 2022. Oestrus detection in dairy cows by using atrous spatial pyramid and attention mechanism. Biosystems Engineering 223, 259–276.

[49] Wang, Y., Kang, X., He, Z., Feng, Y., Liu, G., 2022. Accurate detection of dairy cow mastitis with deep learning technology: a new and comprehensive detection method based on infrared thermal images. animal 16 (10), 100646.

[50] Wang, Z., Hua, Z., Wen, Y., Zhang, S., Xu, X., Song, H., 2024. E-yolo: Recognition of estrus cow based on improved yolov8n model. Expert Systems with Applications 238, 122212.

[51] Werema, C., Yang, D., Laven, L., Mueller, K., Laven, R., 3 2022. Evaluating alternatives to locomotion scoring for detecting lameness in pasture-based dairy cattle in New zealand: In-parlour scoring. Animals 12 (6), 703.

[52] Wu, D., Han, M., Song, H., Song, L., Duan, Y., 2023. Monitoring the respiratory behavior of multiple cows based on

computer vision and deep learning. Journal of Dairy Science 106 (4), 2963–2979.

[53] Wu, D., Wu, Q., Yin, X., Jiang, B., Wang, H., He, D., Song, H., 2020. Lameness detection of dairy cows based on the yolov3 deep learning algorithm and a relative step size characteristic vector. Biosystems Engineering 189, 150–163.

[54] Zhao, K., Bewley, J., He, D., Jin, X., 2018. Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique. Computers and Electronics in Agriculture 148, 226–236.

[55] Zhou, M., Koerkamp, P. W. G., Huynh, T. T., Aarnink, A. J., 2022. Development and evaluation of a thermoregulatory model for predicting thermal responses of dairy cows. Biosystems Engineering 223, 295–308.

Appendix A

A.1 Derivation of Eigen-based correlation feature

For $s(t) = [s_1 \;\; s_2 \;\; s_3 \;\; \cdots\cdots \;\; s_n]$, the correlation signal, $C_{nn-1} = [s_{11} \;\; s_{12} \cdots \;\; s_{1n} \;\;\; s_{21} \;\; s_{22} \;\; \cdots \;\; s_{2n} \;\; s_{n1} \;\; s_{n2} \;\; \cdots \;\; s_{nn}]$ is created which is then converted to a Toeplitz matrix, TC. From the Toepltiz matrix, real Eigen values (λ) are calculated which, when sorted in descending order ($\mathbb{R}$) represent the correlation features.

$$TC = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix}$$

$$Corr(TC) \rightarrow \lambda \rightarrow \mathbb{R} \tag{E1}$$

$$\mathbb{R} = [\mathbb{R}_1 \;\; \mathbb{R}_2 \;\; \mathbb{R}_3 \;\; \cdots\cdots \;\; \mathbb{R}_n] \text{ with } \mathbb{R}_1 > \mathbb{R}_2 > \mathbb{R}_3 \cdots\cdots > \mathbb{R}_n$$

A.2 Internal structure of the deep neural network used for assessment of the protocols

Table A. Implementation details of the deep neural network used for classification.

		Learnable Parameters		
Layer #	**Type**	**Weight (S_fxI_sxN_f)**	**Bias**	**Number of parameters**
1	Input	1x252x1	-	0
2	Convolution	5x252x300	300	378300
3	ReLU	1x300x1	-	-
4	Normalization	1x300x1	300	600
5	Convolution	5x300x260	260	390260
6	ReLU	1x260x1	-	0
7	Normalization	1x260x1	260	520
8	Convolution	7x260x220	220	400620
9	ReLU	1x220x1	-	0
10	Normalization	1x220x1	220	440
11	Convolution	7x220x180	180	277380
12	ReLU	1x180x1	-	0
13	Normalization	1x180x1	180	360
14	Convolution	9x180x140	140	226940
15	ReLU	1x140x1	-	0
16	Normalization	1x140x1	140	280
17	Convolution	9x140x100	100	126100
18	ReLU	1x100x1	-	0
19	Normalization	1x100x1	100	200
20	Convolution	5x100x60	60	30060
21	ReLU	1x60x1	-	0
22	Normalization	1x60x1	60	120
23	Convolution	5x60x20	20	6020
24	ReLU	1x20x1	-	0
25	Normalization	1x20x1	20	40
26	Global Average Pooling	1x20x1	-	-
27	Fully Connected	5x20x1	5	105
28	Softmax	1x5x1	-	0
29	Classification	1x5x1	-	0

A.3 Glossary of acronyms

Given below in Table B are the commonly used acronyms in this paper.

Table B. Acronyms and their definitions.

Acronym	Definition
AI	Artificial Intelligence
BCD	Behavioural Change Detection
CNN	Convolutional Neural Network
CPD	Classification based on Parallel Data
CSD	Classification based on Serial Data
DL	Deep Learning
DNN	Deep Neural Network
ISSWT	Inverse SSWT
LLP-Cow	Lameness Level Predictor for Cow
LSTM	Long short-term memory
NCR	Normalization, Convolution, and Rectified linear unit (ReLU)
PDG	Parallel Data Generation
R-CNN	Region-based Convolutional Neural Network
RGB	Red, Blue, Green
SSA	Singular Spectrum Analysis
SDG	Serial Data Generation
SORT	Simple, Online, and Real-time Tracking
SSWT	Synchrosqueezed Wavelet Transform