Lameness Detection in Cows using Hierarchical Deep Learning and Synchrosqueezed Wavelet Transform

Delaram Jarchi, *Senior Member, IEEE,* Jasmeet Kaler and Saeid Sanei, *Senior Member, IEEE*

Abstract—Objectives: Identification of cow lameness is important to farmers to improve and manage cattle health and welfare. No validated tools exist for automatic lameness detection. In this research, we aim to early detect the cow lameness by identifying the instantaneous fundamental gait harmonics from low frequency (16Hz) acceleration signals recorded using leg-worn sensors. Methods: A triaxial accelerometer has been worn on each cow leg. Synchrosqueezed wavelet transform (SSWT) has been applied to acceleration signals to generate the initial time-frequency spectrum related to the gait. This spectrum is given as an input to a designed deep neural network including time-frequency based long short-term memory (LSTM) to estimate instantaneous frequencies at each time point. An inverse SSWT (ISSWT) is then used to recover the gait harmonic and to estimate an enhanced spectrum. Results: Validation of instantaneous frequencies has been provided for each cow leg (combined signals from 23 cows) and the time-series cross validator across the three folds are provided. The average of mean squared errors in frequencies across 3 folds for each leg is obtained as 0.036, 0.033, 0.044 and 0.042 for left-front, right-front, right-back and left-back legs, respectively. Conclusion: Estimation of instantaneous gait frequencies is proved useful for identification of cow gait phases, lameness detection, accurate estimation of gait speed, coherency in movement among the legs and identification of non-gait episodes. Moreover, the proposed method can be used as a new frequency ridge estimation method exploiting SSWT for many other applications.

Index Terms—Accelerometer, cow lameness, deep neural network, gait analysis, LSTM, SSWT.

I. INTRODUCTION

LAMENESS management, prevention and treatment are crucial for improving animal welfare which contribute to both health and economic implications on dairy farms. Traditionally, visual inspection of cows has been used for locomotion scoring by farmers. There is evidence to suggest that using visual inspection, farmers underestimate the lameness level. Early detection of lameness is thus useful for improving the cow health and welfare and economic benefit to dairy industry.

Accelerometers have been used for human and animal gait analysis [1][2]. It has been documented that the use of cow-worn sensors such as accelerometers is becoming an emerging

D. Jarchi is with the Department of Computer Science and Electronic Engineering, University of Essex, Colchester, UK, e-mail: delaram.jarchi@essex.ac.uk.
J. Kaler is with thee School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK, jasmeet.kaler@nottingham.ac.uk.
S. Sanei is with the School of Science and Technology, Nottingham Trent University, Nottingham, UK, e-mail: saeid.sanei@ntu.ac.uk.

technology in diary industry meanwhile the accelerometry-based gait measurements are still rather underdeveloped [3]. Accelerometers and gyroscopes have been used for sheep behaviour and lameness classification [4][5].

Leg-mounted accelerometers are used in [6] to detect lameness in cows. Collar, leg, and ear-worn tri-axial accelerometer are used for prediction of lameness in sheep using machine learning with high accuracies of above 85% for sheep lameness detection [5]. Monitored activities included walking, standing, grazing and lying [5]. An accuracy of 91.9% has been reported for detection of lameness in diary cattle using acceleration signals [7]. Unilateral hind limb lameness and foot pathologies have been investigated in [8] using two high-frequency accelerometers (400 Hz) enabling extraction of gait cycle, stance phase and swing phase. The relationship between the gait characteristics and locomotion score has been investigated in diary cows with hoof lesion [9].

For measurement of the changes in cow gait or behavior, the use of sensor technologies has been explored in [10] and [11]. Similar to sheep, accelereometers have also been used for classification of behaviours (lying, standing, and feeding) in cows [12]. Leg-mounted accelerometer produced a high accuracy for classification of lying while neck-mounted accelerometer led into a high accuracy for classification of feeding [13]. Neck-mounted sensors measuring both position and accelerations have been used in [14] to quantify the feeding behavior changes associated with the lameness in dairy cattle. Kinematic, kinetic and indirect methods have been reviewed in [15].

Several variables such as walking duration, lying duration, standing duration, total number of steps, and step frequency, have been derived for activity-based lameness analysis and found useful for early identification of lameness. However, studies on lameness detection in cattle using accelerometers have reported varying accuracies from 66%-91%. Moreover, most of the studies reporting higher accuracies have used very high frequency accelerometers (400Hz) which are not very practical for multiple recordings in commercial environment [12].

Better understanding of behavioral gait analysis requires accurate and reliable estimation of gait variables and their correlation. In this research, we propose a robust technique to estimate the main gait harmonic by reconstructing the time-frequency spectrum of the low-frequency (16Hz) acceleration signals exploiting a deep learning technique. This harmonic can be used to estimate gait frequency, duration of gait and

non-gait periods, and the lameness base on correlation analysis of various legs and energy analysis from the reconstructed time-frequency spectrum. The remainder of the paper is as follows. In Section II, the basic methods used here for spectrum reconstruction are first explained. These include synchrosqueezed wavelet transform (SSWT), inverse SSWT (ISSWT) and long short-term memory (LSTM) networks. Then, the overall deep network based structure is explained in Section III. LSTM-based deep network has been used for two different purposes. It is first combined with SSWT to extract the main gait harmonic. The first LSTM-based deep neural network can be considered as an automatic approach to extract the time-frequency ridge. It can also be used for enhancement of the time-frequency spectrum of the acceleration signals. The second LSTM-based deep neural is used to detect the lameness based on the cow gait behaviour history. In Section IV, the results of applying the proposed model to triaxial acceleration data are provided. Finally, Section V concludes the paper.

II. BASIC METHODS

A. Synchrosqueezed Wavelet Transform

Time-frequency transform is recommended for nonstationary signal analysis and feature extraction. SSWT has been often used for auditory signal analysis where a time-frequency reassignment technique is applied for improving the time-frequency spectral presentation of the signal [16]. This approach has been previously used in sleep analysis applications [18][19]. SSWT can be outlined using three major steps which are explained as follows.

Step 1: Let's start from continuous wavelet transform (CWT) of signal **s** defined as:

$$W_s(a,b) = \int_{-\infty}^{\infty} s(t)a^{-1/2}\overline{\psi}(\frac{t-b}{a})dt \qquad (1)$$

where ψ presents the selected mother wavelet ($\overline{\psi}$ is complex conjugate form), t presents the time index, a is the wavelet scale, and b is the position parameter. One necessary condition on the selected mother wavelet is to assure that it is concentrated on the positive side of frequency axis; i.e., $\hat{\psi}(\varepsilon) = 0$ for $\varepsilon < 0$. To simplify understanding of the SSWT, consider defining a purely harmonic input signal presented as $s(t) = A\cos(\omega t)$. By exploiting Plancherel's theorem [16], the CWT can be obtained as:

$$
\begin{aligned}
W_s(a,b) &= \int_{-\infty}^{\infty} s(t)a^{-1/2}\overline{\psi}(\frac{t-b}{a})dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{s}(\varepsilon)a^{1/2}\overline{\hat{\psi}}(a\varepsilon)e^{ib\varepsilon}d\varepsilon \\
&= \frac{A}{4\pi}\int_{0}^{\infty} [\delta(\varepsilon - \omega) + \delta(\varepsilon + \omega)]a^{1/2}\overline{\hat{\psi}}(a\varepsilon)e^{ib\varepsilon}d\varepsilon \\
&= \frac{A}{4\pi}a^{1/2}\overline{\hat{\psi}}(a\omega)e^{ib\omega}
\end{aligned}
$$
$$(2)$$

In the cases where $\hat{\psi}(\varepsilon)$ is concentrated around $\varepsilon = \omega_0$, it is straightforward to observe that $W_s(a,b)$ will be concentrated around $a = \omega_0/\omega$ which is spreading out over a region around

the horizontal line $a = \omega_0/\omega$. When $\omega = \omega_0/a$ is similar but not necessarily exactly identical to the actual instantaneous frequency (IF) of the input signal, some undesired non-zero energy for $W_s(a,b)$ will appear. One main objective of synchrosqueezing is to move these excessive energy away from ω. The reassigning technique has been proposed to tackle this issue.It aims to reassign the frequency locations as close as possible to the actual IF.

Step 2: The candidate IFs ($\omega(a,b)$) are first obtained by applying the following equation, for which $W_s(a,b) \neq 0$:

$$\omega(a,b) = -i(W_s(a,b))^{-1}\frac{\partial}{\partial b}W_s(a,b) \qquad (3)$$

It is easy to demonstrate that for a purely harmonic signal $s(t) = A\cos(\omega t)$, $\omega(a,b)$ are equal to signal frequency ω [16]. Estimation of candidate IFs is useful to recover actual frequencies that is explained in the next step.

Step 3: Apply *synchrosqueezing* to CWT by using $(b,a) \Rightarrow (b,\omega(a,b))$ and performing a frequency re-allocation technique, where the time-scale domain is mapped into the time-frequency domain. Presenting ω_l as the nearest frequency to the original $\omega(a,b)$, each value of $W_s(a,b)$ is transformed into $T_s(\omega_l,b)$:

$$T_s(\omega_l,b) = (\Delta\omega)^{-1} \sum_{a_k:|\omega(a_k,b)-\omega_l|\leq\frac{\Delta\omega}{2}} W_s(a_k,b)a_k^{-3/2}(\Delta a)_k$$
$$(4)$$

where $\Delta\omega$ is the width of frequency bins $[\omega_l - \frac{1}{2}\Delta\omega, \omega_l + \frac{1}{2}\Delta\omega]$, $\Delta\omega = \omega_l - \omega_{l-1}$, $(\Delta a)_k = a_k - a_{k-1}$ and $T_s(\omega_l,b)$ represents the synchrosqueezed transform at the centres ω_l of successive frequency bins. To derive the SSWT transform, first, the reassigned frequencies are calculated for all the scales at each fixed time point b using equation (3). Then, for each desired IF of ω_l, $T_s(\omega_l,b)$ is calculated by summing all $W_s(a,b)$ taking into account the distance between the reassigned frequency $\omega(a,b)$ and ω_l. This distance must be within a specified frequency bin width ($\Delta\omega/2$).

Once SSWT has been applied to a desired signal ($s(t)$), we use $T_s(\omega_l,b)$ (equation (4)), instead of $W_s(a,b)$ (equation (1)). One interesting aspect of the synchronosqueezing stage is that the original signal can be analytically reconstructed [16]. Overall, $T_s(\omega_l,b)$ has sharper concentrations around the actual IFs of the original signal. Therefore, the resulting spectrum from SSWT is expected to be more sparse than $W_s(a,b)$ obtained by CWT.

B. Inverse Synchrosqueezed Wavelet Transform

ISSWT can be used for signal reconstruction using a similar concept of inverting the CWT and integrating over the frequencies that are expected to be associated with a desired component. Consider a fully discretized version of equation (4) which is represented as $\tilde{T}_s(w_l,t_m)$. ISSWT receives two inputs; one is SSWT spectrum and the other one is a set of fixed frequency ranges specified by the user, or frequencies obtained by applying a standard least-squares ridge extraction method [17]. Let's denote these frequencies as $l \in L(t_m)$, where $m = 0,...,n-1$, $t_m = t_0 + m\Delta t$, $a_j = 2^{j/n_v}\Delta t$,

$j = 1, ..., Ln_v$, Ln_v as log-scale samples of scale a and n_v is a user-defined parameter affecting thee number of scales. Then, the kth mode signal can be reconstructed from:

$$s_k(t_m) = 2R_\psi^{-1}\Re(\sum_{l \in L(t_m)} \tilde{T}_s(w_l, t_m))$$ (5)

where R_ψ presents a normalisation constant defined in [20] and $\Re$ refers to real part of a complex number. By using both a set of fixed frequencies (e.g. [0.01-3]Hz) and those by the least-squares ridge extraction method, it is possible to reconstruct the time-domain signal (k^{th} mode).

Fig. 3. Diagram of the proposed model to reconstruct the spectrum containing the main gait harmonic.
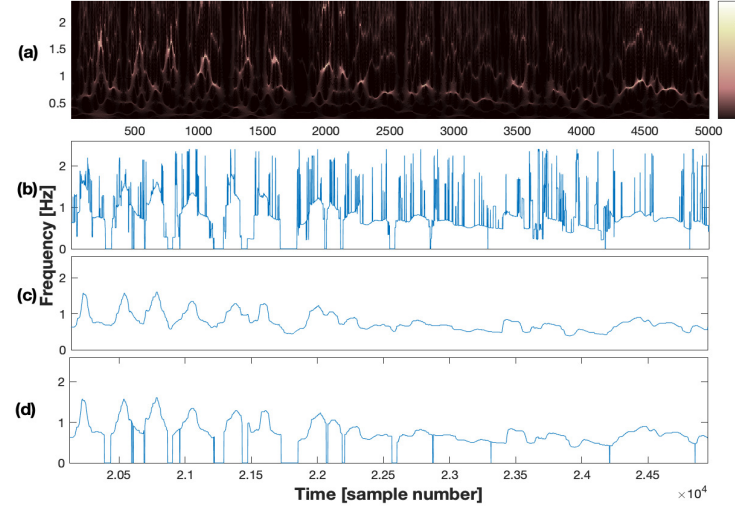
Fig. 4. (a) Input SSWT from combination of two axes (anterior-posterior and superior-inferior). (b) Maximum-energy time-frequency ridge with no penalty for changing frequency. (c) Maximum-energy time-frequency ridge with penalty value equal to 1. (d) Reference frequency obtained by applying a threshold of 0.1 to (b) to find zero frequencies to be combined with (c).
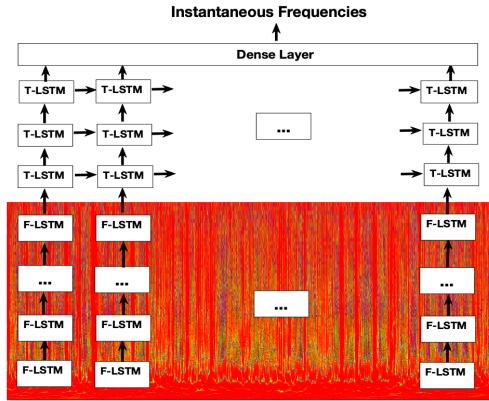

Fig. 1. Architecture of the proposed neural network to estimate the instantaneous frequencies. The TF spectrum given by SSWT method is used where FLSTM scan frequencies first, then, the time is scanned using TLSTM.
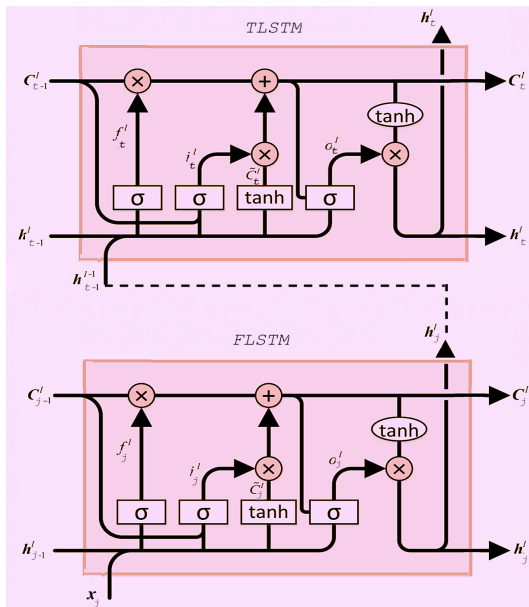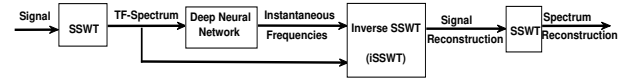

Fig. 2. The structure of the last FLSTM cell output which is used as the input into the first cell of time-frequency LSTM (TF-LSTM) layer for temporal analysis over time.

C. Long short-term memory

LSTM, as a recurrent neural network (RNN) has been proposed by Hochreiter et al. [21]. LSTM differs from ordinary neural network in that the nodes within the same hidden layer are connected. This means that the input to the hidden layer at each time point not only includes the output from the previous layer, but also the output of the same hidden layer at the previous time point [22]. LSTM development aimed to solve the disappearance or explosion of gradient descent problem [23] in basic RNNs. This problem is well modelled by inclusion of the following four components:

- Memory units: store the temporal state of the network
- Input gates: modulate the input activations into the cells
- Output gates: modulate the output activations of the cells
- Forget gates: adaptively reset the cell's memory

LSTM has been used in many applications such as automatic speech recognition [24][25][26] and voice conversion [27]. Processing of the data sequence is a desired application of LSTM exploiting the data history. Therefore, LSTM is a suitable candidate to process acceleration signals as time-series. LSTM is designed based on the following equations:

$$
\begin{aligned}
i_j &= \sigma(W_i.[x_j, h_{j-1}] + b_i) \\
f_j &= \sigma(W_f.[x_j, h_{j-1}] + b_f) \\
o_j &= \sigma(W_o.[x_j, h_{j-1}] + b_0) \\
\hat{c}_j &= tanh(W_c.[x_j, h_{j-1}] + b_c) \\
c_j &= f_j \odot c_{j-1} + i_j \odot \hat{c}_j \\
h_j &= o_j \odot tanh(c_j)
\end{aligned}
$$ (6)

where x_j is the input vector to the LSTM unit, i_j, f_j, $\hat{c}_j$ and o_j denote the input gate's activation vector, forget gate's
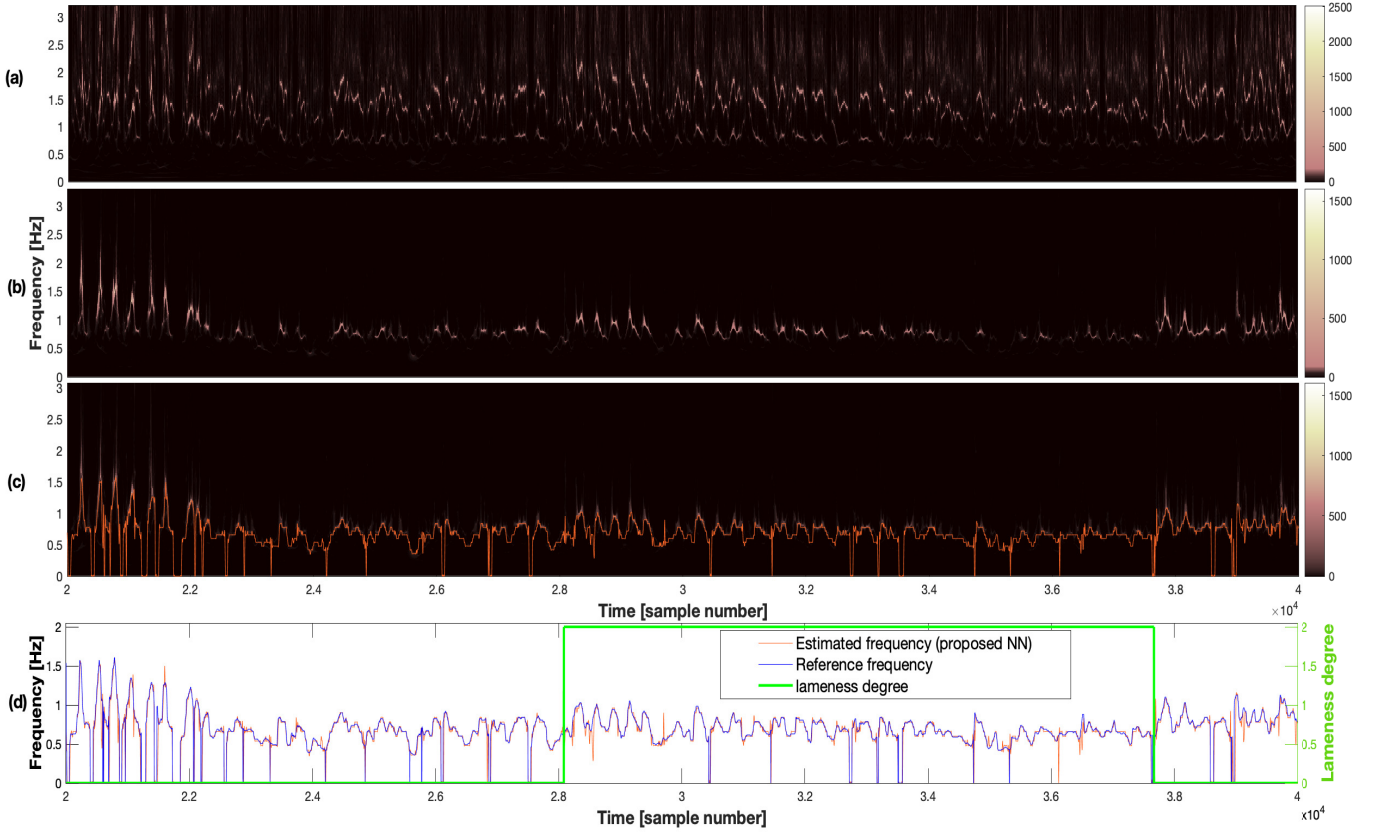
Fig. 5. (a) Noisy input spectrum of the acceleration signal from anterior-posterior axis. (b) Reconstructed spectrum using the proposed DNN and SSWT. (c) Estimated instantaneous frequencies. (d) Actual and estimated instantaneous frequencies with inclusion of lameness degree in the same plot.
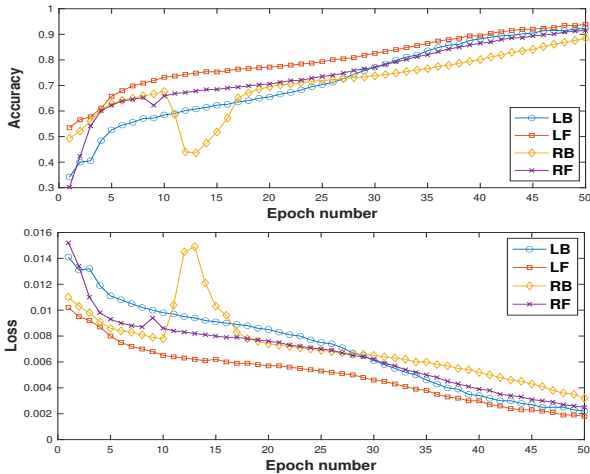

Fig. 6. Loss and accuracy during 50 epochs for four cow legs LB, LF, RB and RF.

activation vector, memory cell input activation vector and output gate's activation vector, respectively, at step j, c_j is the cell state vector, h_j is the output vector of the LSTM unit and

$\odot$ is an element-wise dot product. W presents different weight matrices, i.e. W_f is the weight matrix from the forgetting gate, W_i is the weight matrix of the input gate, and W_o is the weight matrix of the output gate. b_f, b_i, b_o and b_c are bias weights to be learned during the training process and σ is the sigmoid function. In this research, time-based LTSM (TLSTM) and frequency-based LSTM (FLSTM) cells are used for incorporation of both time and frequency information of the acceleration signals which are explained in the following section.

III. PROPOSED FRAMEWORK

One main objective of the proposed method is to provide a robust technique for estimation of cow gait instantaneous frequencies. Here, we have designed a deep learning framework which uses SSWT spectrum to train the model and output instantaneous frequencies. The time-frequency spectrum derived by the SSWT and reference frequencies (obtained by applying various ridge estimation methods with manually set frequencies and thresholds) have been used to train the deep networks. The main motivation behind developing the first deep neural network is to enhance the initial SSWT spectrum (calculated from a noisy/raw signal) for extraction of the desired gait harmonic. The proposed deep network is

TABLE I
STRUCTURE OF THE DNN FOR ESTIMATION OF INSTANTANEOUS
FREQUENCIES, LAYER TYPES, OUTPUT SHAPE AND NUMBER OF
PARAMETERS. NONE REFERS TO THE FIRST DIMENSION THAT WILL BE
REPLACED AS BATCH SIZE DURING FITTING THE MODEL.

TF deep networks		
Layer (type)	Output Shape	Parameter no.
Input Layer	(None, 8, 31, 20)	0
Time Distributed	(None, 8, 256)	28368
LSTM	(None, 8, 256)	525312
LSTM	(None, 8, 256)	525312
LSTM	(None, 256)	525312
Dense	(None, 52)	13364

expected to automatically perform frequency extraction and spectrum enhancement. It can be considered as an objective and automatic frequency ridge estimation technique which can be adjusted for many other applications without requiring to manually setting up parameters of ridge estimation method for each windowed segment of input signal.

Based on the developed network, the frequencies are scanned in the first layer after formation of a 2D time-frequency presentation. Then, the temporal data of several 2D segments (forming 3D data) are used as input to scan the temporal variation of frequencies along the time axis.

In Fig. 1, the architecture of the proposed deep network is shown. From this figure, FLSTM is supposed to represent the segmented frequencies along the frequency axis. The FLSTM also contains frequencies at the corresponding segmented time window. We have used TimeDistributed layer along with FLSTM to manage the time in the FLSTM for many-to-many mappings. In the subsequent TLSTM layers (3 TLSTM layers), the data is scanned over the time domain to explore the temporal variations. The stacked segments of the above 2D presentations form the 3D data to be traced by TLSTM. Finally, a dense layer (with softmax activation layer) has been used to obtain the output instantaneous frequencies. The input and output data are explained in the following sub-sections.

1) Input data to the deep network: To train the deep network presented in Fig. 1, 4D data should be constructed from the initial spectrum of the acceleration data. The 4D data is formatted as $(time\text{-}batch,\ time\text{-}chunk\text{-}len,\ frequency\text{-}batch,\ frequency\text{-}chunk\text{-}len)$. The $time\text{-}batch$ is specified during run-time. For FLSTM, the frequency axis is divided into overlapping fixed width length ($frequency\text{-}chunk\text{-}len$) segments. In our experiment, we have used 50% overlap between the frequency segments. The $time\text{-}chunk\text{-}len$ presents the time window of the 3D data used as input into TLSTM. In the following section, all the parameters of the DNN are provided for detailed illustration. In addition, the relationship between the last cell in the FLSTM layer with its output as the input to the TLSTM layer is depicted in Fig. 2.

2) Output of deep network: Output of the DNN has a dimension equal to the number of frequency bins. To convert the output of the DNN to instantaneous frequency, at each time point, maximum of the DNN outputs of all bins is calculated. Then, the corresponding frequency of the selected bin (with maximum value of DNN output) is considered as the estimated instantaneous frequency.

Once the instantaneous frequencies are estimated, they are used as the input to the ISSWT along with the initial SSWT spectrum for reconstruction of the main harmonic of the acceleration signals recorded from leg-worn sensors. By applying SSWT to the reconstructed signal, a clear spectrum is expected (see Fig. 3). This spectrum can be used for many purposes such as accurate estimation of gait speed, correlation analysis of various legs, detection of lameness and cow gait phases. These are discussed in the next section supported by some examples.

IV. RESULTS

In this section, the results of applying the proposed framework to the acceleration signals recorded from four legs; left-front (LF), left-back (LB), right-front (RF), right-back (RB), are presented. Cows were from one farm where they were housed indoors and walked on a concrete floor for up-to 5 minutes in a straight alleyway. Ethics Approval (number 2137171016) was obtained from School of Veterinary Medicine and Science, University of Nottingham Ethics committee.

A custom-made wearable device based on the Intel$^{®}$ Quark^{TM} SE microcontroller C1000 integrating Bosch BMI160 (Bosch-sensortec.com, 2016), flash memory and a low-power wide-area radio module was used for data acquisition. BMI160 included a low-power inertial measurement unit (IMU), featuring a 16 bit triaxial accelerometer and a 16 bit gyroscope. The sampling frequency of the signals is 16Hz. Lameness is classified as locomotion score 0/1 as non lame and 2 and above as lame [28]. For each leg, acceleration signals of all cows are concatenated. This allows temporal analysis of individual cow legs across subjects. SSWT has been applied to time windowed segments (5000 samples) of the acceleration signals with their mean removed. It is worth noting that the mean removal of accelerations is the only pre-processing done before apply the SSWT to minimise the baseline shift and potential effect of drift. To segment the final time-frequency matrix given by the SSWT, the frequency axis was initially limited to 3.2Hz as the maximum walking-cycle frequency to be applied to the deep network.

At each time point, a window of length 20 frequency bins ($frequency\text{-}chunk\text{-}length$) is formed. The frequency windows are moved for every 10 frequency bins (50% overlap). Considering the initial 310 frequency bins, 31 segments are formed each with a length of 20. Therefore, $frequency\text{-}batch$ is 31. In this case, at each time point a 2D matrix of 31×20 is constructed. The $time\text{-}chunk\text{-}length$ has been set to 8. Therefore, the 3D tensor of $8 \times 31 \times 20$ are formed to be used as input to the deep network. When forming these 3D tensors over time, a 4D data input is formed considering the $time\text{-}batch$. To capture the frequency dynamics of cow gait, all the acceleration signals of one cow leg (LB/LF/RB/RF), from 23 cows, are combined and fed to the deep network. The number of samples for each leg is 204999 where 25624 samples is the total number of total time batches.

A summary of the DNN including the layer type, output shape and number of parameters is shown in TABLE I. In

TABLE II
VALIDATION OF DEVELOPED LSTM-BASED FRAMEWORK TO ESTIMATE
THE INSTANTANEOUS FREQUENCIES USING 3-FOLD (k IS THE FOLD
NUMBER) TIME-SERIES CROSS-VALIDATOR.

Leg	metrics	$k=1$	$k=2$	$k=3$
LB	train#	5095	10188	15281
	test#	5093	5093	5093
	MSE	0.0369	0.0503	0.0403
LF	train#	3587	7171	10755
	test#	3584	3584	3584
	MSE	0.0271	0.0201	0.0608
RB	train#	3898	7794	11690
	test#	3896	3896	3896
	MSE	0.0460	0.0503	0.0348
RF	train#	5731	11462	17193
	test#	5731	5731	5731
	MSE	0.0285	0.0314	0.0392

this table, the final number of frequency bins is set to 52. A frequency resolution of 0.06 has been considered for output frequencies in the training stage considering the reference frequencies where maximum frequency is set to 3.12Hz.

A. Cow Gait Spectral Analysis

1) Deriving reference frequency: We have applied a time-frequency ridge estimation method to the initial spectrum derived by SSWT. First, a function called *tfridge* integrated in Matlab toolkit (The MathWork Inc.) has been used. This function gives both the time-dependent frequency ridge (called *fridge*) and the row-index vector associated with the maximum-energy ridge (called *iridge*). We need an automatic approach to estimate the maximum-energy time-frequency ridges while taking discontinuance in the spectrum (zero frequencies) into account. This is the basis of the first proposed deep neural network. We have manually set a threshold of 0.1 to assign a frequency of zero to the positions of cow stops. This threshold has been set manually with reference to the neighboring frequencies. An example, illustrating the estimated time-frequency ridges and the reference extraction using SSWT spectrum is shown in Fig. 4. In this figure, the SSWT spectrum of the signal derived from two accelerometer axes (the square root of multiplication of anterior-posterior and superior-inferior axes to get a more enhanced spectrum for reference extraction) of LB which is shown in Fig. 4(a). The estimated maximum-energy time-frequency with penalty value equal to zero is shown in Fig. 4(b). This penalty value has been used in the *tfridge* of Matlab toolkit and controls the changes in frequency (the higher value for penalty value; the lower changes in adjacent frequencies). By increasing the penalty value to 1, smoother frequencies can be obtained as shown in Fig. 4(c). In order to obtain the final reference frequencies, smoothed frequencies must be combined with those episodes of zero frequencies. These episodes are obtained from Fig. 4(b) (where the penalty value is zero) by applying a threshold of 0.1. Then, the final reference is shown in Fig. 4(d).

Here, we consider that in places where the transition between hoof contacts is rather slow, frequency of zero has not been used and this is considered as a low speed gait cycle rather than making it as no-motion. These are related to very short stops and can make training of the DNN very difficult. In addition, we avoid very short duration of no-walking. Moreover, to apply *tfridge*, the fixed frequencies are set manually for each leg.

An example illustrating the operation of DNN for instantaneous frequency estimation is shown in Fig. 5. The LB signal is segmented from 20000 to 40000 samples. The initial time-frequency spectrum directly estimated by the SSWT is shown in Fig. 5(a). From this figure, there are two (main) very close harmonics and some noise/artefact related components around them. There is also discontinuance in the spectrum which is related to the case that the cow has stopped walking. It is a difficult task to identify the main gait harmonic from such spectrum while identifying places where the cow stops.
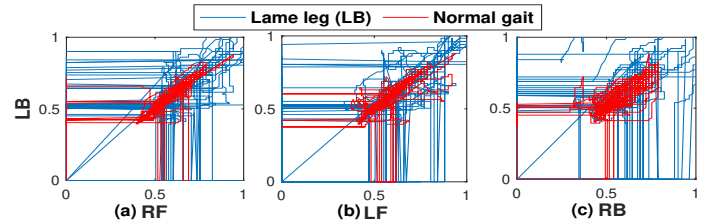

Fig. 7. Pairwise comparison of simultaneous frequencies of LB (lame leg) versus other legs (normal gait).
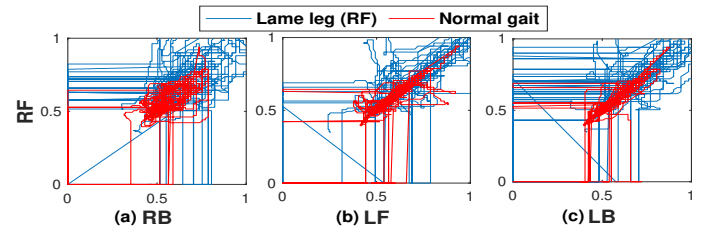

Fig. 8. Pairwise comparison of simultaneous frequencies of RF (lame leg) versus other legs (normal gait).

The output of DNN includes the estimated instantaneous frequencies as shown in Fig. 3. These frequencies have been used to produce a new time-frequency spectrum which is more enhanced and only contains the main cow gait harmonic. To do this, the estimated instantaneous frequencies by the DNN are used to extract the row-index vectors (to re-produce new *iridge*). The row-index vectors are used by the ISSWT along with the initial spectrum to reconstruct the gait harmonic and subsequently the enhanced spectrum which is shown in Fig. 5(b) for the selected segment. In Fig. 5(c), the instantaneous frequencies (output of DNN) are overlaid with the reconstructed spectrum (Fig. 5(b)). These instantaneous frequencies versus the reference frequencies are shown in Fig. 5(d). In this figure, the lameness degree has been also provided. The main gait harmonic can be used for detailed cow data analysis such as speed of cow gait and identification of non-gait episodes as well as measurement of duration of non-gait period. All of these information are useful in detection of any abnormal gait or lameness.

As noted earlier, the acceleration signals from the same leg of all cows (23 cows) are combined together and used to train the DNN. The accuracies and losses for each epoch are shown in Fig. 6. A total of 50 epochs and batch size of 48 are used. For the model settings, we used mean squared error (MSE) for loss, *adam* as optimizer and *accuracy* as the metric to evaluate the DNN performance. The final training accuracies at 50th epoch are 92% , 88% , 93% and 91% for LB, RB, LF and RF, respectively.
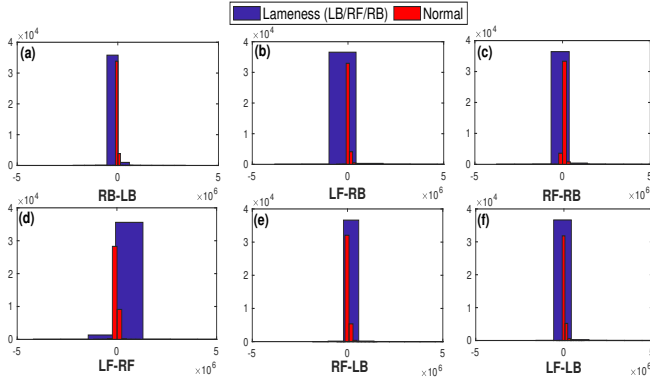

Fig. 9. Histogram of energies for lame and normal gaits for pairwise comparison of three legs (LB, RF, and RB).

2) Time-series cross-validator: For validation of the proposed framework, since a time-series is constructed exploiting temporal dependencies between frequencies, data shuffling seemed inappropriate. Therefore, we have used a specific time-series cross-validator using a python package called *tscv*. Based on the provided *split* function of this package, the time series data samples are split such that the test indices become higher as the folds increase. Also, unlike the standard cross-validation methods, successive training sets are supersets of those that come before them. Considering k-fold cross validation, in the kth split, the first k folds are returned as the training set while the $(k+1)$th fold is considered as the test set. The results of 3-fold cross validation is shown in Table II. For each leg, and in each fold the number of training and test sample size are shown. The MSE of instantaneous frequencies are calculated for each leg and each fold.

3) Correlation analysis and lameness detection: Simultaneous correlation analysis of cow gait can reveal important information regarding lameness. This is useful to identify for which leg the force is more exerted and how symmetrical the cow gait cycles are. Here, a case study for presenting simple correlation among pairwise frequencies of various legs has been presented for lameness monitoring application.

A data segment has been selected where only one leg (LB) is lame and the other legs (i.e. RF, LF and RB) are normal. In addition, another segment (similar size) has been selected where all the legs are normal. Pairwise frequencies for LB versus RF, LF and RB are shown in Fig. 7. From this figure for normal legs, the gait frequencies are towards lesser frequencies (or lower speed) than when the leg is lame. A similar comparison of correlation between various legs is provided in Fig. 8 where RF is lame and other legs are non-

Deep networks for lameness detection		
Layer (type)	Output Shape	Parameter no.
Input Layer	(None, 1, 16)	0
LSTM	(None, 1, 200)	173600
LSTM	(None, 1, 200)	320800
LSTM	(None, 200)	320800
Dense	(None, 100)	20100
Dense	(None, 1)	101

lame. From this figure, it can be seen that there are more non-gait phases for LB versus RF. Although this observation might not be generalised, it can demonstrate the load timings of lame/non-lame legs.

4) Energy analysis: Associated energy of the gait harmonic in spectral domain has shown to be an important feature in lameness classification too. To demonstrate the effectiveness of energy analysis of the main gait harmonic, a simple test was carried out. First, the gait signals of all the legs were segmented where all legs were normal was selected. Then, other data segments with similar size to thee normal gaits were selected while at least one of these legs was lame (LB, RB, RF). The maximum energy of the gait harmonic considering a total of four frequency bins around the estimated frequencies are calculated. The histogram of difference between these energies for each two legs are shown in Fig. 9. It can be clearly seen from this figure that, for the normal segments, the energies are centered around zero while for the case of lameness, the energies are skewed towards one lame leg. Detailed analysis of energies reveals new insight into early lameness detection.

5) Gait phases: Accurate detection of gait phases is crucial to understand the cow gait behaviour, which is possible using the reconstructed signal in time-domain (Fig. 3) and by applying peak detection techniques. This is particularly important while using a low-frequency accelerometer such that used in the presented research and can be further explored in future studies.

B. Lameness detection

For detection of lameness, first, we have directly used the output of developed TF-LSTM presenting instantaneous frequencies (for each leg) as input to another DNN. Then, based on the analysis examples explored in Section IV-A, more features such as spectral energies around instantaneous frequencies were used. We also segmented the data where the simultaneous sensor readings from all legs were available. Therefore, a reduced number of samples were used in the second DNN. Features include: a stream of instantaneous frequencies 4 legs (4D), the associated maximum and mean energy in the spectral domain (8D) and the absolute value of squared root of multiplication of two accelerometer axes for each leg (4D) were used. Therefore, the new input data dimension was 16. As seen from Table III, we used three LSTM layers and two Dense layers with sigmoid activation

TABLE IV
RECEIVER OPERATING CHARACTERISTICS USING TEST DATASET.

|	metric	LSTM output					
		Original	avg (1sec.)	avg (5sec.)	avg (15sec.)	avg (30sec.)	avg (60sec.)
Fold#2	Sensitivity	0.7883	0.8611	0.9080	0.9505	0.9975	0.9897
	Specificity	0.6572	0.8490	0.9396	0.9992	0.9889	0.9835
	F1-score	0.7448	0.8586	0.9238	0.9743	0.9934	0.9870
Fold#3	Sensitivity	0.6347	0.6307	0.6476	0.6502	0.8041	0.9014
	Specificity	0.7271	0.9481	0.9990	1.0000	1.0000	1.0000
	F1-score	0.6613	0.7486	0.7856	0.7880	0.8914	0.9481
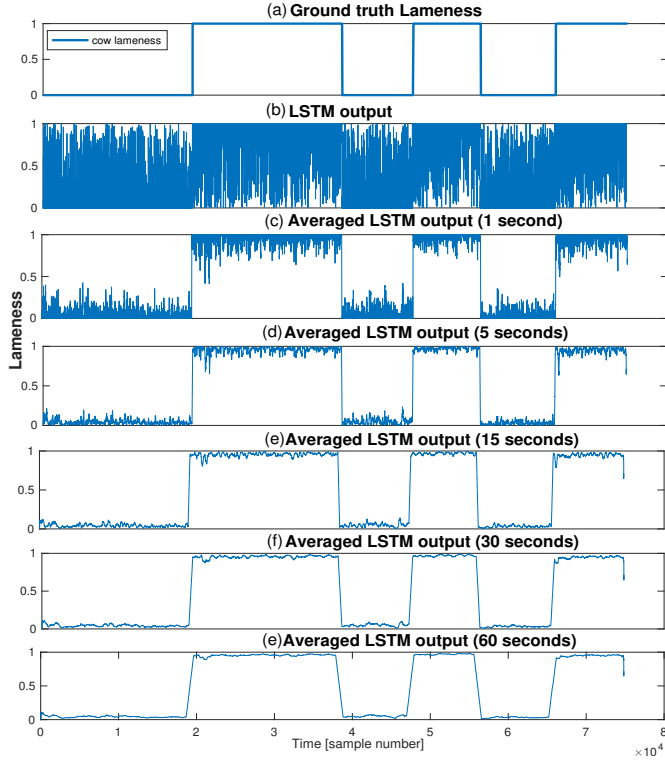
Fig. 10. (a) Ground truth lameness. (b) Original LSTM output using DNN model in Table III. Moving average filter applied to LSTM output using (c) 1 second, (d) 5 seconds, (e) 15 seconds, (f) 30 seconds, and (g) 60 seconds windows.

function. The loss was set to *binary_crossentropy*, optimizer as *adam* and metrics as *accuracy*. The input data was reformed as (74831, 1, 16).

The ground truth lameness is shown in Fig. 10(a) for the selected data segments with all recordings available. In this figure, the interval with amplitude 1 refers to a segment of lame cow walking and the 0 amplitude interval refers to normal walking segment (the non-lame and lame segments can come respectively from a healthy cow and a cow suffering from lameness). In Fig. 10(b), the LSTM output has been shown when streaming the entire data into the LSTM. The final loss and accuracy were obtained as 0.0856 and 0.9673, respectively. However, for binary classification of lameness, it was found that by applying a moving average filter, lameness patterns became more evident. Next, we used time-series cross-validator using 3-fold cross validation. The first fold did

not use a balanced data size for training the non-lame and lame conditions. Therefore, we focused on the test results using the second and third folds.

In the second fold, 37417 samples were used for training while the next 18707 were used for testing. These 18707 samples are shown in a dashed box in Fig. 10(a) which are separated from the data and shown in Fig. 11(b) The noisy output shown in Fig. 11(c) did not provide a good lameness detection. However, after applying a moving average filter, using 1 second (16 samples), 5 seconds (80 sample), 15 seconds (240 samples), 30 samples (480 samples) and 60 seconds (960 samples), lameness detection converged into the ground truth (Fig. 11(d-h)). Similar analysis was performed for the third fold (See Fig. 12), where 56124 samples were used for training while the next 18707 were used for testing. The calculated sensitivity, specificity and F1-score are reported in TABLE IV for fold #2 and #3. It is evident that by increasing the windowed size in averaging, these calculated parameters were approached to 1.

V. CONCLUSION

In this research, a robust frequency ridge estimation using deep learning and SSWT has been proposed. Moreover, an effective approach for automatic detection of cow lameness using accelerometer data has been investigated. The proposed DNN framework has been used to reconstruct the spectrum of the noisy low frequency acceleration signals (16Hz) including the main gait harmonic. The reconstructed spectrum can be used for identification of various gait parameters useful for early detection of lameness in cows. These parameters include, gait speed, duration of non-gait episodes, correlation of various legs from both frequency and energy analysis of the spectrum and gait phases. Detailed gait analysis can be also performed by detection of gait phases using the reconstructed signal in time domain from the fundamental gait harmonic. This allows for better identification of peak amplitudes related to hoof contacts and therefore estimating gait symmetry level of various combination of legs.

Designed DNN framework exploits LSTM networks to estimate instantaneous frequencies. The designed platform can be used in future systems exploiting suitable machine learning techniques to detect lameness in a large cohort of diary cows. Meanwhile, identification of the fundamental gait harmonic from only exploring the spectrum given by the SSWT is challenging as there are extra, closely related and noisy harmonics which need to be moved away from the
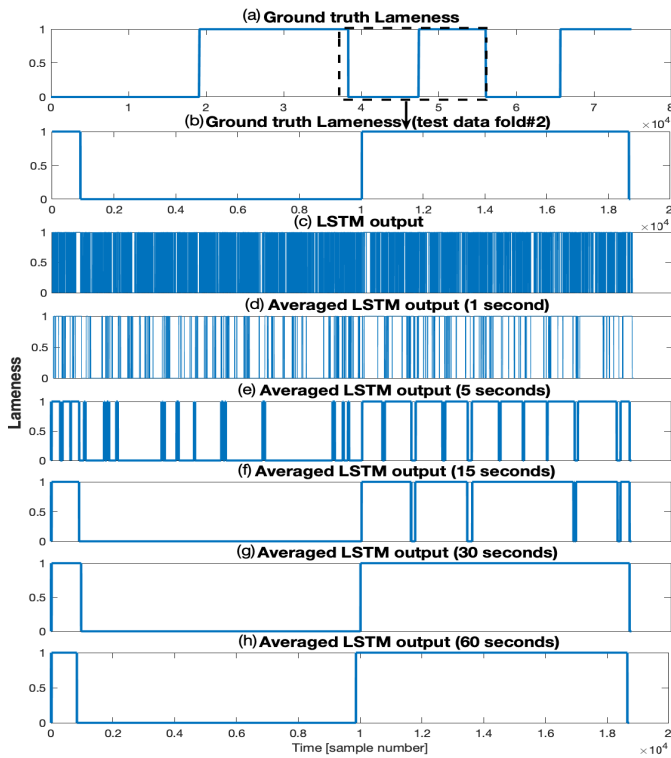
Fig. 11. (a) Ground truth scores for lameness; 0 is for non-lame and 1 is for lame. (b) the zoomed highlighted segment using dashed-rectangle in graph (a) used for fold #2. (c) Original LSTM output using DNN model in Table III. Moving average filter applied to LSTM output using (d) 1 second, (e) 5 seconds, (f) 15 seconds, (g) 30 seconds, (h) 60 second windows.
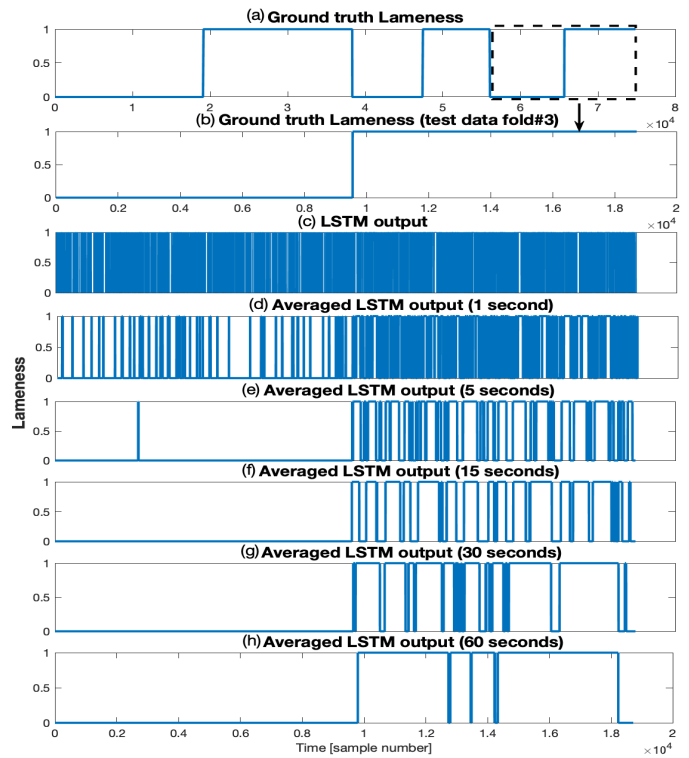
Fig. 12. (a) Ground truth scores for lameness; 0 is for non-lame and 1 is for lame. (b) the zoomed highlighted segment using dashed-rectangle in graph (a) used for fold #3. (c) Original LSTM output using DNN model in Table III. Moving average filter applied to LSTM output using (d) 1 second, (e) 5 seconds, (f) 15 seconds, (g) 30 seconds, (h) 60 second windows.

spectrum. Using the power of deep learning and by training the system using the reference frequencies, it is possible to provide a reliable estimation of instantaneous frequencies. Such system is preferred rather that the use of manual parametric setting in the frequency ridge estimation method applied to SSWT spectrum. Such parameter settings need to be adjusted for many signal segments protecting variations in the frequencies.

The proposed method in this paper can be generalized to solve many problems which require robust frequency estimation. This includes heart rate analysis from wearable sensors including optical sensors and accelerometers.

Automatic methods using machine learning have been applied in gait analysis of cows/cattle to detect lameness with varying success. Our approach for the first time identified novel insights into features useful for lameness detection showing high accuracies. Our system uses low-power consumption sensors with low sampling frequency which makes continuous monitoring more practical. Higher sampling frequencies are beneficial for detailed temporal signal analysis. However, our method exploits both temporal and spectral domain for reliable gait harmonic extraction. Such system is expected to help farmers identify the cows/cattle with lameness to improve their welfare, health and associated costs. However, having only two categories of lame and no-lame may not be sufficiently helpful. Historical observation of cow/diary cattle gait patterns can help prevention and early identification of lameness. One more advantage of our method is that, since

it exploits an LSTM network, it is able to learn long-term dependencies on cow gait behavior which can be very useful for early identification of lameness. In addition, such data can help assignment of severity of lameness which is an important aspect of cow/diary cattle gait analysis which has not been well explored yet.

ACKNOWLEDGMENT

REFERENCES

[1] S. Sanei, D. Jarchi and A. G. Constantinides, *Body Sensor Networking, Design and Algorithms*. Jonh Wiley and Sons, 2020.
[2] D. Jarchi, J. Pope, T. K. M. Lee, L. Tamjidi, A. Mirzaei and S. Sanei *A review on accelerometry-based gait analysis and emerging clinical applications*. IEEE reviews in biomedical engineering, vol. 11, pp. 177-194, 2018.
[3] 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, pp. 1–13, 2020.
[4] E. Walton, C. Casey, J. Mitsch, J. A. Vázquez-Diosdado, J. Yan, T. Dottorini, K. A. Ellis, A. Winterlich and J. Kaler, *Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour*. R. Soc. open sci., vol. 5, no. 2, pp. 1–14, 2019.

[5] J. Barwick, D. Lamb, R. Dobos, D. Schneider, M. Welch and M. Trotter, *Predicting lameness in sheep activity using tri-axial acceleration signals*. animals, vol. 8, no. 12, 2018.

[6] V. M. Thorup, L. Munksgaard, P.-E. Robert, H. W. Erhard, P. T. Thomsen and N. C. Friggens, *Lameness detection via leg-mounted accelerometers on dairy cows on four commercial farms*. Animal, vol. 9, no. 10, pp. 1704-1712, 2015.

[7] J. Haladjian, J. Haug, S. Nüske and B. Bruegge, *A wearable sensor system for lameness detection in dairy cattle*. Multimodal Technologies and Interaction, vol. 2, no. 27, pp. 1704-1712, 2018.

[8] M. Alsaaod, M. Luternauer, T. Hausegger, R. Kredel and A. Steiner, *The cow pedogram—Analysis of gait cycle variables allows the detection of lameness and foot pathologies*. J. Dairy Sci. , vol. 2, no. 100, pp. 1417–1426, 2017.

[9] N. Blackie, E. C. L. Bleach, J. R. Amory and J. R. Scaife, *Association between locomotion score and kinematic measures in dairy cows with a varying hoof lesion type*. Journal of Dairy Science, vol. 96, no. 6 pp. 3564-3572, 2013.

[10] A. Van. Nuffel, I. Zwertvaegher, S. V. Weyenberg, M. Pastell, V. M. Thorup, C. Bahr, B. Sonck and W. Saeys, *Lameness Detection in Dairy Cows: Part 2. Use of Sensors to Automatically Register Changes in Locomotion or Behavior*. animals, vol. 5, no. 3, pp. 861–885, 2015.

[11] A. V. Nuffel, I. Zwertvaegher, L. Pluym, S. V. Weyenberg, V. M. Thorup, M. Pastell, B. Sonck and W. Saeys, *Lameness detection in dairy cows: Part 1. How to distinguish between non-lame and lame cows based on differences in locomotion or behavior*. animals, vol. 5, pp. 838-860, 2015.

[12] N. W. O. Leary, D. T. Byrne, A. H. O. Connor and L. Shalloo, *Invited review: Cattle lameness detection with accelerometers*. Journal of Diary Science, vol. 103, no. 5, pp. 3895-3911, 2020.

[13] S. Benaissa, F. A. M. Tuyttens, D. Plets, T. de. Pessemier, J. Trogh, E. Tanghe, L. Martens, L. Vandaele, A. V. Nuffel, W. Joseph and B. Sonck, *On the use of on-cow accelerometers for the classification of behaviours in dairy barns*. Res Vet Sci , vol. 125, pp. 425-433, 2019.

[14] Z. E. Barker, J. A. V. Diosdado, E. A. Codling, N. J. Bell, H. R. Hodges, D. P. Croft and J. R. Amory, *Use of novel sensors combining local positioning and acceleration to measure feeding behavior differences associated with lameness in dairy cattle*. Journal of Dairy Science, vol. 101, pp. 6310–6321, 2018.

[15] M. Alsaaod, M. Fadul and A. Steiner, *Automatic lameness detection in cattle*. The Veterinary Journal, vol. 246, pp. 35-44, 2019.

[16] I. Daubechies, J. Lu and H. Wu, *Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool*. Applied and Computational Harmonic Analysis, vol. 30, pp. 243–261, 2011.

[17] R. A. Carmona, W. L. Wang and B. Torresani, *Characterization of signals by the ridges of their wavelet transforms*. Applied and Computational Harmonic Analysis, vol. 45, no. 10, pp. 2586–2590, 1997.

[18] D. Jarchi, J. Andreu-Perez, M. Kiani, O. Vysata, J. Kuchynka, A. Prochazka and S. Sanei, *Recognition of patient groups with sleep related disorders using bio-signal processing and deep learning*. Sensors, vol. 20, no. 9, pp. 2594, 2020.

[19] D. Jarchi, S. Sanei and A. Prochazka, *Detection of sleep apnea/hypopnea events using synchrosqueezed wavelet transform*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1199-1203, 2020.

[20] G. Thakur, E. Brevdo, N. S. Fuckar and H. T. Wu, *The Synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications*. Signal Processing, vol. 93, no. 5, pp. 1097–1094, 2011.

[21] S. Hochreiter and J. Schmidhuber, *Long short-term memory*. Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] A. Gers, J. Schmidhuber and F. Cummins, *Learning to forget: Continual prediction with LSTM*. Neural Computation, vol. 12, no. 10, pp. 2451-2471, 2000.

[23] Y. Bengio, P. Simard and P. Frasconi, *Learning long-term dependencies with gradient descent is difficult*. IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 157-166, 1994.

[24] J. Li, A. Mohamed, G. Zweig and Y. Gong, *LSTM time and frequency recurrence for automatic speech recognition* . IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2015.

[25] Y. Liu and D. Wang, *Time and frequency domain long short-term memory for noise robust pitch tracking* . IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.

[26] J. Li, A. Mohamed, G. Zweig and Y. Gong, *Exploring multidimensional lstms for large vocabulary ASR*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4940-4944, 2016.

[27] R. Li, Z. Wu, Y. Ning, L. Sun, H. Meng and L. Cai, *Spectro-Temporal Modelling with Time-Frequency LSTM and Structured Output Layer for Voice Conversion*. Interspeech, pp. 3409-3413, 2017.

[28] P. Y. Lim, N. Huxley, A. Willshirec, M. J. Green, R. Othman and J. Kaler, *Unravelling the temporal association between lameness and body condition score in dairy cattle using a multistate modelling approach*. Preventive Veterinary Medicine, vol. 118, no. 4, pp. 370-377, 2015.