Computers and Electronics in Agriculture 162 (2019) 531–542

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Livestock vocalisation classification in farm soundscapes

James C. Bishop[a,b,*], Greg Falzon[a,b], Mark Trotter[c], Paul Kwan[b], Paul D. Meek[d,e]

[a] University of New England – Precision Agriculture Research Group (PARG), Armidale, NSW, Australia
[b] University of New England, School of Science and Technology, Armidale, NSW, Australia
[c] Institute for Future Farming Systems, CQUniversity, Rockhampton, QLD, Australia
[d] NSW Department of Primary Industries, PO Box 530, Coffs Harbour, NSW, Australia
[e] University of New England, School of Environmental and Rural Science, Armidale, NSW, Australia

ARTICLE INFO

Keywords:
Precision livestock farming
Animal welfare
Vocalisation detection
Wavelets
Mel-frequency Cepstral coefficients
Machine learning
Support vector machines

ABSTRACT

Livestock vocalisations have been shown to contain information related to animal welfare and behaviour. Automated sound detection has the potential to facilitate a continuous acoustic monitoring system, for use in a range Precision Livestock Farming (PLF) applications. There are few examples of automated livestock vocalisation classification algorithms, and we have found none capable of being easily adapted and applied to different species' vocalisations. In this work, a multi-purpose livestock vocalisation classification algorithm is presented, utilising audio-specific feature extraction techniques, and machine learning models. To test the multi-purpose nature of the algorithm, three separate data sets were created targeting livestock-related vocalisations, namely sheep, cattle, and Maremma sheepdogs. Audio data was extracted from continuous recordings conducted on-site at three different operational farming enterprises, reflecting the conditions of real deployment. A comparison of Mel-Frequency Cepstral Coefficients (MFCCs) and Discrete Wavelet Transform-based (DWT) features was conducted. Classification was determined using a Support Vector Machine (SVM) model. High accuracy was achieved for all data sets (sheep: 99.29%, cattle: 95.78%, dogs: 99.67%). Classification performance alone was insufficient to determine the most suitable feature extraction method for each data set. Computational timing results revealed the DWT-based features to be markedly faster to produce (14.81 – 15.38% decrease in execution time). The results indicate the development of a highly accurate livestock vocalisation classification algorithm, which forms the foundation for an automated livestock vocalisation detection system.

1. Introduction

To meet the increasing global demand for livestock products, livestock management practices have shifted towards intensive methods (FAO, 2009). Although output has dramatically increased (Thornton, 2010), it has become more difficult for farmers to observe and monitor individual animals. In parallel, consumers are demanding more transparency in the welfare, environmental impact, and safety reporting of the animal products they purchase (Thornton, 2010; Moynagh, 2000; Grandin, 2014). This poses a dilemma for the modern producer who needs to find a balance between high production efficiency targets, and ethical, sustainability, and safety requirements. Precision Livestock Farming (PLF) aims to address these concerns by facilitating the continuous, automated monitoring of livestock, and enabling more appropriate and timely interventions. In PLF, the biological responses of livestock are constantly measured, providing up-to-date information on 'states' of interest (e.g. welfare indicators, production targets, etc.). These systems can be used to alert farmers to critical events on farm, to aid management decisions, and to augment their existing knowledge.

In practice, PLF systems require state-of-the-art software and hardware systems. To obtain continuous information concerning animal behaviour, various sensors can be used, such as microphones (Berckmans, 2014; Chung, 2013; Exadaktylos et al., 2014), cameras (Sadgrove, 2017), accelerometers (Alvarengaa, 2016; Barwick, 2018; Barwick, 2018), and Global Positioning System (GPS) satellites (Falzon, 2013). The resulting data streams are analysed to discover discriminatory features pertaining to the target behaviour and are subsequently used to train machine learning (ML) models. Well defined models are capable of automatically predicting the state, condition, or trait of interest (Berckmans, 2014). An emerging area of PLF research is concerned with the automated detection and classification of acoustic events, and how audio signals can be used as an input for PLF systems.

Acoustic monitoring provides an accurate and non-invasive way to measure the biological responses, and by extension, welfare states of

* Corresponding author at: University of New England – Precision Agriculture Research Group (PARG), Armidale, NSW, Australia.
E-mail address: jbisho23@myune.edu.au (J.C. Bishop).

https://doi.org/10.1016/j.compag.2019.04.020
Received 31 August 2018; Received in revised form 7 April 2019; Accepted 15 April 2019

livestock (Exadaktylos et al., 2014). Coughs produced by pigs have been shown to contain information related to respiratory disease (Ferrari, 2008; Van Hirtum and Berckmans, 2004). Based on these findings, acoustic monitoring has been used extensively to detect and diagnosis disease in indoor piggeries (Chung, 2013; Jans, 2005; Guarino, 2008; Chedad, 2001), including the localisation of diseased individuals (Exadaktylos et al., 2011). Similar work has been conducted with calves (Vandermeulen, 2016). Feed intake estimation using acoustic analysis has been widely researched, such as with broiler chickens (Aydin et al., 2015), including determining short-term feeding behaviour (Aydin and Berckmans, 2016). In cattle, automated detection and classification of bite and chew activity has been used as means of feed intake estimation (Chelotti, 2016; Clapham, 2011; Milone, 2012; Galli, 2006; Ungar and Rutter, 2006), also providing insight into grazing behaviour (Andriamasinoro, 2016). Similar automatic recognition of jaw movements in free-ranging cattle, goats and sheep has also been demonstrated (Navon, 2013). Automated segmentation and classification of ingestive sounds in sheep has also been achieved (Milone, 2009). Cattle vocalisation analysis has been used to determine estrus (Chung, 2013; Lee, 2014), and as a means of welfare assessment, using "murmuring" during resting and ruminating behaviour as an indicator of "good welfare" (Meen, 2015). Vocalisation-based welfare monitoring has also been demonstrated with chickens (Pereira et al., 2011), including predicting growth (Fontana, et al., 2014). The majority of PLF acoustic monitoring research has focused on algorithms highly optimised for specific vocalisation detection, usually in an indoor production environment. There is an absence of a general purpose, noise-robust vocalisation detection algorithm, which can be readily retrained and adapted to identify different livestock classes.

Numerous audio-specific features have been proposed for use in PLF vocalisation detection and classification. These include mean maximum frequency (Meen, 2015), relative sound intensity (Moura, 2008), power spectra (Chedad, 2001), peak frequency (Fontana, et al., 2014), formant-based (Lee, 2014), energy envelope (Exadaktylos et al., 2011; Aydin et al., 2015), and Mel-Frequency Cepstral Coefficients (MFCCs) (Chung, 2013; Chung, 2013; Bishop et al., 2017). In particular, MFCCs have shown considerable success in a diverse array of applications (Tiwari, 2010; Sharan and Moir, 2017; Ahmad, 2015), making them a good starting point for audio-specific feature extraction. The use of a Fast Fourier Transform (FFT) is central to many audio feature extraction techniques, including MFCCs (Davis and Mermelstein, 1980). As an alternative to the FFT, the Discrete Wavelet Transform (DWT) has been applied to acoustic detection tasks (Abdalla et al., 2013; Rabaoui, 2008; Ramalingam and Dhanalakshmi, 2014). DWTs have the advantage of retaining temporal information, in addition to frequency (Olkkonen and Wavelet, 2011), and have been shown to be noise-robust (Virtanen et al., 2012). The use of DWTs in PLF audio applications has not been widely demonstrated (Banakar et al., 2016).

ML models capable of successfully classifying livestock vocalisations are a key component in developing a livestock vocalisation detection algorithm. Artificial Neural Networks (ANNs) (Chedad, 2001), Support Vector Data Descriptions (SVDDs) (Chung, 2013a; b), AdaBoost.M1 (Lee, 2014), and Support Vector Machines (SVMs) (Bishop et al., 2017; Banakar et al., 2016) have all shown promising results. SVMs excel at binary classification problems (Bishop et al., 2017), and the combination of DWT-based features with an SVM model has not been widely explored in the PLF space (Banakar et al., 2016; Deng, 2010). This is even more pronounced when surveying multi-purpose vocalisation detection algorithms, capable of being adapted to different target the vocalisations of different livestock species.

This work presents a multi-purpose farm animal vocalisation classification algorithm, which forms the foundation of a future automated livestock vocalisation detection system (Fig. 1). Classification is determined using a binary C-SVM classification model (Vapnik et al., 1997). A comparison of MFCC and DWT-based features was conducted, taking into account both classification and timing performance. The latter is especially important for end-deployment, where computational resources may be considerably constrained. The algorithm's performance is tested in three separate vocalisation detection scenarios (sheep, cattle, and guardian dogs), using data collected from audio sensor nodes in operational outdoor farm environments at different locations. Classification performance results demonstrate the development of a highly accurate vocalisation detection algorithm (95.78 – 99.67% accuracy), capable of being easily adapted to different livestock species and deployments.

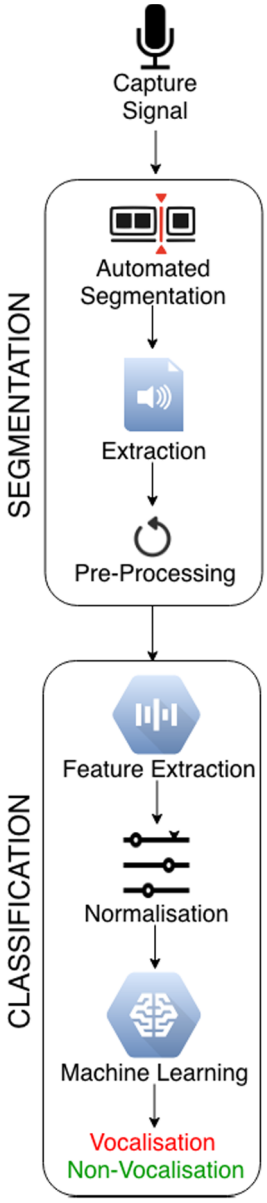

Fig. 1. An overview of the proposed multi-purpose livestock vocalisation detection algorithm. This work focuses on the development of the classification component.

2. Materials and methods

2.1. Data

2.1.1. Data collection

Continuous field recordings were conducted at three different Australian livestock enterprises. Each deployment targeted a specific animal vocalisation: sheep (*Ovis aries)*, cattle (*Bos taurus*), and

Fig. 2. Wildlife Acoustics Song Meter SM3 audio recording unit in the field.

Maremma sheepdogs *(Canis familiaris)*. Sheep and cattle were chosen due to Australia's large production of these animals (AWI, Australian Wool Production Forecast Report., 2017; MLA, Australian Cattle Industry Projections, 2018), while Maremma sheepdogs were selected due to their wide use as livestock guardians, particularly against wild dog predation (Smith, 2000). Wildlife Acoustics Song Meter SM3 recording units (Wildlife Acoustics, 2014) (Fig. 2) were used for sheep and cattle data collection, and a Wildlife Acoustics Song Meter SM2 unit (Wildlife Acoustics, 2010) was used for the Maremma sheepdogs deployment. Recording units were placed in static, outdoor locations, and animals were free to move around following natural behaviours: the distance between the recording units and livestock was not actively controlled. The recording units were subjected to outdoor weather conditions, in some cases including rain and high winds. The acoustic environment on-site was not controlled, in that regular livestock management practices and routines continued to be carried out. As it was unknown how far target animals would be from the recording source, the input level (i.e. gain) was set to 'auto' on each unit (24 dB) (Wildlife Acoustics, 2014). All these factors resulted in a large variation in amplitude-based metrics between and within data sets (see Section 2.1.6).

2.1.2. Data extraction

Using Audacity audio editing software (Audacity, 2017), each data set (Tables 1, 3, and 5) was segmented into 1 h intervals, and converted to the spectral domain using a Fast Fourier Transform (FFT). The FFT used a window size of 1024 and applied a Hanning window function, which was selected for its balanced frequency resolution, side lobe roll-off rate, and side lobe level reduction (Gaberson, 2006). From the data produced by the FFT, spectrogram images were produced, and visually inspected to ascertain the overall level of vocalisations, using Sonic Visualiser software (Cannam et al., 2010). It was found that vast amounts of the recordings contained very little activity. The time periods identified by the operator as containing the highest vocalisation density were selected for individual sound extraction. All data instances were extracted in single-channel (mono), 16-bit/44.1 KHz quality, in WAVE format; this differs from the original capture quality. Based on the observed average duration of vocalisations of each target animal, a 1 s extraction window was used.

To simulate automated segmentation, all data sets contained both positive and negative instances taken from the same time period. The positive class was comprised of vocalisations by the target animal. All vocalisations within an identified high activity period were extracted, with occurrences determined both aurally and visually (i.e. spectrogram) by the operator. All encountered vocalisations were accepted, regardless of origin (e.g. age or sex of the animal), the number of individuals simultaneously vocalising, or the type of vocalisation. Amplitude was not considered when assessing a vocalisation for extraction; all vocalisations which could be aurally identified as the target species were included. This raises a limitation of the study: only human-identifiable vocalisations were included, and these were only ascertained through aural and spectrogram inspection. Direct observation of behaviour would have provided more certainty and evidence of what sounds were occurring, and their source. The negative class for each data set was composed of the three most frequently occurring non-vocalisation sound types, as determined by the operator. Examples of each of the negative subclasses were taken from throughout the time period used, with the number of extractions based on the observed prevalence of each sound. As with the positive class, overall amplitude was not taken into account when extracting each instance. As each negative subclass can contain a broad range of individual sounds, the frequency distribution of each subclass is not uniform. The "bird" subclass may include many different species, and types of bird calls. The "noise" subclass can contain any non-descript audio occurrence, such as human activity (e.g. speaking, working, etc.), vehicles (e.g. cars, trucks, aeroplanes), and other farm-related sounds. (Figs. 3–5) show examples of each class / subclass: further examples are given in (Supplementary 6.1). In many cases, individual negative instances contained a significant amount of spectral overlap with target vocalisations (Figs. 3–5).

2.1.3. Sheep

Tables 1 and 2

2.1.4. Cattle

Tables 3 and 4

2.1.5. Dogs

Tables 5 and 6

2.1.6. Audio analysis

Following the creation of each data set, audio analysis was conducted in MATLAB 2017a software (The MathWorks, 2017) to ascertain the average level and standard deviation of amplitude, noise, and clipping in each class and subclass. Decibels relative to full-scale (dBFS) is a unit of measurement for amplitude in digital audio systems, based on the root mean square (RMS) value of the signal (Table 7). Mean dBFS was used to define signal power. The signal-to-noise (SNR) ratio was determined by taking 5 samples of background noise from each time period used, calculating the SNR as shown in (Table 7), using MATLAB's 'snr()' function, and taking the mean. Clipping was defined as any 3 consecutive values that contained the maximum value allowed by the bit-rate (i.e. the highest signed 16-bit value is 32,767). Clip rate (CR) was expressed as the number of clips divided by the number of samples (i.e. 44,100 at extraction sample rate) and derived by the formula in (Table 7).

The results of the analysis are given in (Table 8). Overall, there was a lot of variation in dBFS, SNR, and CR both between and within each data set. Standard deviation for all metrics is large, illustrating the diversity in amplitude of each class/subclass. In most cases, a higher dBFS and CR was associated with a more desirable SNR, as extracted samples are higher in amplitude. In the sheep data set, many animals were in close proximity to the capture device, resulting in a marked difference in dBFS, CR and SNR between the positive and negative subclasses: the difference is less pronounced in the wind subclass. The cattle data set contained high levels of wind, causing comparatively higher CRs. The site where this data was collected was also near a major highway, under a flightpath, and was in constant use by farmers and workers. The gain settings on the recording device appeared to be too high for the level of acoustic interference encountered. The dog data set had the lowest average dBFS and CR, but the least variation between classes/

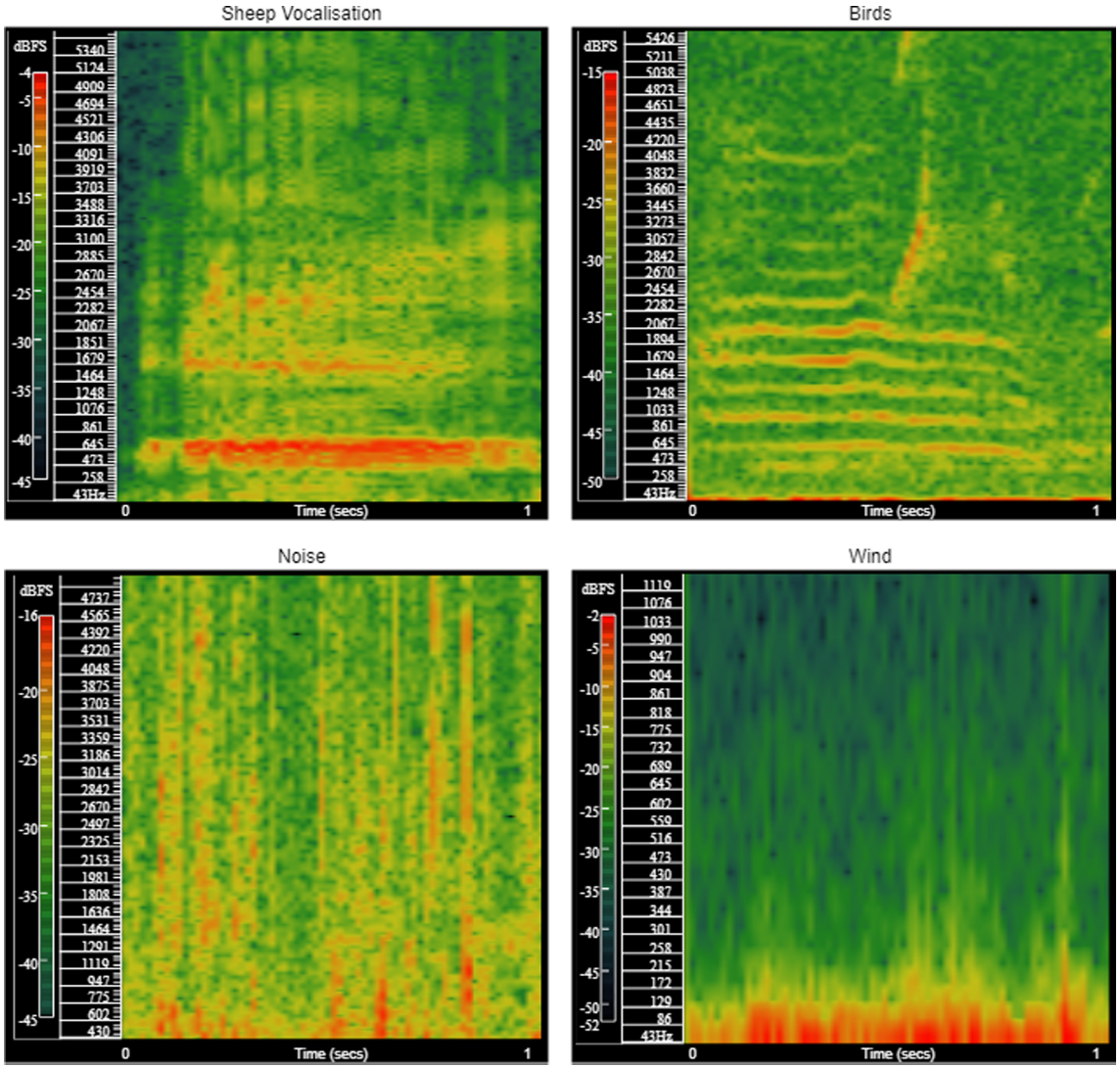

Fig. 3. Spectrogram examples of each class and subclass for the sheep data set.

subclasses. The site where dog data was collected had high levels of insect activity (cicadas), and all sound sources were distant from the capture device. Average dBFS could have been improved by increasing the capture device's gain, to a level more suitable to the low amplitude sounds encountered. The SNR for the dog class shows a less noisy signal than its corresponding negative subclasses. When deploying recording devices in the future, some of the variation in dBFS, SNR, and CR can be reduced by setting a more appropriate gain level for the acoustic environment. The gain should be set based on analysis of test recordings, and ideally, checked and adjusted during the duration of the deployment.

2.1.7. Automated segmentation

A fully automated vocalisation detection system must be capable of not only classifying sounds, but also segmenting them from the audio stream (Fig. 1). The algorithm presented is the classification component of a proposed system, focusing on using audio-specific feature extraction and machine learning to classify livestock vocalisations. To test the classification accuracy of the algorithm, segmentation was performed manually: this is a limitation of the current design. Future work will focus on the development of an automated segmentation component.

The goal of automated audio segmentation (AAS) is to find the boundaries of homogeneous acoustic content in an audio stream (Castán et al., 2015). By ascertaining the points where the signal significantly changes (e.g. temporally or spectrally), individual sounds (Huang et al., 2013) or portions of audio derived from the same source (Bhandari et al., 2013), can be identified. A basic livestock vocalisation AAS component would need to achieve two main goals: identify when possible sounds of interest occur, and segment individual sounds that exist close together. Energy-based methods are typically employed to determine initial segmentation boundaries, by applying a threshold function to audio windows based on a metric, such as RMS or zero-crossing rate (Panagiotakis and Tziritas, 2005). This may result in large section of segmented audio that contain many different sounds in close proximity. Metric-based methods can be used to locate the acoustic change points in the signal, the points of most dissimilarity or 'distance' (Kemp et al., 2000), most commonly by employing a distance function to evaluate the correlation between adjacent windows (Rybach et al., 2009). Demonstrated distance functions include Gish distance (Kemp et al., 2000), Kullback-Leibler divergence (Virtanen and Helen, 2007), and Baynesian Information Criterion (BIC) (Ozan et al., 2014).

2.2. Feature extraction

2.2.1. Mel-Frequency Cepstral Coefficients (MFCCs)

The Mel-Scale is designed to mimic human auditory response, being

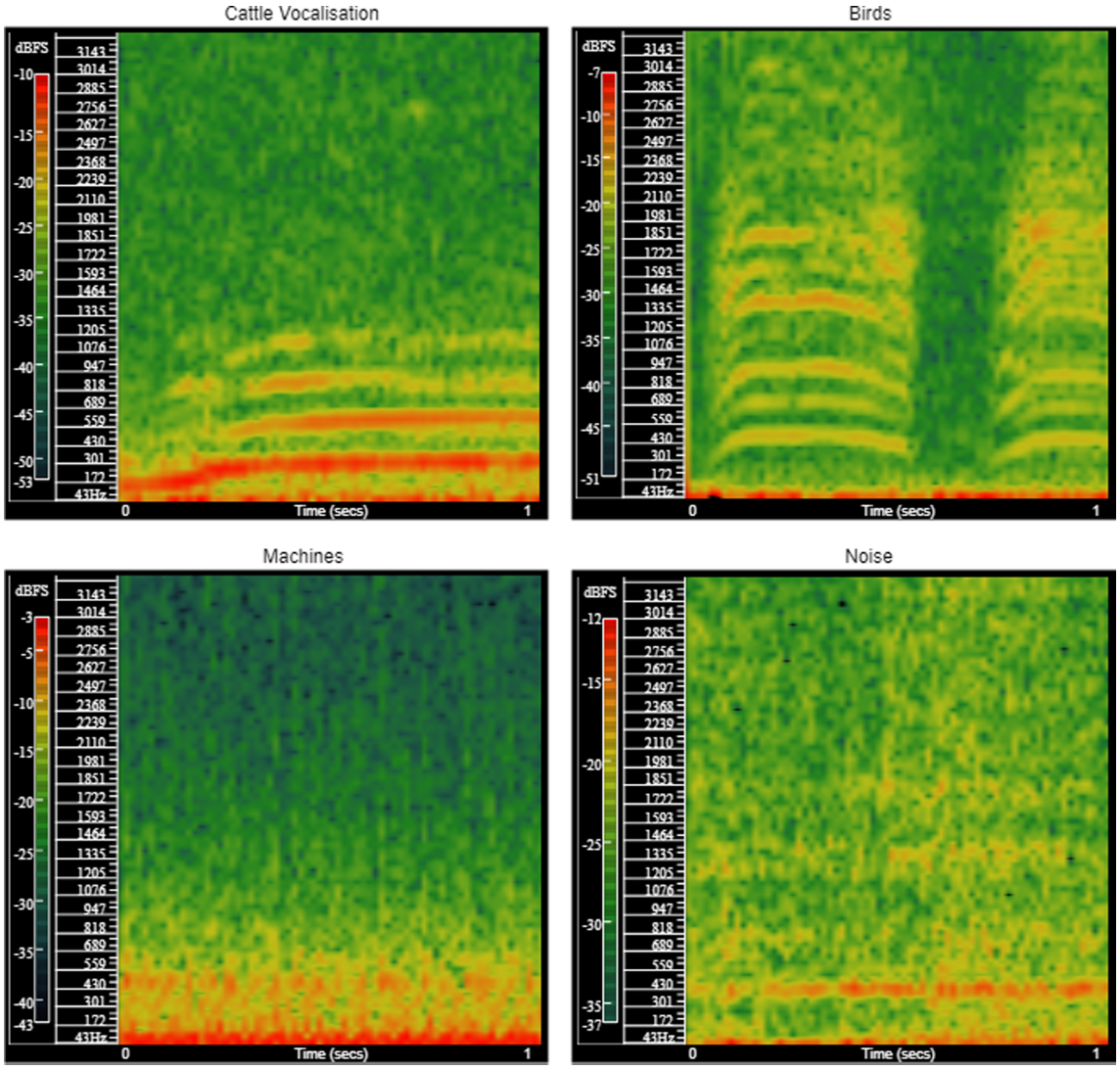

Fig. 4. Spectrogram examples of each class and subclass for the cattle data set.

based on the relationship between actual and perceived frequency (Patel, 2013). A cepstrum provides information about spectral distribution, envelope, and changes between bands (Sharan and Moir, 2016). This combination makes MFCCs an effective acoustic classification feature. The steps required to extract MFCCs from a digital audio signal are shown in (Fig. 6), and outlined in detail in (Sharan and Moir, 2015). MATLAB 2017a was used to extract MFCCs from each instance. The parameters used for each data set are given in (Table 9). The low and high cut-off values were found by conducting a grid search, utilising stratified 10-fold cross-validation (10CV) (see Section 2.3.2), with accuracy as the performance metric. The search bounds were determined by visual spectrogram inspection of typical vocalisation ranges (Figs. 3–5). The value of pre-emphasis alpha (Young et al., 2002; Solera-Urena, 2007), number of channels (Sharan and Moir, 2016), liftering type (Paliwal, 1999), and liftering value (Young et al., 2002) were all selected based on review. Using a 1 s window length, and omitting the 0th coefficient (Zheng et al., 2001), the final feature vector has a dimensionality of 12.

2.2.2. Discrete Wavelet Transform (DWT)

The DWT (Daubechies, 1992) provides information about both the spectral and temporal content of a signal (Olkkonen and Wavelet, 2011). The DWT operates by decomposing a discrete signal into two coefficient sub-bands, detail and approximation, using a particular wavelet type. The Daubechies wavelets (Daubechies, 1992) are commonly used wavelets in acoustic detection (Olkkonen and Wavelet, 2011). The DWT can be performed at different decomposition levels by successively applying the transform to the approximation sub-band produced at each level, using a cascading filter bank scheme (Mallat, 1989) (Fig. 7).

MATLAB 2017a was used to implement a Daubechies DWT. Wavelet type and decomposition level were selected using a parameter grid search, and performing stratified 10CV on the training set, with accuracy as the criterion. The search bounds and selected values are shown in (Table 10). To reduce the dimensionality of the final feature vector, the features shown in (Table 11) were extracted from each detail coefficient sub-band. The last approximation sub-band was omitted, as it was not found to improve results. The final feature vector had a dimensionality of 24.

2.2.3. Timing

It was deemed important to evaluate the timing performance of each feature extraction technique, as it was highly likely that the end deployment would benefit from any reduction in computational time, as this translates to less required resources, and faster operation. Feature extraction timing experiments were conducted in MATLAB

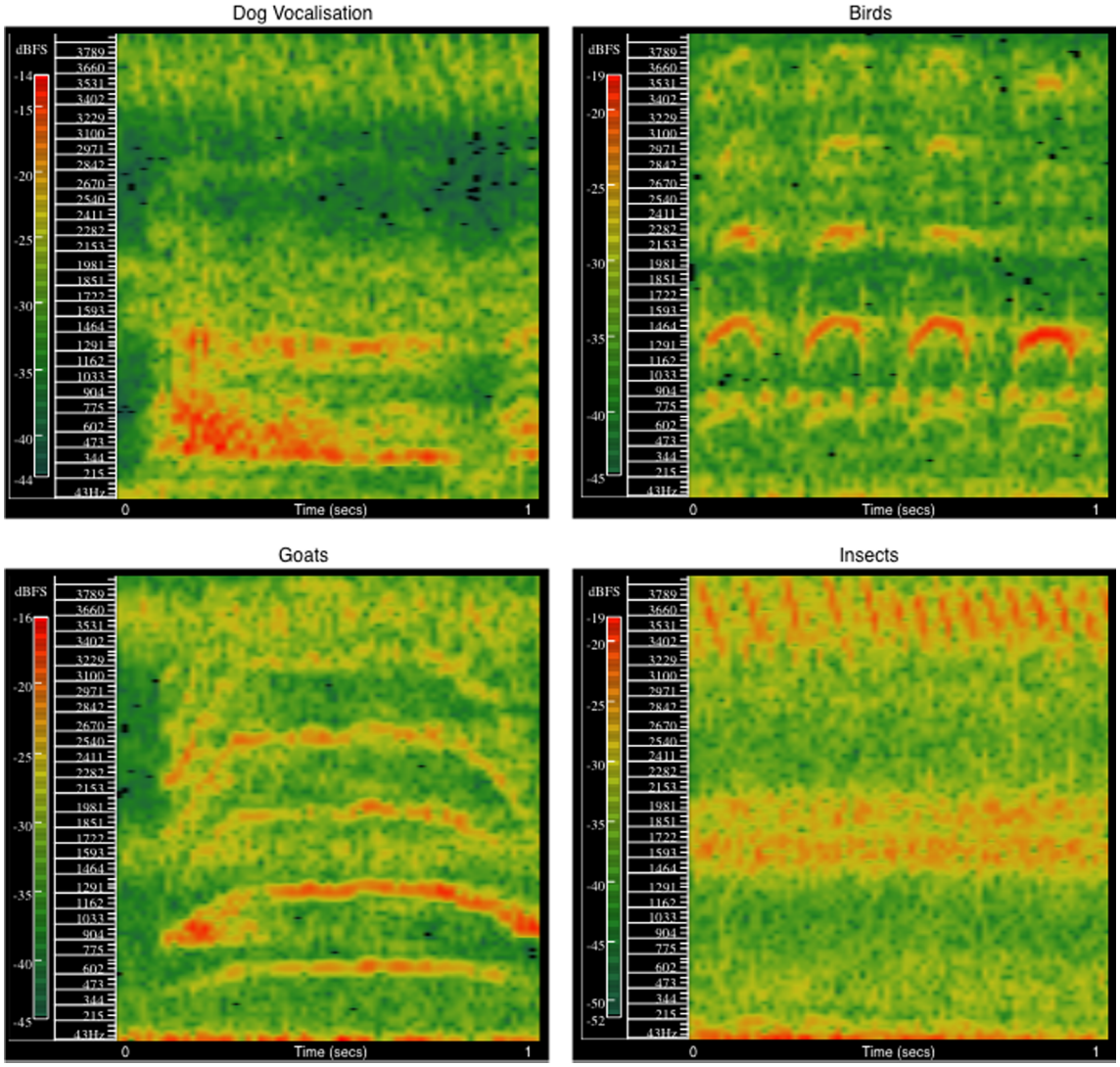

Fig. 5. Spectrogram examples of each class and subclass for the dog data set.

Table 1
Sheep data acquisition information.

Location:	Northern New South Wales, Australia
Enterprise:	Wool production
Capture Quality:	16-bit/16KHz
Total Hours:	720
Hours Processed	6

Table 2
Sheep vocalisation data set.

Positive	Negative			Total	
Sheep	Birds	Noise	Wind	Train	Test
1681	579	663	439	3026	336

2017a (The MathWorks, 2017) using the 'timeit()' function. This function calls a target function multiple times and returns the median of the timing results. The positive sheep vocalisation class was used during testing, with 'timeit()' called once for each instance, resulting in 1681 measurements. The system used was a mid-2015 MacBook Pro (Apple. MacBook Pro, 2015), with a 2.2 GHz Intel Core i7 processor, 16 GB of 1600 MHz DDR3 RAM, and a solid state hard drive.

Table 3
Cattle data acquisition information.

Location	Northern New South Wales, Australia
Enterprise	Beef production
Capture Quality	32-bit/24KHz
Total Hours	144
Hours Processed	5

Table 4
Cattle vocalisation data set.

Positive	Negative			Total	
Cattle	Birds	Machines	Noise	Train	Test
1020	340	340	340	1836	204

Table 5
Dog data acquisition information.

Location	North Coast New South Wales, Australia
Enterprise	Goat meat production
Capture Quality	32-bit/48KHz
Total Hours	284
Hours Processed	2

Table 6
Dog vocalisation data set.

Positive	Negative			Total	
Dogs	Birds	Goats	Insects	Train	Test
1200	538	147	515	2160	240

Table 7
Formulas for dBFS, signal-to-noise ratio (SNR), and clip rate (CR) for a given signal (X).

Metric	Definition
dBFS:	$dBFS_X = 20log_{10}\left(\sqrt{\frac{1}{N}\sum_{i=1}^{n} x_i^2}\right)(\sqrt{2})$
Signal-to-noise ratio:	$SNR_X = 10log_{10}\left(\frac{P_{X\,signal}}{P_{X\,noise}}\right)$ where P = power of signal
Clip rate:	$CR_X = 100\left(\frac{clips}{totalsamples}\right)$ where clip = 3 successive max values for bit-rate (e.g. 32,767 for 16-bit)

2.3. Support vector Machines (SVMs)

SVMs can be defined as supervised, non-probabilistic, binary classifiers (James, 2014). SVMs operate by creating a separating hyperplane, a decision boundary that aims to maximise the division between classes (Bishop, 2006). The position of the hyperplane is defined by its margins, delineated by the closest instances to the boundary: the support vectors (James, 2014). A 'slack' variable (C) is present in SVMs, allowing the width of the margins to be adjusted, which provides some control over the bias-variance trade-off of the model (Hsu et al., 2010). Kernel functions are commonly employed to allow for non-linear decision boundaries, whilst reducing the higher computational complexity associated with enlarging the feature space (Vapnik et al., 1997). A Radial Basis Function (RBF) kernel was used during experimentation, due to its purported balanced performance (Hsu et al., 2010). The RBF kernel has an adjustable gamma (γ) value, allowing greater flexibility when tuning the model (Hsu et al., 2010). To implement the SVM portion of the algorithm, software was developed in C++ utilising the LIBSVM library (Chang and Lin, 2011).

2.3.1. Training and testing

To rigorously test an ML model's classification performance, separate training and testing data must be obtained (James, 2014). (Fig. 8) provides an overview of the process used to test classification performance for each data set. Due to the stochastic nature of the data sets, stratified 10CV was used. A stratified approach was used to ensure that the distribution of both the positive and negative classes remained equal when splitting each fold of the data. As each negative subclass may differ in size, a consistent ratio of each negative subclass is present in both the training and testing portion. Prior to splitting, all data sets were shuffled using a predefined input seed, facilitating reproducibility. For each fold, the data set was split into a training and testing set, with 90% and 10% of the total instances respectively. The test set for each fold of data is only used to test the performance of the model. The aggregate results are presented in (Section 3), and individual fold results are shown in (Supplementary 6.3).

The SVM parameters C and γ were found by conducting a parameter grid search on the training data portion of each fold (Chang and Lin, 2011). For each increment of log_2(C) and log_2(γ), 10-fold cross-validation (10CV) was performed on the training set, with the resulting accuracy used to rank performance. The results for all folds are presented in (Supplementary 6.2).

The Clopper-Pearson method (Clopper and Pearson, 1934) was used to calculate the 99% binomial proportional confidence intervals (CIs) for each experiment. Using this method, a range of values is obtained (i.e. the Maximum Likelihood estimate of proportion or accuracy, and an upper and lower bounds) for which, in 99% of estimated intervals, the true proportion (accuracy) lies within this range. The Clopper-Pearson method was also selected due to its suitability in the estimation of intervals when the proportions are close to either 0 or 1. This facilitates a more detailed comparison of feature extraction methods within each data set, rather than only using average performance metrics. By comparing the CIs of each method, within each data set, the statistical significance of the difference between results can be ascertained.

2.3.2. Normalisation

It has been demonstrated by (Hsu et al., 2010) that SVM performance can be improved by applying min-max normalisation to training and testing data as a pre-processing step. Common scaling bounds include (FAO, 2009) or [−1 – 1] (Hsu et al., 2010). Min-max normalisation was applied firstly to the training data, with each feature being normalised independently: the minimum and maximum values were subsequently stored to be used for scaling testing data. Min-max normalisation was applied during all experiments, using [−1, 1] scaling bounds.

3. Results

All results are derived from performing stratified 10CV on each data set, using both MFCC and DWT-based features. Aggregate performance metrics for each data set are given in (Table 12). Results for each individual fold are shown in (Supplementary 6.3). To further compare feature extraction methods within each data set, binomial proportion confidence intervals (99%) were derived for classification accuracy, using the Clopper-Pearson method: these are displayed in (Fig. 9). The

Table 8
dBFS, signal-to-noise (SNR), and clip rate (CR) means, with standard deviation (STD) for each class and subclass. SNR is expressed as decibels (dB), and clip rate as clips per second (cl/s).

		dBFS	dBFS STD	SNR	SNR STD	CR	CR STD
Sheep	Sheep	−2.41 dBFS	7.67 dBFS	25.32 dB	17.70 dB	0.462	0.760
	Birds	−20.89 dBFS	1.31 dBFS	−30.03 dB	2.35 dB	0	0
	Noise	−19.64 dBFS	3.20 dBFS	−6.86 dB	6.31 dB	0	0
	Wind	−12.87 dBFS	5.19 dBFS	18.04 dB	14.31 dB	0.034	0.168
Cattle	Cattle	−6.10 dBFS	6.85 dBFS	12.97 dB	16.71 dB	0.725	2.472
	Birds	−7.50 dBFS	6.63 dBFS	9.676 dB	16.43 dB	0.163	0.710
	Machines	−7.87 dBFS	6.19 dBFS	8.55 dB	15.84 dB	0.066	0.378
	Noise	−3.30 dBFS	8.55 dBFS	17.21 dB	18.64 dB	1.352	2.349
Dogs	Dogs	−25.46 dBFS	3.74 dBFS	10.91 dB	11.45 dB	0	0
	Birds	−29.86 dBFS	5.50 dBFS	−0.98 dB	14.81 dB	0	0
	Goats	−28.67 dBFS	3.87 dBFS	−10.55 dB	11.76 dB	0	0
	Insects	−29.40 dBFS	4.13 dBFS	−7.41 dB	12.32 dB	0	0

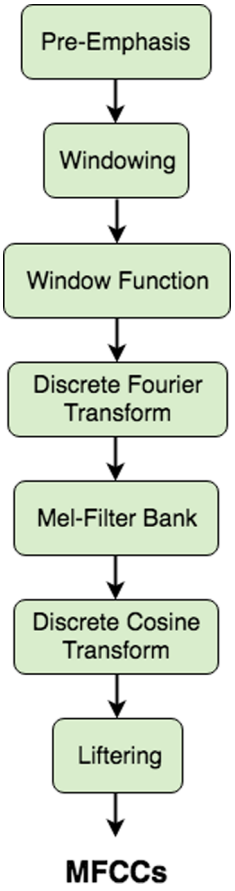

Fig. 6. The steps required to extract MFCCs from a digital audio signal.

Table 9
MFCC parameters for each data set.

	Sheep	Cattle	Dogs
Pre-Emphasis Alpha:	0.97		
Window Length:	1 s		
Window Function:	Hanning		
Filter Channels:	20		
Liftering Type:	Sinusoidal		
Liftering Value:	22		
Low Cut-Off:	300 Hz	0 Hz	250 Hz
High Cut-Off:	5000 Hz	1000 Hz	2500 Hz

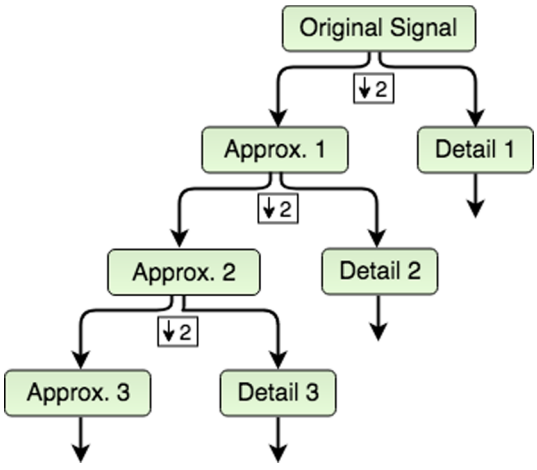

Fig. 7. Cascading filter bank scheme, where ↓2 represents a down-sampling of 2.

Table 10
DWT parameter search bounds, and selected values.

	Search Bounds		
Parameter	Lower	Upper	Selected
Wavelet type:	db2	db14	**db8**
Decomposition level:	1	8	**6**

Table 11
Features extracted from each of the DWT detail coefficient (x) sub-bands.

Feature	Definition
Non-normalised Shannon entropy (Safty and El-Zonkoly, 2008):	$E_{(x)} = -\sum_{i=1}^{n} x_i^2 \log x_i^2$
Log energy (Rabaoui, 2008):	$E_{(x)} = \sum_{i=1}^{n} x_i^2$
Mean (Tzanetakis et al., 2002):	$\bar{x} = \frac{1}{n}(\sum_{i=1}^{n} x_i)$
Standard deviation (Tyagi and Panigrahi, 2017):	$\sigma = \sqrt{\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}\right)}$

computational timing of each feature extraction method was also considered, with results provided in (Fig. 10).

4. Discussion

As can be seen from Section 3, the proposed algorithm produced excellent results across all performance metrics (Table 12), for all data sets, indicating that it is both accurate and flexible. High accuracy was observed for both the MFCC and DWT-based feature extraction methods (Table 12). Precision results were comparable to TNR in all data sets (Table 12), demonstrating that the algorithm has a balanced response to both vocalisation and non-vocalisation acoustic events. The relatively high levels of noise present in each data set (Table 8) did not reduce the performance of the algorithm, demonstrating that it is noise-robust. It is difficult to compare the obtained results to others in the field, as there are few examples of automated sheep (Bishop et al., 2017), cattle (Chung, 2013), or dog vocalisation (Yeo et al., 2012) detection, and no examples of an algorithm that can be easily applied and retrained to multiple vocalisation types. It is therefore necessary to compare performance results obtained in other livestock-related acoustic recognition tasks. This includes the automated detection and classification of the ingestive sounds of cattle (97.4% (Chelotti, 2016) and 94% (Milone, 2012)), pig coughs (97.8% (Chung, 2013) and 86% (Jans, 2005)), cattle oestrus (94% (Chung, 2013)), and sheep, cattle and goat jaw movements (94% (Navon, 2013)). For all data sets, the proposed algorithm achieved comparable (i.e. cattle 95.78%) or higher (i.e. sheep 99.29% and dogs 99.67%) accuracy, when compared to the aforementioned examples.

Classification performance results for the sheep data set were very high, with 98.54% and 99.29% accuracy recorded for MFCC and DWT-based methods respectively (Table 12). The accuracy obtained for MFCCs was comparable to previous research (Bishop, et al., 2017) (99.22%), but the DWT-based features outperformed this work (99.29%). When compared to MFCCs, the DWT method also showed a slight advantage in precision (+0.43%) (Table 12), which can be interpreted as an increased ability to predict the positive class (i.e. sheep vocalisations). Although the DWT-based technique proved superior (+0.75% accuracy), further examination of the confidence intervals revels a substantial overlap between results (Fig. 9). There was only 0.9% difference in upper bounds, and a 0.58% difference in lower bounds, meaning there is no significant statistical difference between the results of each method. It should be noted that the confidence bounds for both methods were very small, with −0.62%/+0.48% for MFCCs, and −0.47%/+ 0.31% for the DWT-based method. (Table 8)

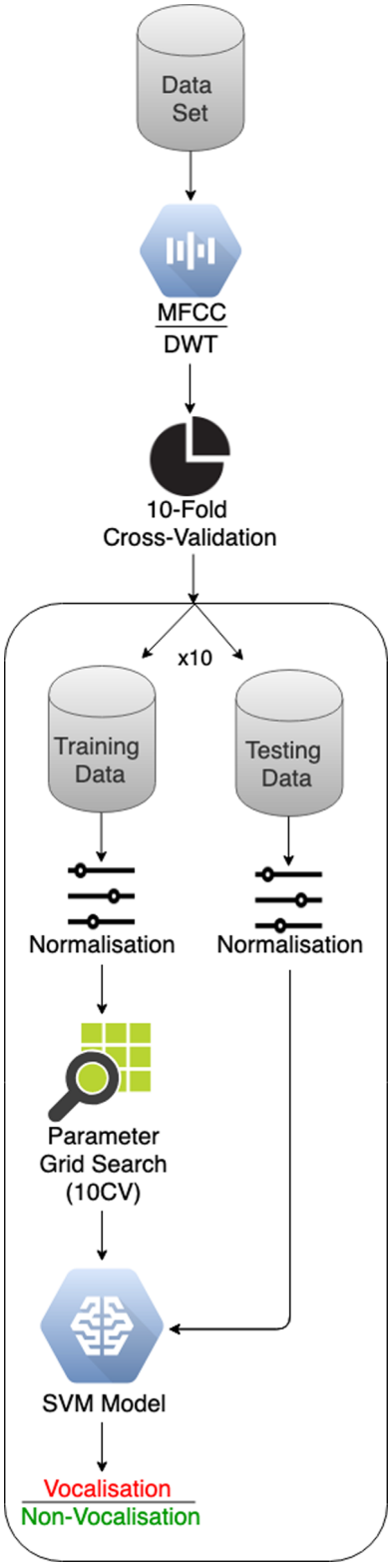

Fig. 8. Overview of algorithm training and testing procedure.

shows that within the sheep data set, the sheep class had a significantly higher average dBFS and SNR as compared to the negative subclasses, which may have had a positive effect on results. The high CR of the positive class might have been reduced by setting a more appropriate level of gain on the recording device. The composition of the sheep data set is suitable for testing some possible deployments, such as an on-collar device, as vocalisations will be far more likely to be louder than

Table 12
Aggregate performance results from stratified 10CV on each data set, using both MFCC and DWT-based features. Bold indicates the highest result for each metric, within each data set.

	Sheep		Cattle		Dogs	
	MFCC	DWT	MFCC	DWT	MFCC	DWT
TPR:	97.86	**98.93**	**94.61**	90.10	**99.58**	98.75
FPR:	0.77	**0.36**	**3.04**	4.22	**0.25**	1.08
TNR:	99.23	**99.64**	**96.96**	95.78	**99.75**	98.92
FNR:	2.14	**1.07**	**5.39**	9.90	**0.42**	1.25
ACC:	98.54	**99.29**	**95.78**	92.94	**99.67**	98.83
P:	99.22	**99.65**	**96.97**	95.56	**99.75**	98.92
F1:	**0.99**	**0.99**	**0.96**	0.94	**1.00**	0.99
AUC:	0.98	**1.00**	**0.97**	0.93	**1.00**	0.99

TPR:	True Positive Rate	ACC:	Accuracy
FPR:	False Positive Rate	P:	Precision
TNR:	True Negative Rate	F-1:	F-1 Score
FNR:	False Negative Rate	AUC:	Area Under Curve

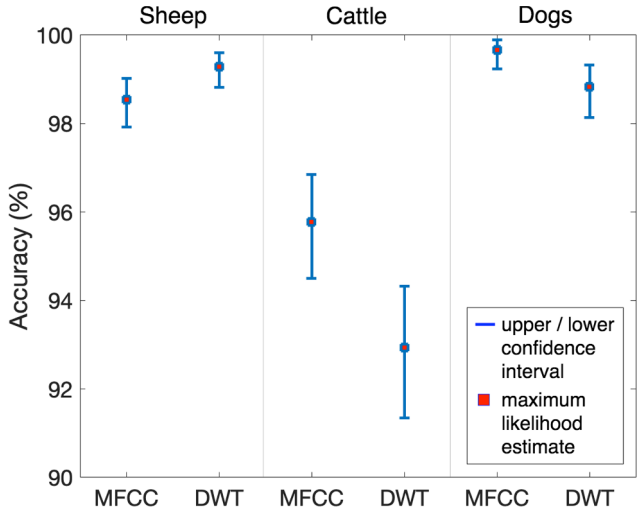

Fig. 9. Binomial proportion confidence intervals (99%), derived using the Clopper-Pearson method (Clopper and Pearson, 1934).

other sounds when the microphone is closer to the source. Performance in a more static deployment may require further evaluation using a more comprehensive data set. It is difficult to make an extensive comparison of the obtained results to other work in the field, as there appear to be few examples of automated sheep vocalisation detection in the literature (Bishop et al., 2017).

For the cattle data set, the MFCC feature extraction method was clearly superior to the DWT-based method, showing better performance over all metrics. MFCCs produced higher accuracy (+2.84%), precision (+1.41%), and F1-score (+0.02%) (Table 12). A comparison of the CIs for each method (Fig. 9) show no overlap between bounds, therefore it can be concluded that there is a small, though statistically significant difference between results. The overall recognition performance for the cattle data set was high, and was comparable to similar work in determining cattle oestrus via vocalisation detection (Chung, 2013). The proposed algorithm achieved a slightly lower accuracy of 95.78% (−1.92%), a better FPR of 3.04% (−1.06%), but an inferior FNR of 5.39% (+3.09%). Despite the algorithm's high performance, results from the cattle data set were distinctly lower than those obtained from the sheep and dog data sets. When comparing the highest result for each performance metric per data set, the cattle data set recorded 3.51 – 3.89% reduced accuracy, 2.68 – 2.78% diminished precision, 0.03 – 0.04 lower F-1 score, and 0.03 decreased AUC. An analysis of the

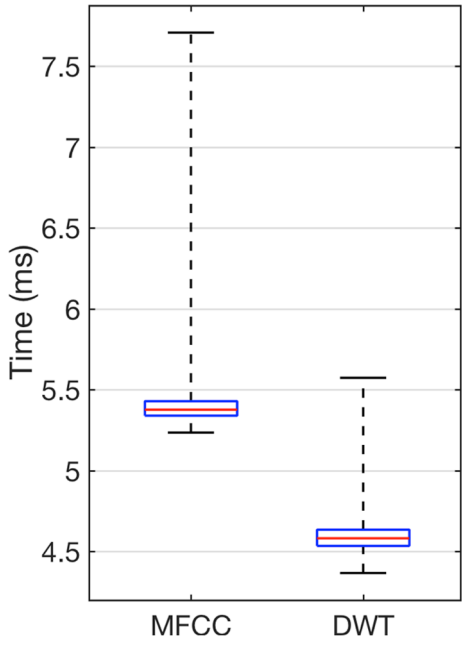

Fig. 10. Computational timing results (ms), for MFCC and DWT-based feature extraction techniques, calculated on a mid-2015 MacBook Pro (Apple. MacBook Pro (15-inch, Mid, 2015).

calculated CIs also reveals larger overall confidence bounds, as compared to the other data sets, with an average −0.75/+0.72% for MFCCs, and −1.01/+0.98% for DWT-based features. Based on these results, it appears that the cattle data set posed a much more challenging acoustic classification problem.

Based on spectrogram observation (e.g. Fig. 4 and Supplementary 6.1.2), there appears to be a higher level of audio frequency collisions between the positive and negative classes in the cattle data set, as compared to that of sheep and dogs. Cattle vocalisations typically have a large low frequency component (Fig. 4). The negative class of the cattle data set contained many instances of overhead planes and passing vehicles, both of which produce significant low frequency activity (Fig. 4). It is hypothesised that the algorithm had slight trouble differentiating between these similar types of sounds. The cattle data set had high rates of clipping in all classes (Table 8), and it is possible this contributed to the reduced classification performance. The gain setting on the recording device appears to have been set too high for the acoustic environment. The cattle data set contained a negative subclass (noise) that had a higher dBFS, SNR, and CR compared to the positive vocalisation class (Table 8): this may also have had a negative effect on performance. It can be difficult for an algorithm to discern between sounds with similar envelope or frequency composition, as was shown in similar work in the automatic recognition of jaw movements in cattle, goats and sheep (Navon, 2013). The SVM model applied equal misclassification costs to both the positive and negative classes. This cost could be altered to favour a lower FPR or FNR, depending on nature of the deployment. Overall, results for the cattle data set were high (> 95% accuracy), but further work is required to identify and explain the deviation of results from the other data sets.

The algorithm performed very well on the dog data set, with high accuracy obtained for both MFCCs (99.67%) and DWT-based features (98.83%) (Table 12). Accuracy results were markedly higher than those reported for other dog vocalisation detection problems (Yeo et al., 2012), but it is difficult to make comparisons as the problems are inherently different. Precision results were also high for both methods (MFCCs: 99.67%, DWT: 98.92%) (Table 12), demonstrating that the algorithm is good at detecting the positive (dog vocalisation) class. The dog data set had the lowest average dBFS, SNR, and CR (Table 8). This is expected when sounds have originated far from the capture device but could have been improved by setting a more appropriate level of gain on the recording unit. These results are encouraging for a static vocalisation detection deployment, where distance from source may be significant and variable. It may be beneficial to expand the goat negative subclass, due to the relatively low number of instances (Table 6), and the frequency overlap observed in spectrogram images (Fig. 5, Supplementary 6.1.3). Although MFCCs recorded better results in all performance metrics, an investigation of the CIs shows a large amount of overlap between CI bounds (Fig. 9). From this, it can be surmised that there is no significant statistical difference between the results for each feature extraction method, for the dog data set.

It is difficult to compare the feature extraction techniques based on classification performance alone. As has been shown, there is significant overlap between the CIs for each method in the sheep and dog data sets (Fig. 9). Whilst the higher results obtained in the cattle set for MFCCs were statistically significant, it was only a marginal difference. The results from the preliminary timing tests provide clearer distinction between the two compared methods. The DWT-based features were markedly faster to produce than the MFCCs, with an almost 1 ms difference between the median, minimum, and first and third quartile extraction times (Fig. 10). This equates to a 14.81 – 15.38% decrease in execution time. The difference in maximum execution time was even more pronounced, with an over 2 ms (27.26%) discrepancy between the MFCC and DWT-based methods. This shows that the DWT-based feature extraction technique is both computationally faster, and more consistent in terms of processing time when compared to MFCCs. It must be noted that the timing comparison of these two techniques was performed in MATLAB 2017a, using a relatively high-powered laptop. In order to target low-powered and embedded systems, minimisation of the algorithm's computational requirements would become more critical, as these systems are usually constrained by low CPU speeds and RAM capacities. Under these conditions, it would be ideal to perform comparisons with significantly less overhead (e.g. by using C or C++ modules executed directly on target hardware).

There were some limitations with the methodology of this study. In order to focus on a performance comparison between the presented feature extraction methods, and on the applicability of an SVM to livestock vocalisation classification, audio segmentation was performed manually. To simulate automated segmentation, all identified instances within a specified time period were extracted (see Section 2.1), leading to a large variation in dBFS, SNR, and CR both between and within data sets (Table 8). As animal behaviour was not directly observed, identification of target vocalisations was done aurally and by spectrogram inspection, which may lead to instances being missed, incorrect labelling of sounds, and introduction of selection bias. Future audio data collection should include visual recording, or direct observation of animal behaviour, in order to confirm when vocalisation have occurred. As hundreds of hours of recordings were collected (720 h for the sheep data set alone), it was infeasible to manually process all the data. As only a few hours of data, identified as containing high-levels of vocalisation activity, were used to produce each data set, there may be some selection bias present. The time-consuming nature of manual segmentation only further reinforces the need for an automated livestock vocalisation detection algorithm capable of segmenting and classifying sounds, and which can be easily retrained to target different animals. The proposed algorithm is capable of accurately classifying livestock vocalisations but would require an automated segmentation component in order to function as a livestock detection algorithm in a real-world deployment. Future research will focus on the development of this component, using a combination of an energy-based method for determining initial segmentation boundaries, and a metric-based method for acoustic change point detection, in order to segment sounds that occur in close proximity. The performance of the presented classification algorithm can then be re-evaluated when combined with automated segmentation.

The high classification results obtained must be interpreted in the context of the study. The classification models were trained and tested using data from the same acoustic environments, captured during the same deployments. This represents the 'ideal' conditions for the model, in terms of similarities in the soundscape. This study did not test generalisation to other farms and deployments, so does not fully address the suitability and adaptability of the algorithm to 'real-world' scenarios. It is expected that classification performance will be reduced when testing with sounds captured from new locations, but the level of reduction is yet to be ascertained. The results presented are very encouraging, but further work is needed to test and develop an algorithm truly capable of accurately detecting livestock vocalisations across different locations.

5. Conclusion

A multipurpose livestock vocalisation classification algorithm was developed using a combination of an SVM ML model, and either MFCC or DWT-based feature extraction. Performance of the algorithm was evaluated using 3 different data sets, each targeting a different animal (i.e. sheep, cattle, dogs), and under different acoustic conditions (Table 8). Data was manually extracted from audio acquired from fully operational outdoor farm environments. MFCC and DWT-based feature extraction techniques were compared, using both classification and computational timing results. 10CV was used to train and test an SVM model, using a grid search for hyperparameter selection (Fig. 8). High classification performance was observed across all 3 data sets, using both feature extraction methods (Table 12, Fig. 9). For the sheep data set, the DWT-based method achieved higher performance metrics, but an analysis of the CI's revealed no statistical difference between each method (Fig. 9). The cattle data set recorded relatively lower overall performance than the sheep and dog data sets, suggesting that it posed a more challenging acoustic classification problem. MFCCs obtained higher classification performance for the dog data set (Table 12), but as with the sheep data set, the CI's showed no statistical difference between each method (Fig. 9). The results for the computational timing required for each method were clearer, with the DWT-based method proving to be faster and with much less variation in timing (Fig. 10).

The high classification performance achieved paves the way for further development of an automated livestock vocalisation detection algorithm. A limitation of the proposed algorithm was the manual segmentation and extraction of sound instances. Future work will focus on the development of an automated segmentation component which can be used in conjunction with the presented algorithm to facilitate further testing of classification performance. This includes testing the model and algorithm in other acoustic environments, with an emphasis on developing a model capable of generalising to different deployment. Once this 'offline' capacity has been achieved, then the algorithm can be extended to real-time applications, with simulation testing on target hardware (Stover, 2017), using either static nodes or on-animal devices.

Acknowledgements

James Bishop is supported by an Australian Postgraduate Award (APA).

Funding support was in part provided by Australian Wool Innovation (AWI), and the Australian Federal Government Department of Agriculture and Water Resources.

Ethics approval was granted by the University of New England Animal Ethics Committee (AEC15-135).

Thanks to Joshua Stover and Arachchige Surantha Ashan Salgadoe for their assistance during the editing process, to Huw Nolan and Derek Schneider for their technical assistance, and to all of the livestock producers who allowed access to their properties.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2019.04.020.

References

FAO, 2009. The State of Food and Agriculture: Trends in the Livestock Sector. Food and Agriculture Organization of the United Nations., pp. 198–213.
Thornton, P.K., 2010. Livestock production: recent trends, future prospects. Phil. Trans. London Royal Soc. B: Biol. Sci. 365, 2853–2867.
Moynagh, J., 2000. EU regulation and consumer demand for animal welfare. AgBioForum 107–114.
Grandin, T., 2014. Animal welfare and society concerns finding the missing link. Meat Sci. 98, 461–469.
Berckmans, D., 2014. Precision livestock farming technologies for welfare management. Revue Scientifique et Technique 33 (1), 189–196.
Chung, Y., et al., 2013. Automatic detection of Cow's oestrus in audio surveillance system. Asian-Australas J. Anim. Sci. 26 (7), 1030–1037.
Exadaktylos, V., Silva, M., Berckmans, D., 2014. Automatic identification and interpretation of animal sounds, application to livestock. Prod. Opt.
Sadgrove, E.J., et al., 2017. Fast object detection in pastoral landscapes using a colour feature extreme learning machine. Comp. Electron. Agri. 139, 204–212.
Alvarengaa, F.A.P., et al., 2016. Using a three-axis accelerometer to identify and classify sheep behaviour at pasture. Appl. Animal Behav. Sci. 181, 91–99.
Barwick, J., et al., 2018. Categorising sheep activity using a tri-axial accelerometer. Comp. Electron. Agri. 145, 289–297.
Barwick, J., et al., 2018. Predicting lameness in sheep activity using tri-axial acceleration signals. Animals 8 (12), 1–16.
Falzon, G., et al., 2013. A relationship between faecal egg counts and the distance travelled by sheep. Small Ruminant Res. 111 (1–3), 171–174.
Ferrari, S., et al., 2008. Analysis of cough sounds for diagnosis of respiratory infections in intensive pig farming. Am. Soc. Agricult. Biol. Eng. 51 (3), 1051–1055.
Van Hirtum, A., Berckmans, D., 2004. Objective recognition of cough sound as a biomarker for aerial pollutants. Indoor Air 14, 10–15.
Chung, Y., et al., 2013. Automatic detection and recognition of pig wasting diseases using sound data in audio surveillance systems. Sensors (Basel) 13 (10), 12929–12942.
Jans, P., et al., 2005. Evaluation of an algorithm for cough detection in pig houses. 16th IFAC World Congress. 2005. Prague, Czech Republic 2005.
Guarino, M., et al., 2008. Field test of algorithm for automatic cough detection in pig houses. Comp. Electron. Agri. 62 (1), 22–28.
Chedad, A., et al., 2001. Recognition system for pig cough based on probabilistic neural networks. J. Agricult. Eng. Res. 79 (4), 449–457.
Exadaktylos, V., et al., 2011. Sound localisation in practice: an application in localisation of sick animals in commercial piggeries. Adv. Sound Local. 575–590.
Vandermeulen, J., et al., 2016. Early recognition of bovine respiratory disease in calves using automated continuous monitoring of cough sounds. Comp. Electron. Agri. 129, 15–26.
Aydin, A., Bahr, C., Berckmans, D., 2015. A real-time monitoring tool to automatically measure the feed intakes of multiple broiler chickens by sound analysis. Comp. Electron. Agri. 114, 1–6.
Aydin, A., Berckmans, D., 2016. Using sound technology to automatically detect the short-term feeding behaviours of broiler chickens. Comp. Electron. Agri. 121, 25–31.
Chelotti, J.O., et al., 2016. A real-time algorithm for acoustic monitoring of ingestive behavior of grazing cattle. Comp. Electron. Agri. 127, 64–75.
Clapham, W.M., et al., 2011. Acoustic monitoring system to quantify ingestive behavior of free-grazing cattle. Comp. Electron. Agri. 76 (1).
Milone, D.H., et al., 2012. Automatic recognition of ingestive sounds of cattle based on hidden Markov models. Comp. Electron. Agri. 87, 51–55.
Galli, J.R., et al., 2006. Acoustic monitoring of chewing and intake of fresh and dry forages in steers. Animal Feed Sci. Technol. 128, 14–30.
Ungar, E.D., Rutter, S.M., 2006. Classifying cattle jaw movements: comparing IGER behaviour recorder and acoustic techniques. Appl. Animal Behav. Sci. 98, 11–27.
Andriamasinoro, A.L.H., et al., 2016. A review on the use of sensors to monitor cattle jaw movements and behavior when grazing. Biotechnol., Agronomy, Soc. Environ. 20 (S1), 273–286.
Navon, S., et al., 2013. Automatic recognition of jaw movements in free-ranging cattle, goats and sheep, using acoustic monitoring. Biosyst. Eng. 114, 474–483.
Milone, D.H., et al., 2009. Computational method for segmentation and classification of ingestive sounds in sheep. Comp. Electron. Agri. 65 (2), 228–237.
Lee, J., et al., 2014. Formant-based acoustic features for cow's estrus detection in audio surveillance system. 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS).
Meen, G.H., et al., 2015. Sound analysis in dairy cattle vocalisation as a potential welfare monitor. Comp. Electron. Agri. 118, 111–115.
Pereira, E.M., Naas, I.A., Jacob, F.C., 2011. Using vocalization pattern to assess broiler's well-being. Precision Livestock Farming 11, 11–14.
Fontana, I., et al., 2014. Broiler vocalisation to predict the growth. Measuring Behavior. Wageningen, The Netherlands.
Moura, D.J., et al., 2008. Real time computer stress monitoring of piglets using vocalization analysis. Comp. Electron. Agri. 64 (1), 11–18.
Bishop, J.C., et al., 2017. Sound analysis and detection, and the potential for precision livestock farming - a sheep vocalization case study. In: 1st Asian-Australasian Conference on Precision Pastures and Livestock Farming. 2017: Hamilton, New

Zealand, pp. 1–7.
Tiwari, V., 2010. MFCC and its applications in speaker recognition. Int. J. Emerging Technol. 1 (1), 19–22.
Sharan, R.V., Moir, T.J., 2017. Robust acoustic event classification using deep neural networks. Inform. Sci. 396, 24–32.
Ahmad, J., et al., 2015. Gender identification using MFCC for telephone applications - a comparative study. Int. J. Comp. Sci. Electron. Eng. 3 (5), 351–355.
Davis, S.B., Mermelstein, P., 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoustics, Speech, Signal Process. 28 (4), 357–366.
Abdalla, M.I., Abobakr, H.M., Gaafar, T.S., 2013. DWT and MFCCs based feature extraction methods for isolated word recognition. Int. J. Comp. Appl. 69, 21–26.
Rabaoui, A., et al., 2008. Using one-class SVMs and wavelets for audio surveillance. IEEE Trans. Inform. Forensics Security 3 (4), 763–775.
Ramalingam, T., Dhanalakshmi, P., 2014. Speech/music classifcation using wavelet based feature extraction techniques. J. Comp. Sci. 10 (1), 34–44.
Olkkonen, H., 2011. Discrete wavelet transforms: algorithms and applications. In: Olkkonen, H. (Ed.), Rijeka, Croatia: InTech.
Virtanen, T., Singh, R., Raj, B., 2012. Techniques for noise robustness in automatic speech recognition. Wiley, New York, NY, USA.
Banakar, A., Sadeghi, M., Shushtari, A., 2016. An intelligent device for diagnosing avian diseases: newcastle, infectious bronchitis, avian influenza. Comp. Electron. Agri. 127, 744–753.
Deng, X., et al., 2010. Eggshell crack detection using a wavelet-based support vector machine. Comp. Electron. Agri. 70 (1), 135–143.
Vapnik, V., Golowich, S.E., Smola, A., 1997. Support vector method for function approximation, regression estimation, and signal processing. Adv. Neural Inform. Process. Syst. 9, 281–287.
AWI, 2017. Australian wool production forecast report. Australian Wool Innovation Limited (AWI).
MLA, 2018. Australian Cattle Industry Projections 2018. Meat & Livestock Australia (MLA).
Smith, M.E., et al., 2000. Review of methods to reduce livestock depradation: I. Guardian animals. Acta Agriculturae Scandinavica Section A — Animal Science 50 (4), 279–290.
Wildlife Acoustics, I. Song Meter SM3. 2014; Available from: < https://www.wildlifeacoustics.com/images/documentation/SM3-USER-GUIDE.pdf > .
Wildlife Acoustics, I. Song Meter SM2. 2010; Available from: < http://media.nhbs.com/equipment/sm2-manual.pdf > .
Audacity. Audacity: free, open source, cross-platform software for recording and editing sounds. 2017; Available from: < http://audacityteam.org/ > .
Gaberson, H., 2006. A Comprehensive Windows Tutorial. Sound Vibration 14–23.
Cannam, C., Landone, C., Sandler, M., 2010. Sonic visualiser: an open source application for viewing, analysing, and annotating music audio files. International ACM Multimedia Conference. 2010: Firenze, Italy.
The MathWorks, I. MATLAB R2017a. 2017; Available from: < https://au.mathworks.com/products/matlab.html > .
Castán, D., et al., 2015. Albayzín-2014 evaluation: audio segmentation and classification in broadcast news domains. EURASIP J. Audio, Speech, Music Process. 33, 1–9.
Huang, Z., et al., 2013. A blind segmentation approach to acoustic event detection based on i-vector. In: 14th Annual Conference of the International Speech Communication Association. 2013: Lyon, France, pp. 2282–2286.
Bhandari, G.M., Kawitkar, R.S., Borawake, M.C., 2013. Audio segmentation for speech recognition using segment features. Int. J. Comp. Technol. Appl. 4 (2), 182–186.
panagiotakis, C., tziritas, g., 2005. A speech/music discriminator based on RMS and zero-crossings. IEEE Trans. Multimedia 7 (1), 155–166.
Kemp, T., et al., 2000. Strategies for automatic segmentation of audio data. IEEE International Conference on Acoustics, Speech, and Signal Processing.
Rybach, D., et al., 2009. Audio segmentation for speech recognition using segment features. IEEE International Conference on Acoustics, Speech and Signal Processing. 2009: Taipei, Taiwan.
Virtanen, T., Helen, M., 2007. Probabilistic model based similarity measures for audio query-ByExample. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. 2007: New Paltz, NY.
Ozan, E.C., et al., 2014. An unsupervised audio segmentation method using Bayesian information criterion. 6th International Symposium on Communications, Control and Signal Processing (ISCCSP). 2014: Athens, Greece.
Patel, K., 2013. Speech recognition and verification using MFCC and VQ. Int. J. Emerging Sci. Eng. (IJESE) 1 (7), 33–37.
Sharan, R.V., Moir, T.J., 2016. An overview of applications and advancements in automatic sound recognition. Neurocomputing 200, 22–34.
Sharan, R.V., Moir, T.J., 2015. Subband spectral histogram feature for improved sound recognition in low SNR conditions. In: IEEE International Conference on Digital Signal Processing. IEEE, Singapore, pp. 432–435.
Young, S., et al., 2002. The HTK Book (version 3.2). Cambridge University: Engineering Department, pp. 1–355.
Solera-Urena, R., et al., 2007. Robust ASR using support vector machines. Speech Commun. 49 (4), 253–288.
Paliwal, K.K., 1999. Decorrelated and Liftered Filter-Bank Energies for Robust Speech Recognition. EUROSPEECH, Budapest, Hungary.
Zheng, F., Zhang, G., Song, Z., 2001. Comparison of different implementations of MFCC. J. Comp. Sci. Technol. 16 (6), 582–589.
Daubechies, I., 1992. Ten lectures on wavelets. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 11 (7), 674–693.
Safty, S.E., El-Zonkoly, A., 2008. Applying wavelet entropy principle in fault classification. World Acad. Sci., Eng. Technol. 40, 133–135.
Tzanetakis, G., Essl, G., Cook, P., 2002. Audio analysis using the discrete wavelet transform. WSES International Conference of Acoustics and Music: Theory and Applications. 2002. Skiathos, Greece.
Tyagi, S., Panigrahi, S.K., 2017. A DWT and SVM based method for rolling element bearing fault diagnosis and its comparison with artificial neural networks. J. Appl. Comp. Mech. 3 (1), 80–91.
Apple. MacBook Pro (15-inch, Mid 2015). 2015; Available from: https://support.apple.com/kb/sp719?locale=en_AU.
James, G., et al., 2014. *Introduction to Statistical Learning with applications in R*. Second ed. Wiley.
Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. Springer.
Hsu, C.W., Chang, C.C., Lin, C.J., 2010. A Practical Guide to Support Vector Classification. National Taiwan University, Taiwan.
Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intelligent Syst. Technol. 2 (3), 1–27.
Clopper, C., Pearson, E.S., 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26, 404–413.
Yeo, C.Y., Al-Haddad, S.A.R., Ng, C.K., 2012. Dog Voice identification (ID) for detection system. Second International Conference on Digital Information Processing and Communications (ICDIPC). IEEE, Klaipeda City, Lithuania.
Stover, J., et al., 2017. Hardware and embedded algorithms for real time variable rate fertiliser applications. 7th Asian-Australasian Conference on Precision Agriculture. Hamilton, New Zealand.