Potential of Accelerometers and GPS Tracking to Remotely Detect Perennial Ryegrass Staggers in Sheep

Article · February 2022

DOI: 10.1016/j.atech.2022.100040

9 authors, including:

Ly Trieu
New Mexico State University

Derek W. Bailey
New Mexico State University

Tran Cao Son
New Mexico State University

David Scobie
Lincoln University New Zealand

Smart Agricultural Technology 2 (2022) 100040

Contents lists available at ScienceDirect

Smart Agricultural Technology

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

Potential of Accelerometers and GPS Tracking to Remotely Detect Perennial Ryegrass Staggers in Sheep

Ly Ly Trieu[a], Derek W. Bailey[a,*], Huiping Cao[a], Tran Cao Son[a], David R. Scobie[b], Mark G. Trotter[c], David E. Hume[d], B. Lee Sutherland[b], Colin T. Tobin[e]

[a] New Mexico State University, Las Cruces, NM, USA
[b] AgResearch, Lincoln Research Centre, Lincoln, New Zealand
[c] Central Queensland University, Rockhampton, QLD, Australia
[d] AgResearch, Grasslands Research Centre, Palmerston North, New Zealand
[e] North Dakota State University, Carrington, ND, USA

ARTICLE INFO

Keywords:
Accelerometer
On-animal sensor
Machine learning
Perennial ryegrass staggers

ABSTRACT

Perennial ryegrass staggers (staggers) is a neurotoxic condition in livestock that is caused by consumption of ryegrass (*Lolium perenne*) infected with specific strains of *Epichloë* fungal endophytes. These grass-endophyte associations produce toxins that can adversely affect animals and can in some cases lead to death. In sheep, symptoms typically include head shaking, changes in gait, stiffness and falling. Affected sheep can recover after removing them from pastures containing toxic strains of endophyte. A pilot case study was conducted in Lincoln, New Zealand to determine if ryegrass staggers could be identified with data collected through GPS tracking and accelerometers. Fourteen sheep per treatment grazed in either a toxic endophyte-infected ryegrass paddock or an endophyte-free control paddock for 17 days in late March and early April 2017. Randomly selected sheep were fitted with collars containing a 3-axis accelerometer recording movements at 12 Hz (10 collars in endophyte infected paddock and 6 in the control paddock). Three sheep per treatment were also tracked at 3-minute intervals with GPS receivers. Sheep were scored by an experienced observer for symptoms of staggers weekly and at the end of the study using a 0 to 5 scale. Control sheep did not display any symptoms of staggers and 10 sheep in the infected pasture displayed little or no symptoms (0 or 1 score). The other 4 sheep in infected pasture had scores from 2 to 4 at the end of the study. Sheep grazing in the infected pasture (2.91 m/min ± 0.04 SE) moved slower ($P=0.04$) than sheep in the control pasture (3.12 m/min ± 0.05 SE). Distance travelled varied among days, but there did not appear to be any temporal trends. Machine learning analyses of accelerometer data showed that the behavior of affected sheep changed during the study. Activity of sheep displaying symptoms (scores ≥ 2) increased more in the morning and midday during the latter part of the study than control sheep and sheep with few or no symptoms (score < 2). However, behavior of individual sheep at night remained relatively consistent during the study. Accelerometers may be useful for remotely detecting perennial ryegrass staggers.

1. Introduction

Sheep raised on perennial ryegrass (*Lolium perenne* L) pastures can suffer from perennial ryegrass staggers [10]. It is a neurotoxic condition that afflicts several species of grazing livestock that is the result of toxins produced when specific strains of an asexual, mutualistic *Epichloë* fungal endophyte infect the host ryegrass. When animals consume forage with these toxins and become ill, symptoms develop gradually over time. In most cases, animals recover when moved to pastures with no endophyte or to pastures infected with selected, non-toxic endophytes.

It is crucial to monitor sheep, and other livestock, grazing endophyte infected pastures to ensure animal welfare. In addition, on-animal sensing may be useful for understanding the complex interactions of grazing and prescribed fire on toxin levels and livestock responses in rangelands invaded by or dominated by endophyte infected forages [29–31] such as tall fescue *(Schedonorus arundinaceus* [Schreb.] Dumort. nom. cons.). Visual monitoring or periodic testing for staggers is time consuming and labor intensive, especially when sheep are kept in large groups or in extensively farmed pastures. On-animal sensors have potential to remotely monitor and identify changes in animal behavior [2]. The behaviors expressed by the animal can be utilized to detect when livestock are becoming ill. In this paper, we explore this concept via sensors applied to sheep grazing endophyte infected perennial ryegrass.

* Corresponding author. Derek W. Bailey, Animal and Range Sciences, New Mexico State University, PO Box 30003 MSC 3-I, Las Cruces, NM 88003, USA.
E-mail address: dwbailey@nmsu.edu (D.W. Bailey).

https://doi.org/10.1016/j.atech.2022.100040
Received 6 January 2022; Received in revised form 7 February 2022; Accepted 7 February 2022

Remote monitoring of livestock well-being using GPS tracking and accelerometers have been evaluated in different environments and applications. Lambing can be identified using GPS tracking [12, 15] and accelerometers [16]. Machine learning (ML) can be used to classify sheep behavior from accelerometer data [16]. Accelerometers can be used to detect diseases such as bovine ephemeral fever in cattle [34]. Recent developments have promised to provide GPS tracking and accelerometer data in real time or near real time such as the internet of things, LoRaWAN [32, 35] and satellite technologies. Real time monitoring of livestock may allow farmers and ranchers to respond more rapidly when animals become ill.

The objective of this paper is to provide a "proof of concept study" to evaluate the potential of detecting perennial ryegrass staggers through remote monitoring of sheep with GPS tracking and accelerometers. We apply classification ML algorithms to accelerometer data from monitored sheep. The study hypothesizes that accelerometer and GPS tracking data will be able to identify the changes of behavior that occur when sheep are affected by perennial ryegrass staggers.

2. Material and methods

2.1. Study site

The study was conducted at AgResearch Lincoln, New Zealand from 23 March to 10 April 2017. It used two experimental paddocks each 0.1625 ha which had been sown as pure stands of 'GA66' diploid perennial ryegrass in October 2016. One paddock contained endophyte infected perennial ryegrass, with the endophyte being a 'standard' (also referred to as wild-type or common toxic) strain of *Epichloë festucae* var. *lolii*. The standard strain is known to cause ryegrass staggers in sheep and other livestock [10]. The other paddock was a control, with 'GA66' perennial ryegrass that did not contain endophyte.

2.2. Animals

The protocol for this study was approved by the AgResearch Invermay Animal Ethics Committee (Dunedin, NZ; application # 14103).

A total of 28 crossbred female sheep that were 18 months old, were randomly assigned to the two paddocks (14 sheep per paddock). The mean live weight of the sheep was 56.3 kg ± 0.6 SE at the beginning of the study.

All 14 sheep in the endophyte infected paddock and eight of the 14 sheep in the control paddock were fitted with Gulf Coast X-16-4 Accelerometers (Gulf Coast Data Concepts, LLC, Waveland, MS USA). Accelerometers were fitted on collars around the neck of the sheep and recorded movements of the x, y and z axes. Movement of the head fore and aft was associated with the x axis. Left and right head movements were associated with the y axis and up and down movements were associated with the z axis.. Movement data were recorded at a rate of 12 Hz and were stored on the accelerometer until retrieval at the end of the study. Three sheep in each paddock were tracked with igotU 600 GPS receivers (MobileAction, New Taipei City, Taiwan) on collars. Positions were recorded at 3-minute intervals. One GPS receiver in the control paddock did not record any positions for the entire study period most likely due to battery failure. In this experiment, positions recorded outside the paddocks were deleted from the data set. ArcMap geographical information software (ESRI ArcGIS Suite, Redlands, California USA) was used to identify positions recorded outside of the paddocks.

2.3. Study design

The trial was based on protocols developed and modified over many years in the evaluation of novel grass-endophyte associations, as best described by [14]. Sheep grazed control and endophyte paddocks during the autumn of 2017 from March 23 to April 10. Sheep were scored for stagger symptoms using a 0 to 5 scale [20] in which score 0 means no symptoms, score 5 means severe tremors after minor disturbance or exercise. The sheep were chased out of their treatment paddock and down the raceway between paddocks for a few minutes at a modest run. Stagger scores were recorded on March 30, April 3 and April 10. Most sheep (21 of 28) were weighed at the beginning of the study, and all sheep were weighed at the end of the study.

3. GPS data analysis

Distance travelled each hour (m/hour) was calculated by summing the distances between all positions recorded during an hour. The hourly average travel rate (m/min) was calculated by averaging the velocity recorded by the GPS on an hourly basis. Distance travelled per day (m/day) was calculated by summing all the distances between positions recorded during a day (24 h). The daily average travel rate (m/min) was calculated by averaging all the velocities recorded by the GPS each day (24 h). Separate statistical analyses of GPS data were conducted for hour and day data.

Hourly metrics were analyzed using the repeated measures procedure of PROC MIXED in SAS (SAS Institute Inc., Cary, NC, USA; [23]). The fixed effects of the model consisted of treatment (endophyte or control), day, hour as well as the interactions of treatment by day and treatment by hour and treatment by hour by day. However, the two-way interactions and three-way interaction were not significant ($P > 0.05$). Consequently, they were removed from the final model. Thus, the final model was reduced to treatment, day and hour as fixed effects. The subject of repeated measures analyses was sheep. Covariance of repeated measures was modelled using auto-regressive order of 1 [AR(1)] covariance structure because it had the lowest Akaike's Information Criteria (AIC) value among three structures evaluated, AR(1), compound symmetry and unstructured [23].

Daily metrics were also analysed using repeated measures procedure of PROC MIXED in SAS [23]. The fixed effects of the model consisted of treatment (endophyte or control), day and the interaction of treatment by day. The subject of repeated measures analyses was sheep. Similar to the analysis of hourly metrics, AR(1) was selected because its AIC value was lower than the other structures evaluated, compound symmetry and unstructured [23].

4. Accelerometer data analyses

4.1. Data preprocessing

The raw accelerometer data recorded from 0600 h to 1800 h on the 3 days when staggers was scored and 1 day before the last sensor's record, were extracted and initially partitioned into 1-minute epochs. For one sheep, the accelerometer stopped just before the last stagger score was recorded so the day the third stagger score was recorded was not available. Accelerometer data from the day (0600 h to 1800 h) were used because sheep activity at night is lower than during the day and less likely to be informative (see results below). The 1-minute epochs were then averaged into 30-minute periods for machine learning classification. Thirty 1-minute epochs, without any missing data, were averaged for each 30-minute period. The 1-minute epoch does not miss any timestamps if it contains more than or equal to 700 timestamps; otherwise, the epoch is considered as a missed timestamp epoch. Only epochs without any missing timestamps were utilized in our experiment. The stagger scores were associated with the accelerometer epochs based on the timestamps. Data from sheep in the control paddock were not used for the machine learning analyses.

4.2. Metrics description

Nineteen metrics from [16] were calculated for each 1-minute epoch. Because the four metrics A_x, A_y, A_z, and A_{xyz} are perfectly collinear with each other via Variance Inflation Factor in our dataset, only averages of

Table 1
Sixteen features and the equation used to calculate values.

Feature	Equation
Average all-axis (A_{xyz})	$A_{xyz} = \frac{1}{T}\sum_1^T (x(t) + y(t) + z(t))$
Minimum X-axis (Min_x)	The minimum X-axis value in the epoch
Minimum Y-axis (Min_y)	The minimum Y-axis value in the epoch
Minimum Z-axis (Min_z)	The minimum Z-axis value in the epoch
Maximum X-axis (Max_x)	The maximum X-axis value in the epoch
Maximum Y-axis (Max_y)	The maximum Y-axis value in the epoch
Maximum Z-axis (Max_z)	The maximum Z-axis value in the epoch
Standard Deviation X-axis (SD_x)	$SD_x = \sqrt{\frac{1}{T}\sum_{t=1}^T (x(t) - \bar{x})^2}$ where $\bar{x}$ is the mean of X-axis value in the epoch
Standard Deviation Y-axis (SD_y)	$SD_y = \sqrt{\frac{1}{T}\sum_{t=1}^T (y(t) - \bar{y})^2}$ where $\bar{y}$ is the mean of Y-axis value in the epoch
Standard Deviation Z-axis (SD_z)	$SD_z = \sqrt{\frac{1}{T}\sum_{t=1}^T (z(t) - \bar{z})^2}$ where $\bar{z}$ is the mean of Z-axis value in the epoch
Average Standard Deviation all-axis (SD_{xyz})	$VAR_{xyz} = \frac{1}{3*T} * (SD_x^2 * T + SD_y^2 * T + SD_z^2 * T)$ $SD_{xyz} = \sqrt{VAR_{xyz}}$
Movement Intensity (MI)	$MI = \frac{1}{T}\sum_{t=1}^T \sqrt{x(t)^2 + y(t)^2 + z(t)^2}$
Signal Magnitude Area (SMA)	$SMA = \frac{1}{T}\sum_{t=1}^T (\lvert x(t)\rvert + \lvert y(t)\rvert + \lvert z(t)\rvert)$
Energy (Energy)	$Energy = \frac{1}{T}\sum_{t=1}^T (x(t)^2 + y(t)^2 + z(t)^2)^2$
Entropy (Entropy)	$Entropy = \frac{1}{T}\sum_{t=1}^T (1 + (x(t) + y(t) + z(t))^2) * ln\,(1 + (x(t) + y(t) + z(t))^2)$
Movement Variation (MV)	$MV = \frac{1}{T}\sum_{t=2}^T (\lvert x(t-1) - x(t)\rvert + \lvert y(t-1) - y(t)\rvert + \lvert z(t-1) - z(t)\rvert)$

the 3 axes (A_{xyz}) were evaluated and A_x, A_y, and A_z were not evaluated. Therefore, sixteen metrics in Table 1 were calculated for each 1-minute epoch. For datasets 1 and 2, the 1-minute epochs were averaged together into 30-minute instances.

- Dataset 1 has 30-minute instances and 2 classes:
  - label 0 corresponds to stagger scores 0, 1, and 2;
  - label 1 corresponds to stagger scores 3 and 4.
- Dataset 2 has 30-minute instances and 3 classes:
  - label 0 corresponds to stagger scores 0;
  - label 1 corresponds to stagger scores 1, and 2;
  - label 2 corresponds to stagger scores 3 and 4.
- Dataset 3 has 1-minute epochs and 5 classes with labels 0, 1, 2, 3, and 4, which correspond to stagger scores 0, 1, 2, 3, and 4, respectively.

Dataset 3 is much larger than datasets 1 and 2 because the duration of the instance is much shorter in dataset 3 and there are more classes (Table 2).

4.3. Machine learning of accelerometer data

Classification of perennial ryegrass staggers scores using accelerometer data falls within the scope of machine learning (ML). In typical classification problems, the 30-minute instances in datasets 1 and 2, or the 1-minute epochs in dataset 3 (Table 2) are regarded as data instances or objects. The stagger scores describing the level of sickness are denoted as the class/target of one data instance/object. The metrics describing the movement of a sheep (Table 1) are denoted as features or attributes. Each data instance consists of multiple features. In this study, ML approaches are utilized to classify a data instance to a class which indicates whether a sheep has staggers.

In addition to the missing data point problem, another major challenge in applying ML techniques to solve this classification problem is that the data are not balanced — there is a big difference in the number of instances from the two different classes: stagger and no-stagger, in dataset 1 (Table 2). Imbalanced data can result in poor predictive performance of classification models, specifically a low accuracy for the class with few examples (called the minority class). For example, in dataset 1, the number of instances of the stagger class is dramatically lower than that of non-stagger class, so the stagger class is the minority class. We are more interested in detecting the instances belonging to the stagger class than those in the no-stagger class; thus, the stagger class is more important than non-stagger class in dataset 1.

4.3.1. Model selection and resampling method

To deal with the class imbalance problem, the first technique we apply is *stratified cross validation* [21]. In building classification models, data are split into two sets, training and testing. The former set is used in ML algorithms to create a model. After building the model, the latter set is used to check the effectiveness of ML algorithms by calculating an evaluation metric (e.g. accuracy, F1 score, etc.). The division of the dataset is random. However, there is a chance that the minority class does not exist in the training set but appears in the testing set. If so, the model will not predict the test data correctly, resulting in a low evaluation metric. This is because, during the training process, the model does not learn and capture the information of the minority class in the training set and fails to represent the actual dataset population. Stratified sampling can help cope with the imbalance, because the technique splits data such that the percentage of instances for each class of the dataset is preserved in each set. To enhance the robustness of the model, we apply *k-fold cross-validation* in which the whole dataset is divided into k folds. Among the k folds, k-1 folds are used for training and one fold is used for testing. In our experiment, we use k = 10. With a combination of stratified sampling, the class distribution of each fold is equal to the proportion of instances of each class in the dataset.

Table 2
Number of instances and classes of the three datasets.

Dataset	The total number of instances				
	Class 0	Class 1	Class 2	Class 3	Class 4
1	970	47			
2	435	535	47		
3	13115	11745	4386	754	675

Secondly, to rebalance the class distribution in a dataset, oversampling and under-sampling approaches have been utilized in different machine learning applications [19, 25, 33]. In our study, we applied the Synthetic Minority Over-sampling Technique (SMOTE) [8] that has shown incredible performance in dealing with imbalanced datasets. SMOTE is an oversampling method that is used to increase the number of instances in the minority class such that the number of instances in minority class was equal to the number of instances in majority class. Instead of duplicating the instances in minority, SMOTE increases the number of instances of the minority class by constructing "synthetic" periods. By using SMOTE, the training set is not only more balanced but also diversified.

4.3.2. Features

Two types of features were developed to classify potential stagger scores. The first type is called domain features (Table 1) calculated directly from the accelerometer data. In addition, some classical feature extraction techniques such as Principal Component Analysis and Linear Discriminant Analysis (LDA) can improve the quality of domain features [13]. We applied LDA [27] in our study as the second type of feature. LDA is a supervised method that is used to reduce the number of dimensions by projecting data points in a high-dimensional dataset into a new low-dimensional feature space. LDA not only maximizes the separation between multiple classes but also minimizes the distances between instances within each class. In LDA, at most c-1 feature projections are produced, where c is the number of classes in a dataset. Therefore, we set up the numbers of features in our study to be 1, 2 and 4, which were the maximum for datasets 1, 2 and 3, respectively (because the numbers of classes in datasets 1, 2 and 3 are 2, 3 and 5, respectively). By using LDA, the number of features is reduced from 16 domain features to 1, 2 and 4 features in datasets 1, 2 and 3, respectively.

4.3.3. Machine learning classifiers

Many machine learning classification algorithms have been developed such as Decision Trees, Random Forests, Support Vector Machine (SVM), and k-nearest neighbors. In our study, we leverage the power of the ensemble approach [11] that combines different classifiers to produce a better predictive classifier than each individual classifier alone (called the non-ensemble approach). Among ensemble methods, Random Forests [6] is one of the most popular and powerful algorithms that has shown outstanding predictive performance in imbalanced datasets and was used in this study. Random Forests consists of multiple decision trees. Each individual tree provides a class prediction, and the class with the most votes from the many decision trees becomes the selected model's prediction. Furthermore, to compare the performance between ensemble and non-ensemble approaches, SVM [9] was chosen as a representative algorithm for non-ensemble approaches because it is a good-working supervised ML algorithm for classification from the literature and has been applied in classifying sheep behavior [16]. In SVM, a hyperplane (decision boundary) is computed to separate the data. The model chooses the best hyperplane, which maximizes the distance between the hyperplane and the training samples that are closest to the hyperplane. Moreover, when datasets are not linearly separate, SVM is kernelized to deal with non-linear classification problems. The idea for kernel SVM classification is as follows. First, non-linear data are transformed into the higher dimensional space in which the data becomes linearly separable. Then, a linear SVM model is applied to classify the data in the new space. To determine whether applying kernel SVM was necessary for our study, we calculated the least square error of a linear model developed from linear regression. The result was extremely low; thus, our dataset was non-linear. As a result, we applied kernel SVM in our study. To reduce the computational expensive transformation features in the new space, Radial Basis Function (RBF) kernel was utilized in kernel SVM. All the above algorithms are implemented using the *scikit-learn* library of Python [27].

4.3.4. Machine Learning approach evaluation

Different fundamental evaluation metrics such as accuracy, precision, recall, and F1-score, have been proposed to evaluate the performance of classification models. Because of the imbalanced class distribution in our study, F1-score is an effective metric to evaluate our model [33]. F1-score is computed using the following equation:

$$F1 = 2 * \frac{Precision\ * Recall}{Precision + Recall}$$

Where $Precision = \frac{TP}{TP+FP}$ and $Recall\ = \frac{TP}{TP+FN}$

In the above equations, true positive (TP) is the number of instances which are correctly classified as stagger classes by the models; false positive (FP) is the number of instances which are incorrectly classified as stagger classes by the models; false negative (FN) is the number of instances which are incorrectly as non-stagger class by the models.

4.4. Repeated measures statistical analyses of accelerometer data

The top four important features resulting from using Random Forests algorithm in *scikit-learn*, based on Gini impurity, are Minimum Z-axis (Min_z), Maximum Z-axis (Max_z), Average of all axes (A_{xyz}), and Movement Intensity (MI). To discover another possible set of top important features, we utilized PROC HPForest in SAS [26], which focuses on out-of-bag (OOB) error to select the most significant classification variables. The top four important features resulting from the HPForest algorithm in SAS, based on OOB Gini values, were (Maximum Z-axis (Max_z), Movement Intensity (MI), Maximum X-axis (Max_x), and Energy ($Energy$).

We conducted a separate statistical analysis using two common features between the two classification processes (*scikit-learn* library [27] and SAS): Max_z and MI. The maximum of the Z axis and MI were utilized as dependent variables and analyzed using the repeated measures of PROC MIXED in SAS [23]. The hourly means of Max_z and MI from all sheep monitored by accelerometers during the study period were used in the analysis. Fixed effects in the model consisted of period (period 1 – March 23 to March 27, period 2 – March 28 to 30 and period 3 – March 31 to April 6), Julian date within period, diurnal time class (morning - 0500 to 1000 h, midday - 1000 to 1700 h, and night - 1700 to 0500 h), hour within diurnal time class, last stagger score (LSS), the interaction of LSS by period, LSS by diurnal time class and the 3-way interaction of LSS by period by diurnal time class. The subject of the repeated measures analyses was the monitored sheep. Covariance of repeated measures was modelled using the compound symmetry structure because it had the lowest AIC value of the AR(1), compound symmetry and unstructured covariance structures evaluated [23].

5. Results

5.1. Sheep health

Most sheep did not display severe ryegrass staggers symptoms. No control sheep displayed symptoms. The mean last stagger score in the endophyte infected paddock was 1.14 (range 0 to 4). Control sheep (-2.14 kg ± 0.66 SE) lost more (P = 0.002) weight than sheep in the endophyte paddock (0.78 kg ± 0.42 SE).

5.2. GPS data

In daily analyses of distance travelled per day, control sheep (4403 m/day ± 93 SE) tended (P = 0.080) to travel farther each day than sheep in the endophyte infected paddock (4089 m/day ± 76 SE). Control sheep (3.08 m/min ±0.06 SE) tended (P = 0.075) to move faster (greater velocity) than sheep in the endophyte paddock (2.88 m/min ± 0.05 SE). Travel and velocity varied (P < 0.05) among days of the study. No interaction between treatment (endophyte vs control) and days was detected (P > 0.10) for either travel or velocity.

In hourly analyses, no differences in travel (m/hour) were detected between sheep in the control and endophyte paddocks (P = 0.17). Travel

Table 3
F1 score for accuracy when predicting perennial ryegrass staggers scores when applying Single Vector Machines (SVM) and Random Forests with domain features in three datasets. Bold indicates the highest F1 score.

	SVM	Random Forests
Dataset 1	0.887	**0.965**
Dataset 2	0.640	0.814
Dataset 3	0.646	0.871

Table 4
F1-score for accuracy when predicting perennial ryegrass staggers scores when applying Single Vector Machines (SVM) and Random Forests with Linear Discriminant Analysis (LDA) in three datasets. Bold indicates the highest F1 score.

	SVM	Random Forests
Dataset 1	0.813	**0.836**
Dataset 2	0.630	0.615
Dataset 3	0.596	0.669

per hour varied ($P < 0.001$) among hours of the day. No differences in travel per hour were detected ($P = 0.22$) among days and no interactions of day by treatment ($P = 0.57$) and hour by treatment ($P = 0.43$) were detected. The velocity of control sheep (3.12 m/min ± 0.05 SE) was greater ($P = 0.049$) than for sheep in the endophyte paddock (2.91 m/min ± 0.04 SE). Velocity varied ($P = 0.044$) among days. Velocity also varied ($P < 0.001$) among hours. No interactions between day and treatment ($P = 0.51$) and hour and treatment ($P = 0.31$) were detected for velocity.

5.3. Accelerometer data

5.3.1. Machine Learning analyses

The efficacy of SVM and Random Forests classification with domain features is shown in Table 3. With domain features, performance of Random Forests was much higher than SVM (Table 3). The highest performance (F1-score of 0.965) was with dataset 1 and Random Forests.

The use of SVM and Random Forests with LDA resulted in inferior performance than using domain features (Tables 3 and 4). Random Forests had the highest performance overall (Table 4). For datasets 1 and 3, the pattern of results using LDA was similar to the use of features (Table 3) where Random Forests shows higher F1-scores than SVM, but the result for dataset 2 shows that Random Forests is slightly worse than SVM. This is mainly because of the number of features and instances in these datasets. The detailed discussions about these results and inconsistencies can be found in Section 6.1.

5.3.2. Repeated measures analyses of accelerometer data

No differences in LSS were detected ($P = 0.66$) for Max_z. Maximum Z-axis varied among periods ($P < 0.0001$) and days within a period ($P = 0.0024$). Maximum Z-axis differed ($P < 0.0001$) among the diurnal time classes (morning, midday, night) and differed ($P < 0.0001$) among hours within diurnal time classes. A strong interaction was detected ($P < 0.0001$) between LSS and diurnal time class and between LSS and period ($P = 0.0092$). Moreover, there was a strong ($P < 0.0001$) 3-way interaction between LSS, period and diurnal time class (Fig. 1).

No differences in MI were detected for LSS ($P = 0.80$). Movement intensity varied ($P < 0.001$) among periods and days within period. In addition, MI varied ($P < 0.001$) among the diurnal time classes and hours with diurnal time class. There were strong interactions between LSS and period ($P = 0.001$) and LSS and diurnal time classes ($P < 0.001$). The 3-way interaction between LSS, period and diurnal time class was also important ($P < 0.001$) for MI (Fig. 2).

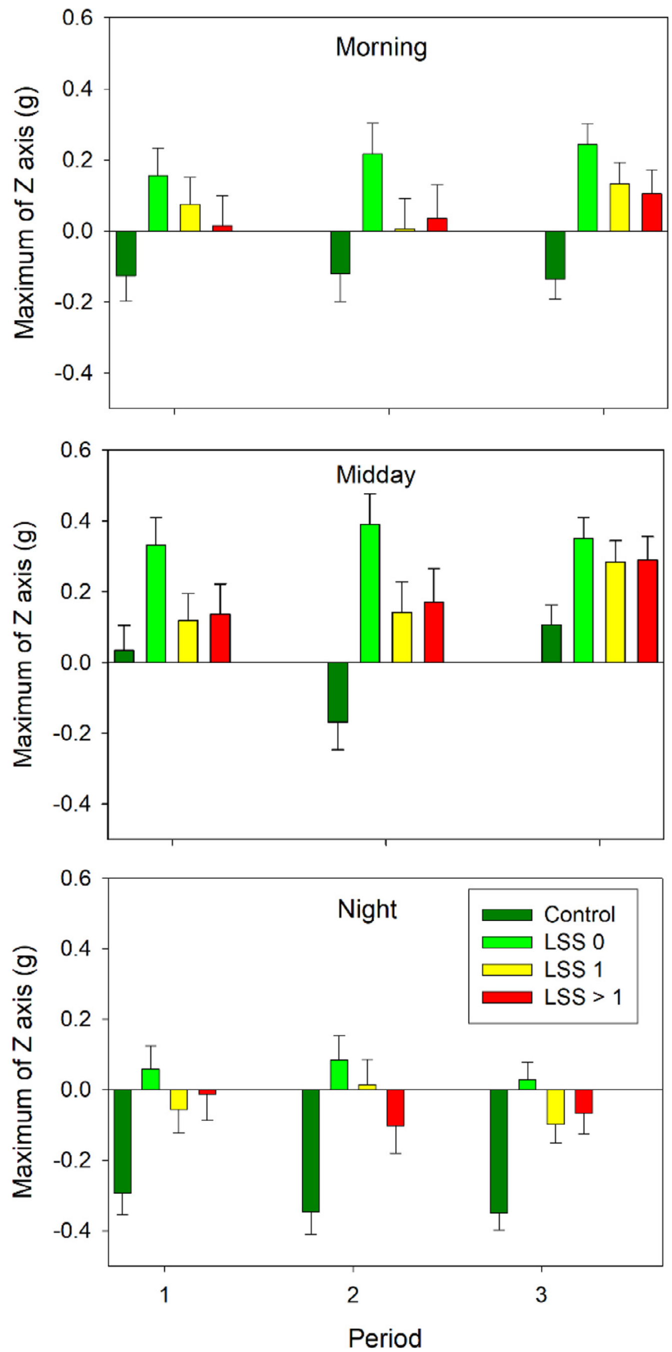

Fig. 1. Mean of the maximums of the Z axis by period during the morning (0500 to 1000), midday (1000 to 1700), and night (1700 to 0500). Means and standard errors (error bars) are provided for the control pasture (C) and last stagger scores (LSS) 0, 1, and > 1 in the endophyte infected pasture during period 1 (March 23 to 27), period 2 (March 28 to 31), and period 3 (April 1 to 6).

The diurnal activity patterns of two sheep that displayed moderate to severe ryegrass stagger symptoms (LSS 3 and 4) at the end of the study changed compared to the beginning of the study when their stagger score was 0 (Fig. 3 and 4). The periods of relatively constant values near zero (likely inactivity) of the y- and especially z-axis were longer at the end of the study than at the beginning. Although all 3 accelerometer axes showed changes in activity patterns, the y-axis (side to side head movements) and especially the z-axis (up and down head movements) showed the most notable changes.

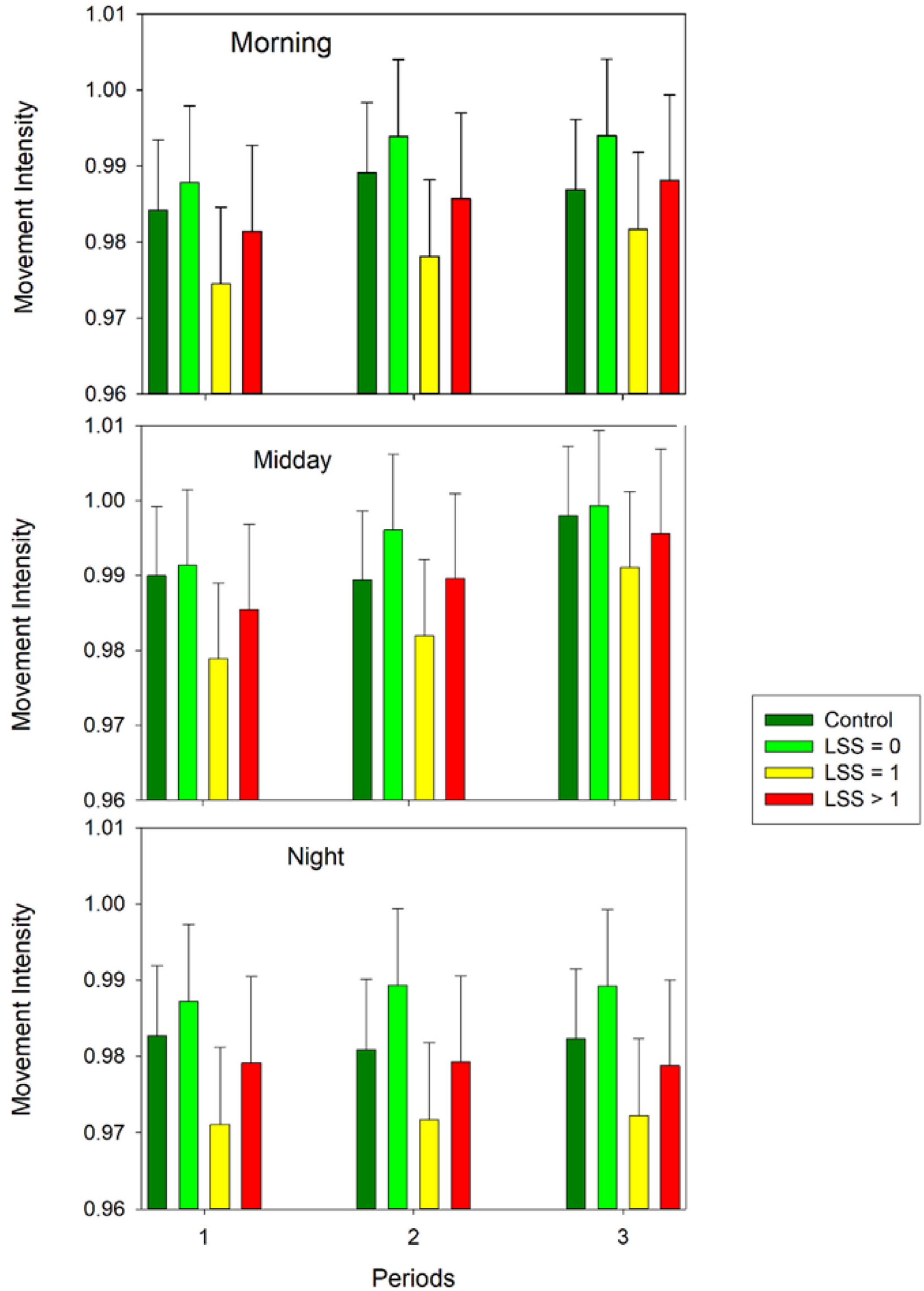

Fig. 2. Mean movement intensity (MI) by period during the morning (0500 to 1000), midday (1000 to 1700) and night (1700 to 0500). Means and standard errors (error bars) are provided for the control pasture (C) and last stagger scores (LSS) 0, 1 and > 1 in the endophyte infected pasture during period 1 (March 23 to 27), period 2 (March 28 to 31), and period 3 (April 1 to 6).

6. Discussion

In this study, only two sheep showed any moderate to severe ryegrass stagger symptoms (LSS 3 or 4). Weight gain of sheep grazing endophyte infected pasture was higher than control sheep supporting the assertion that ryegrass symptoms were minimal. The weather was cool during the study, which may have reduced herbage endophyte toxin concentrations in comparison to late summer when staggers are typically high at this trial site [14]. Although there is evidence that sheep in the endophyte paddock travelled slower and slightly less distance each day than control sheep, GPS tracking does not appear to be useful for detecting staggers. No patterns in the daily differences among days

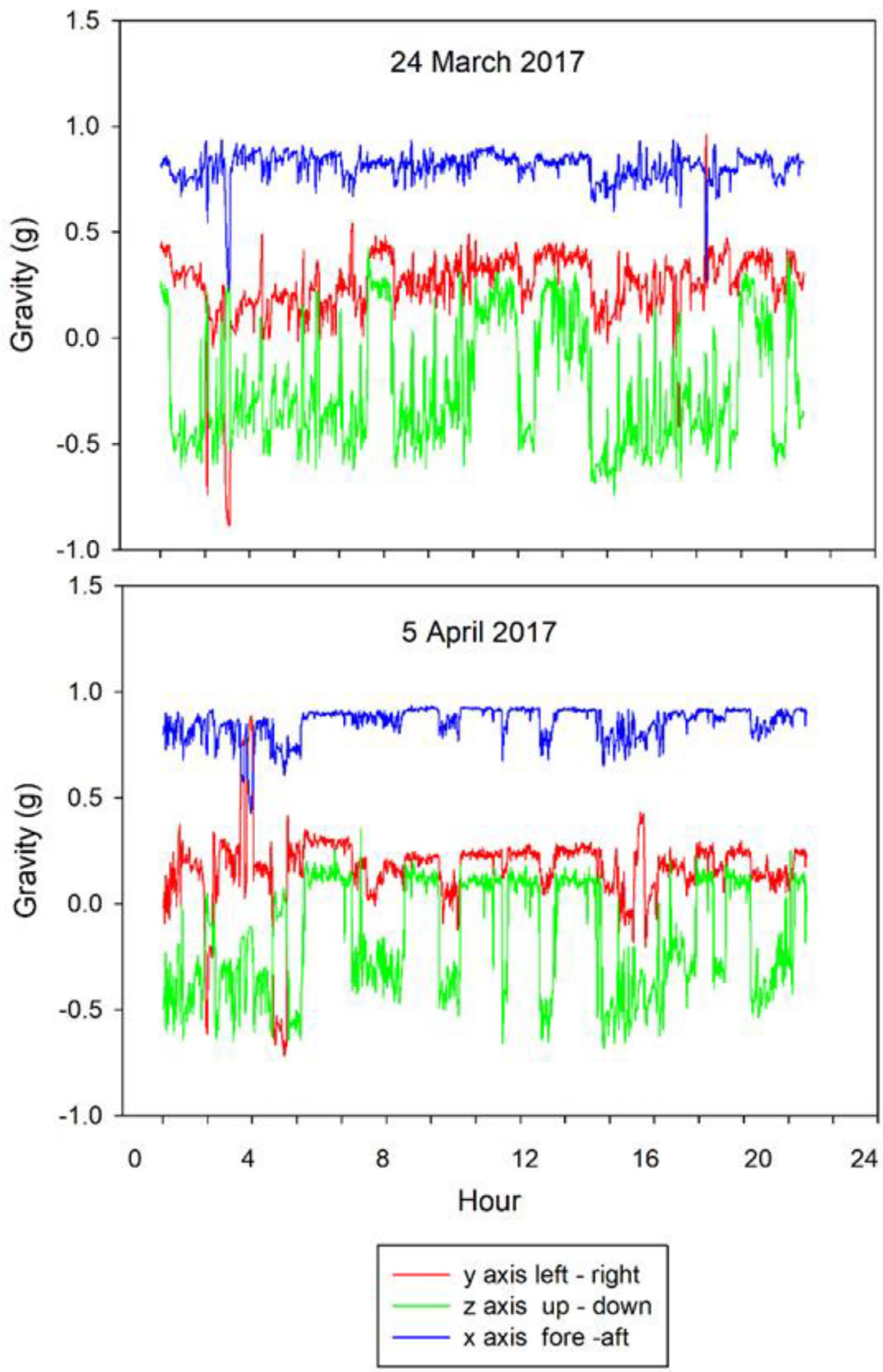

Fig. 3. Average 1-minute movement data from the x, y and z axes during the first and last 24-hour periods in study for sheep 17011 with the highest last stagger score (LSS) of 4. Accelerometer readings are given from 0000 h to 2359 h.

for velocity were detected. Distance travelled and velocity did not increase or decrease as the study progressed, which would be expected if GPS tracking was a potential indicator of ryegrass staggers. Instead, the differences in travel among days were erratic. Paddock size in this study was very small (0.1625 ha), which limited the ability of sheep to fully express spatial movements. Differences in travel between the endophyte-infected and control paddocks is likely an artifact of differences among sheep or differences in the forage characteristics of the paddocks. However, the amount of available forage was similar in both paddocks.

6.1. Machine Learning of accelerometer data explained a large amount of the stagger scores

The LDA generated features did not perform better than the domain features used in both Random Forests and SVM. The domain features for accelerometer data were adapted and combined by [16] based on equations from [1, 4, 5, 7, 24] to detect sheep behavior, and they performed well in this study. In contrast, LDA, in our study, projected the data points into perpendicular axes, but failed to help in classification because the features were directly informative. Such feature reduction

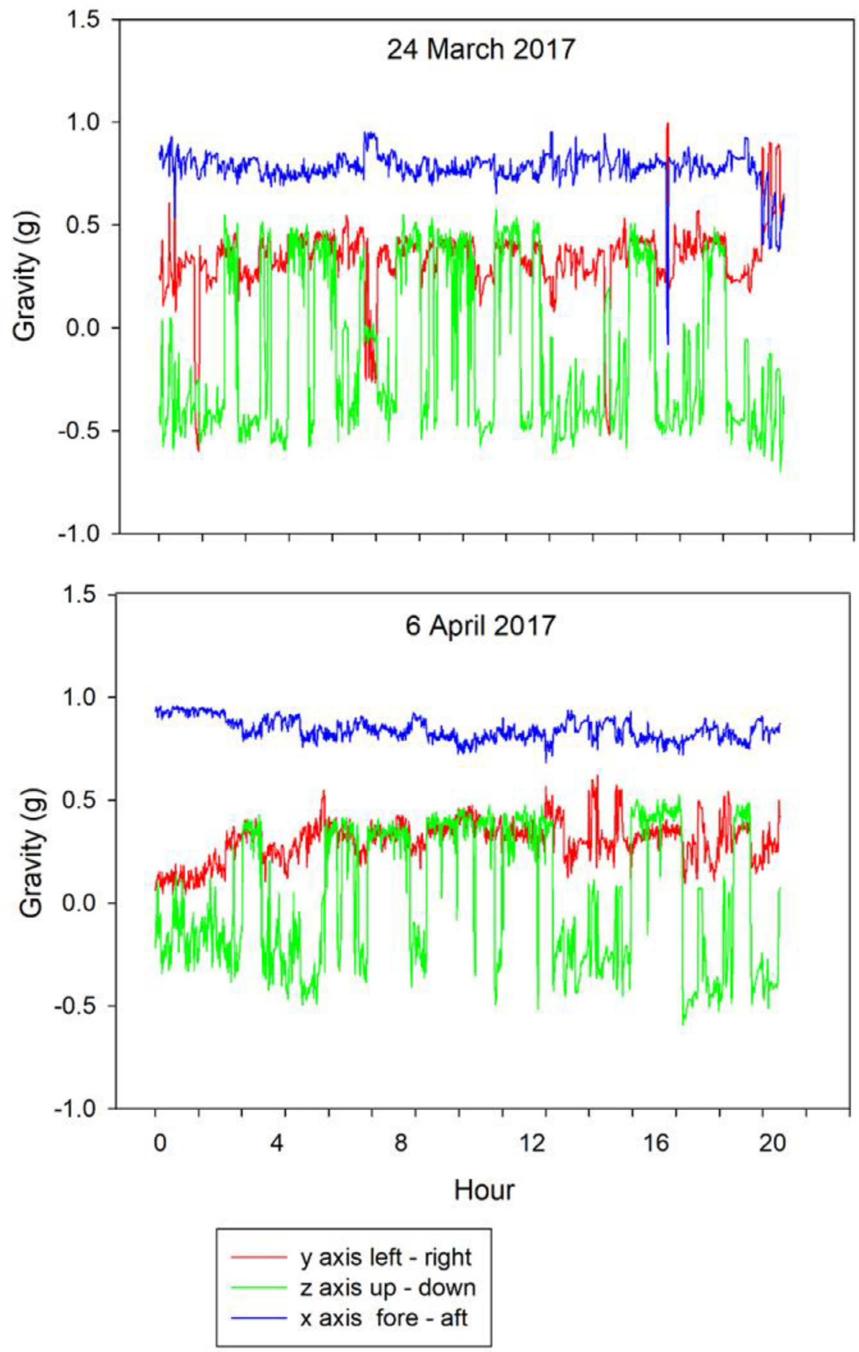

Fig. 4. Average 1-minute movement data from the x, y and z axes during the first and last 24-hour periods in study for sheep 17008 with the second highest last stagger score (LSS) of 3. Accelerometer readings are given from 0000 h to 2359 h.

in LDA negatively affects the datasets with smaller numbers of instances (in our case), even though it does not affect datasets with large numbers of instances. LDA is usually good in cases where there are thousands of features and some of them are redundant or highly related to other features [28]. In this study, we removed variables that were highly correlated before using LDA in Random Forests and SVM which likely reduced the effectiveness of using LDA.

In general, Random Forests outperforms SVM due to the power of ensemble approach. Ensemble approaches give a better result than the corresponding non-ensemble approach, because the error probability of an ensemble approach is always better than the error (if < 0.5) of its individual classifier [33]. However, there is an exception in the result of applying LDA to dataset 2 because the smaller number of instances in each class negatively affected the accuracy. Dataset 2 has the same total number of instances as dataset 1 but, dataset 2 has one more class compared with dataset 1 because the number of instances in class 0 in dataset 1 is divided into classes 0 and 1 in dataset 2. Thus, dataset 2 has less information in each class to learn and capture, to classify data, than in dataset 1. The F1-score in dataset 1 is higher than dataset 3 because a longer duration of an instance may have provided more information and improve the F1-score. Dataset 3 has more instances than dataset 1 but the duration of instances in dataset 1 is 30 times longer than those in dataset 3 which may have resulted in a higher F1-score.

To confirm our experimental result that Random Forests was the best classifier algorithm for our study, Waikato Environment for Knowledge Analysis (WEKA) [18], a well-known ML software, was applied to our datasets. WEKA provides ML algorithms for different ML tasks such as classification, clustering, and regression. We utilized the Auto-WEKA package [22] to search and identify ML algorithms and their hyperparameters to achieve the best performance. The result obtained using Auto-WEKA also showed that Random Forests was the best classifier. Even though Auto-WEKA gave the same result as our work, we do not recommend it due to the following reasons:

- The maximum depth of the tree used in Auto-WEKA could be unlimited; thus, Random Forests model in WEKA can be prone to overfitting.
- Auto-WEKA does not contain resampling methods to account for imbalanced dataset, and K-folds cross validation used in Auto-WEKA may not fit when the class distribution is not equal, specifically in our datasets.

In brief, two ML classifiers and two types of features from accelerometer data were used to classify whether a sheep had staggers. Overall, Random Forests classification provided the best performance. The result indicates that ensemble methods have potential to deal with problems that have small numbers of instances and imbalanced data.

6.2. Statistical analyses of accelerometer metrics (features) support the results of machine learning

Diurnal activity patterns monitored by accelerometer metrics (Max_z and MI) appeared to change during the latter part of the study for sheep displaying symptoms (LSS > 1) based on the strong interaction of period, diurnal time class and LSS. Activity levels indicated by MI increased for all sheep at the end of the study, but for the sheep with symptoms, the activity increase was smaller during the last few days of the study. This may be due to losses in motor function resulting from the endophyte neurotoxins. Although few sheep showed symptoms, accelerometer metrics of sheep changed from the beginning of the study compared with the end of the study. The most notable changes occurred during the daylight hours. At night, when sheep were less active, no differences in accelerometer metrics were apparent among sheep with differing LSS. The most notable change in behavior from the beginning to the end of the study occurred in the z axis that records up and down movements of the head (Fig. 3 and 4). Up and down head movements (z axis) occur while sheep are grazing, as do the other head movements (x and y axes) to a lesser degree. The timing of the largest differences in activity (midday) and the clear differences in the z-axis between the beginning and end of the study suggest that grazing behavior is affected by perennial ryegrass staggers.

Accelerometer metric levels and patterns clearly varied among sheep even at the beginning of the study (e.g., Fig. 3 and 4). The variation among sheep was likely a result of individual difference in behavior and variability in the accelerometers [34]. This variation among sheep emphasizes the need for developing algorithms that are based on changes in individual animal behavior patterns rather than groups of livestock [3]. In addition to accelerometers, other technologies such as nose band sensors may also be useful for monitoring livestock grazing endophyte infected forages, quantifying grazing time and biting rate [29, 30, 36]. Technological advancements, such as Bluetooth® (Bluetooth SIG, Inc., Kirkland, WA USA) can facilitate data transfer from these monitoring devices in real time. Monitoring of individual sheep and development of machine learning and potentially "change-point" algorithms may become a valuable method for detecting the onset of perennial ryegrass staggers and other diseases.

Use of remote monitoring systems such as accelerometers has potential to help producers detect when sheep should be moved from an endophyte-infected paddock to a paddock with a non-toxic endophyte-infected ryegrass or a paddock without endophyte infected forage. Store-on-board accelerometers used in this study store movement data on the device and do not provide managers with the data until the accelerometer is removed from the animal and downloaded from the device. However, real-time monitoring of livestock with accelerometers is becoming commercially available. Herddogg (https://herddogg.com, accessed 1 February 2022) ear tags provide an activity index based on 3-axis accelerometers every 6 minutes. The tags can transmit this activity index to a reader termed a "dog bone" via Bluetooth® technologies. The dog bone reader then transmits that data to the internet using cellular technology. With further development, near real time or real time technologies could transmit data to the internet where it could be evaluated using machine learning techniques and/or change point algorithms, and the results could be processed and interpreted using artificial intelligence technologies [17]. An alarm could then be sent to the producer that a sheep may be affected by endophyte toxins and should be evaluated.

Monitoring livestock with on-animal sensors should improve our understanding of animal behaviour in pastures with endophyte infected forages. Scasta et al. [31] found that cattle intake of alkaloids changed if tall grass prairies containing endophyte infected tall fescue were patch burned. The presence of tall fescue affects structural heterogeneity of rangelands by altering the influence of patch burning and grazing [29, 30]. Monitoring livestock grazing behaviour and the potential influence of endophyte infected forage on animal heath may improve our understanding of these systems and ability to manage them.

7. Conclusion

The application of ML classification to accelerometer data was successfully used in this case study to identify changes in behavior associated with ryegrass staggers in sheep. The results of our study show that ML is a good tool for developing algorithms to detect staggers from accelerometer data because it combines variable patterns from different metrics. Although more research is needed, the combination of ML and real-time monitoring of sheep behavior with accelerometers has potential to detect when endophyte toxin levels affect their well-being (ryegrass staggers) and the animals should be moved to a different pasture. In addition, monitoring of livestock health and grazing behaviour with accelerometers and other on-animal sensors may provide data to help manage endophyte infected pastures and facilitate practices such as patch-burn grazing.

Funding

This project was funded in part by the Harold James Family Trust. the Australian – American Fulbright Commission, and Strategic Science Investment Funding from the New Zealand Ministry of Business, Innovation and Employment.

Declaration of Competing Interest

The authors have no conflict of interest to declare.

References

[1] F.A.P. Alvarenga, I. Borges, L. Palkovič, J. Rodina, V.H. Oddy, R.C. Dobos, Using a three-axis accelerometer to identify and classify sheep behaviour at pasture, Applied Animal Behaviour Science 181 (2016) 91–99.
[2] D.W. Bailey, M.G. Trotter, C.W. Knight, M.G. Thomas, Use of GPS tracking collars and accelerometers for rangeland livestock production research, Translational Animal Science 2 (1) (2018) 81–88.
[3] D.W. Bailey, M.G. Trotter, C. Tobin, M.G. Thomas, Opportunities to apply precision livestock management on rangelands, Frontiers in Sustainable Food Systems 5 (2021) 93.
[4] J. Barwick, D. Lamb, R. Dobos, D. Schneider, M. Welch, M. Trotter, Predicting lameness in sheep activity using tri-axial acceleration signals, Animals 8 (1) (2018) 12.
[5] J. Barwick, D.W. Lamb, R. Dobos, M. Welch, M. Trotter, Categorising sheep activity using a tri-axial accelerometer, Computers and Electronics in Agriculture 145 (2018) 289–297.
[6] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.
[7] H.A. Campbell, L. Gao, O.R. Bidder, J. Hunter, C.E. Franklin, Creating a behavioural classification module for acceleration data: using a captive surrogate for difficult to observe species, Journal of Experimental Biology 216 (24) (2013) 4501–4506.
[8] N.v Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, Journal of Artificial Intelligence Research 16 (2002) 321–357.
[9] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297.

[10] M.E. di Menna, S.C. Finch, A.J. Popay, B.L. Smith, A review of the Neotyphodium lolii/Lolium perenne symbiosis and its associated effects on animal and plant health, with particular emphasis on ryegrass staggers, New Zealand Veterinary Journal 60 (6) (2012) 315–328.
[11] T.G. Dietterich, Ensemble methods in machine learning, International Workshop on Multiple Classifier Systems (2000) 1–15.
[12] R.C. Dobos, D.B. Taylor, M.G. Trotter, B.E. McCorkell, D.A. Schneider, G.N. Hinch, Characterising activities of free-ranging Merino ewes before, during and after lambing from GNSS data, Small Ruminant Research 131 (2015) 12–16.
[13] R.A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Eugenics 7 (2) (1936) 179–188.
[14] L.R. Fletcher, S.C. Finch, B.L. Sutherland, G. deNicolo, W.J. Mace, C. van Koten, D.E. Hume, The occurrence of ryegrass staggers and heat stress in sheep grazing ryegrass-endophyte associations with diverse alkaloid profiles, New Zealand Veterinary Journal 65 (5) (2017) 232–241.
[15] E.S. Fogarty, D.L. Swain, G.M. Cronin, L.E. Moraes, D.W. Bailey, M.G. Trotter, Potential for autonomous detection of lambing using global navigation satellite system technology, Animal Production Science 60 (9) (2020) 1217–1226.
[16] E.S. Fogarty, D.L. Swain, G.M. Cronin, L.E. Moraes, M. Trotter, Behaviour classification of extensively grazed sheep using machine learning, Computers and Electronics in Agriculture 169 (2020) 105175.
[17] E. Gurarie, R.D. Andrews, K.L. Laidre, A novel method for identifying behavioural changes in animal movement data, Ecology Letters 12 (5) (2009) 395–408.
[18] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, ACM SIGKDD Explorations Newsletter 11 (1) (2009) 10–18.
[19] J. Hernandez, J.A. Carrasco-Ochoa, J.F. Martínez-Trinidad, An empirical study of oversampling and undersampling for instance selection methods on imbalance datasets, Iberoamerican Congress on Pattern Recognition (2013) 262–269.
[20] R.G. Keogh, Induction and prevention of ryegrass staggers in grazing sheep, New Zealand Journal of Experimental Agriculture 1 (1) (1973) 55–57.
[21] R. Kohavi, A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, Proceedings of the 14th International Joint Conference on Artificial Intelligence 2 (1995) 1137–1143.
[22] L. Kotthoff, C. Thornton, H.H. Hoos, F. Hutter, K. Leyton-Brown, Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA, The Journal of Machine Learning Research 18 (1) (2017) 826–830.
[23] R.C. Littell, G.A. Milliken, W.W. Stroup, R.D. Wolfinger, O. Schabenberger, SAS for mixed models, SAS Institute. Inc., Cary, NC, 2006, p. 814.
[24] J. Marais, S.P. le Roux, R. Wolhuter, T. Niesler, Automatic classification of sheep behaviour using 3-axis accelerometer data, in: Proceedings of the Twenty-Fifth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA), 2014, pp. 97–102.
[25] R. Mohammed, J. Rawashdeh, M. Abdullah, Machine learning with oversampling and undersampling techniques: overview study and experimental results, in: 2020 11th International Conference on Information and Communication Systems (ICICS), 2020, pp. 243–248.
[26] C. Nord, J. Keeley, An Introduction to the HPFOREST Procedure and its Options, in: Midwest SAS Users Group Conference Proceedings; Cincinnati, OH. Paper AA20, 2016.
[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.
[28] S. Raschka, V. Mirjalili, Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow, 2nd Edition, Packt Publishing, 2017 (2nd ed.).
[29] E.J. Raynor, J.D. Derner, K.J. Soder, D.J. Augustine, Noseband sensor validation and behavioural indicators for assessing beef cattle grazing on extensive pastures, Applied Animal Behaviour Science 242 (2021) 105402.
[30] E.J. Raynor, H.L. Hillhouse, D.M. Debinski, J.R. Miller, W.H. Schacht, Recoupling cross-scale interactions in tall fescue-invaded tallgrass prairie, Landscape Ecology (2021), doi:10.1007/s10980-021-01348-x.
[31] J.D. Scasta, R.L. McCulley, D.M. Engle, D. Debinski, Patch burning tall fescue invaded grasslands alters alkaloids and tiller defoliation with implications for cattle toxicosis, Rangeland Ecology & Management 75 (2021) 130–140.
[32] S.J. Suji Prasad, M. Thangatamilan, M. Suresh, H. Panchal, C.A. Rajan, C. Sagana, B. Gunapriya, A. Sharma, T. Panchal, K.K Sadasivuni, An efficient LoRa-based smart agriculture management and monitoring system using wireless sensor networks, International Journal of Ambient Energy (2021) 1–4.
[33] P.-N. Tan, M. Steinbach, V. Kumar, Introduction to data mining, Pearson Education India, 2016.
[34] C. Tobin, D.W. Bailey, M.G. Trotter, L O'Connor, Sensor based disease detection: A case study using accelerometers to recognize symptoms of Bovine Ephemeral Fever, Computers and Electronics in Agriculture 175 (2020) 105605.
[35] A. Tzounis, N. Katsoulas, T. Bartzanas, C. Kittas, Internet of Things in agriculture, recent advances and future challenges, Biosystems Engineering 164 (2017) 31–48.
[36] J.R. Weinert, J. Werner, C.A. Williams, Validation and Implementation of an Automated Chew Sensor–Based Remote Monitoring Device as Tool for Equine Grazing Research, Journal of Equine Veterinary Science 88 (2020) 102971, doi:10.1016/j.jevs.2020.102971.