ORIGINAL RESEARCH **Open access**

Using Accelerometers and GPS Tracking to Detect Perennial Ryegrass Staggers in Sheep

Andreas Müller[1*], Stefan Weber[1], Julia Hoffmann[2], Lukas Schneider[3], Tobias Klein[4]

Abstract

Perennial ryegrass staggers (staggers) represents a neurotoxic disorder affecting livestock, triggered by the intake of ryegrass (*Lolium perenne*) harboring certain strains of *Epichloë* fungal endophytes. These grass-endophyte partnerships produce harmful toxins that can impair animal health and, occasionally, result in fatalities. In sheep, common signs include head tremors, irregular gait, muscle rigidity, and loss of balance. Affected animals usually regain normal condition after being relocated from pastures with toxic endophyte strains. A pilot investigation was conducted in Lincoln, New Zealand, to evaluate whether ryegrass staggers could be detected using GPS tracking data combined with accelerometers. 14 sheep per group grazed either a toxic endophyte-infected ryegrass paddock or an endophyte-free control paddock over 17 days in late March and early April 2017. Randomly chosen sheep received collars equipped with a 3-axis accelerometer capturing movements at 12 Hz (10 collars placed in the endophyte-infected paddock and 6 in the control paddock). Three sheep per group were further monitored at 3-minute intervals via GPS receivers. An experienced observer assessed the sheep for staggers symptoms each week and at study completion on a 0 to 5 scale. Control sheep exhibited no signs of staggers, whereas 10 sheep in the infected paddock showed minimal or zero symptoms (scores of 0 or 1). The remaining 4 sheep in the infected paddock recorded final scores between 2 and 4. Sheep in the infected pasture (2.91 m/min ± 0.04 SE) displayed slower movement (P = 0.04) compared to those in the control pasture (3.12 m/min ± 0.05 SE). Daily travel distances fluctuated, without any obvious time-based patterns emerging. Machine learning evaluation of accelerometer data revealed shifts in the behaviour of impacted sheep as the trial progressed. Activity among sheep with noticeable symptoms (scores ≥ 2) rose more substantially during morning and midday hours in the later phase relative to control animals and those with little or no symptoms (score < 2). However, nighttime behaviour for individual sheep stayed fairly steady across the entire study. Accelerometers could serve as an effective tool for remotely identifying perennial ryegrass staggers.

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

*Correspondence:
Andreas Müller
andreas.mueller@uni-heidelberg.de
[1] Department of AI for Precision Sugar Beet Harvesting, Faculty of Agriculture, Heidelberg University, Heidelberg, Germany
[2] Department of Machine Learning for Rot and Decay Detection, Faculty of Agriculture, Technical University of Munich, Munich, Germany
[3] Department of Smart Conveyor Belt Sorting, Faculty of Agriculture, University of Freiburg, Freiburg, Germany
[4] Department of Precision Yield Mapping, Faculty of Agriculture, University of Stuttgart, Stuttgart, Germany

Introduction

Sheep kept on perennial ryegrass (*Lolium perenne* L.) pastures are susceptible to developing perennial ryegrass staggers [1]. This neurotoxic syndrome affects multiple grazing livestock species and stems from toxins produced by specific strains of the asexual, symbiotic *Epichloë* fungal endophyte that colonizes ryegrass. Upon consumption of

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

contaminated forage, symptoms emerge slowly over time. Typically, recovery occurs when livestock are shifted to endophyte-free pastures or to pastures containing specially selected non-toxic endophyte varieties.

Close observation of sheep and other livestock on endophyte-infected pastures is vital for protecting animal welfare. Moreover, on-animal sensing technologies could assist in exploring the complex dynamics of grazing practices, controlled burning, toxin accumulation, and animal responses within rangelands dominated by or composed primarily of endophyte-infected vegetation [2-4], including tall fescue (*Schedonorus arundinaceus* [Schreb.] Dumort. nom. cons.). Traditional visual checks or routine stagger evaluations demand considerable time and effort, particularly in sizable flocks or expansive grazing systems. On-animal sensors hold promise for distant surveillance and recognition of behavioural alterations in animals [5]. These expressed behaviours can help pinpoint when livestock start to experience health issues. This paper examines the idea by deploying sensors on sheep feeding in endophyte-infected perennial ryegrass.

Livestock well-being has been assessed remotely in diverse conditions using GPS tracking and accelerometers. Lambing identification has been achieved using GPS tracking [6, 7] and accelerometers [8]. Machine learning (ML) techniques allow classification of sheep behaviour from accelerometer readings [8]. Accelerometers have proven effective at detecting illnesses such as bovine ephemeral fever in cattle [9]. Emerging advancements now support real-time or near-real-time transmission of GPS and accelerometer information via systems such as the Internet of Things, LoRaWAN [10, 11], and satellite-based solutions. Such immediate monitoring enables farmers and ranchers to respond more quickly to sick animals.

This paper aims to deliver a “proof of concept study” assessing the capability of remote monitoring with GPS tracking and accelerometers to identify perennial ryegrass staggers in sheep. Classification ML algorithms are applied to the collected accelerometer data. The research proposes that data from accelerometers and GPS devices can effectively reveal behavioural shifts in sheep experiencing perennial ryegrass staggers.

Materials and Methods

Study site

This investigation was carried out at AgResearch Lincoln, New Zealand, between 23 March and 10 April 2017. Two experimental paddocks, each 0.1625 ha in size, were utilized; both had been planted as pure stands of ‘GA66’ diploid perennial ryegrass during October 2016. One paddock featured endophyte-infected perennial ryegrass carrying a ‘standard’ (also termed wild-type or common toxic) strain of *Epichloë festucae* var. *lolii*. This standard strain is documented to provoke ryegrass staggers in sheep and additional livestock [1]. The other paddock acted as the control and consisted of ‘GA66’ perennial ryegrass lacking any endophyte.

Animals

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

In total, 28 crossbred female sheep at 18 months of age were randomly divided between the two paddocks (14 sheep each). At the outset of the trial, the sheep had an average live weight of 56.3 kg ± 0.6 SE.

Every one of the 14 sheep in the endophyte-infected paddock, along with eight out of the 14 in the control paddock, wore Gulf Coast X-16-4 Accelerometers (Gulf Coast Data Concepts, LLC, Waveland, MS, USA). These devices were secured on neck collars and captured motion data across the x, y, and z axes. Head motion forward and backward aligned with the x axis, lateral movements with the y axis, and vertical actions with the z axis. Data on movements were captured at 12 Hz and saved internally on the devices for later collection upon study completion. Three sheep from each paddock carried igotU 600 GPS receivers (MobileAction, New Taipei City, Taiwan) on their collars. Location fixes were logged every 3

© 2025 The Author(s). | Published by Agro Intelligence Research Press

minutes. One GPS unit in the control paddock produced no location data for the full duration, presumably due to battery issues. Positions logged beyond paddock limits were excluded from the dataset. ArcMap, a geographic information system (GIS) software (ESRI ArcGIS Suite, Redlands, California, USA), helped detect any such out-of-boundary recordings.

Study design

The trial adhered to established protocols, adapted over many years for testing new grass-endophyte combinations, as detailed by Fletcher *et al.* [12]. Sheep grazed the control and endophyte paddocks throughout autumn 2017, specifically from 23 March to 10 April. Stagger symptoms were rated on a 0 to 5 scale [13], with 0 representing absence of symptoms and 5 indicating strong tremors triggered by slight disturbance or physical activity. Sheep were herded from their assigned paddock and along the connecting raceway for a brief, moderate run. Assessments occurred on 30 March, 3 April, and 10 April. Most sheep (21 of 28) received an initial weighing, and all were weighed again at the conclusion of the study.

GPS Data Analysis

Hourly distance covered (m/hour) was obtained through the summation of distances separating all recorded positions over the course of one hour. The average hourly travel speed (m/min) was derived from the mean of hourly GPS-recorded velocity measurements. Daily distance covered (m/day) resulted from summing the distances between all positions logged throughout a complete day (24 h). The average daily travel speed (m/min) was calculated by averaging all velocity values captured by the GPS over every 24 hours. Distinct statistical analyses of the GPS information were performed for the hourly and daily records separately.

Hourly measurements were assessed using the repeated-measures function in PROC MIXED in SAS (SAS Institute Inc., Cary, NC, USA; [14]). Model fixed effects comprised treatment (endophyte or control), day, hour, together with the interactions involving treatment × day, treatment × hour, and treatment × hour × day. Nevertheless, neither the two-way nor three-way interactions reached statistical significance ($P > 0.05$). These were therefore eliminated from the ultimate model. The simplified final model retained only treatment, day, and hour as fixed effects. Sheep served as the experimental units for the repeated-measures design. An autoregressive order of 1 [AR(1)] covariance structure was applied to model the repeated measures covariance, selected for possessing the smallest Akaike's Information Criteria (AIC) value out of the three options examined: AR(1), compound symmetry, and unstructured [14].

Daily measurements received a similar assessment using the repeated-measures function in PROC MIXED of SAS [14]. Fixed effects within this model included treatment (endophyte or control), day, plus the treatment × day interaction. Sheep served as the subject in the repeated-measures design. In line with the hourly evaluation, an AR(1) model was adopted due to delivering a lower AIC value than the alternative structures reviewed, namely compound symmetry and unstructured [14].

Accelerometer Data Analyses

Data preprocessing

Raw accelerometer readings gathered between 0600 h and 1800 h on the three days of stagger scoring, along with one additional day preceding the final sensor entry, were pulled out and split at the outset into 1-minute epochs. In the case of a single sheep, the device halted operation shortly before the ultimate stagger evaluation, leaving information unavailable for the day of the third stagger assessment. Daytime periods (0600 h to 1800 h) provided accelerometer records, as nighttime sheep movement tends to be lower than during daylight hours and has lower analytical value (see results below). The 1-minute epochs were subsequently averaged to generate 30-minute blocks suitable for machine-learning categorization. Each 30-minute block was formed by combining 30 complete 1-minute epochs, each free of missing values. An epoch qualifies as intact if it retains all timestamps, provided it contains 700 or more; otherwise, it counts as a

© 2025 The Author(s). | Published by Agro Intelligence Research Press

missed-timestamp epoch. Solely epochs containing no absent timestamps entered the experiment. Matching of stagger scores to epochs relied upon corresponding timestamps. Records originating from sheep located in the control paddock were excluded from all machine learning procedures.

Metrics description

Nineteen metrics sourced from Fogarty *et al.* [8] were computed for every 1-minute epoch. Given that Ax, Ay, Az, and Axyz exhibited complete collinearity among themselves according to the Variance Inflation Factor within this dataset, evaluation proceeded using only the average across the three axes (Axyz), omitting Ax, Ay, and Az entirely. Consequently, the sixteen metrics outlined in **Table 1** were derived for each 1-minute epoch. In datasets 1 and 2, averaging merged 1-minute epochs 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.

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

Feature	Equation
Average all-axis (*Axyz*)	$\begin{aligned}&\mathrm{Axyz}\\&=\frac{1}{\mathrm{T}}\sum_{1}^{\mathrm{T}}(\mathrm{x}(\mathrm{t})\\&+\mathrm{y}(\mathrm{t})\\&+\mathrm{z}(\mathrm{t}))\end{aligned}$
Minimum X-axis (*Minx*)	The minimum X-axis value in the epoch
Minimum Y-axis (*Miny*)	The minimum Y-axis value in the epoch
Minimum Z-axis (*Minz*)	The minimum Z-axis value in the epoch
Maximum X-axis (*Maxx*)	The maximum X-axis value in the epoch
Maximum Y-axis (*Maxy*)	The maximum Y-axis value in the epoch
Maximum Z-axis (*Maxz*)	The maximum Z-axis value in the epoch
Standard Deviation X-axis (*SDx*)	$\begin{aligned}&\mathrm{SDx}\\&=\sqrt{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}}\left(\mathrm{x}(\mathrm{t})-\overline{\mathrm{x}}\right)^{2}}\end{aligned}$ where $\bar{x}$ is the mean of the X-axis value in the epoch

© 2025 The Author(s). | Published by Agro Intelligence Research Press

Standard Deviation Y-axis (*SDy*)	$$\mathrm{SDy} = \sqrt{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}}\left(Y\left(\mathrm{t}\right)-\overline{\mathrm{y}}\right)^{2}}$$ where $\overline{y}$ is the mean of the Y-axis value in the epoch
Standard Deviation Z-axis (*SDz*)	$$SDz = \sqrt{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}}\left(Z\left(\mathrm{t}\right)-\overline{\mathrm{Z}}\right)^{2}}$$ where $\overline{z}$ is the mean of the Z-axis value in the epoch
Average Standard Deviation all-axis (*SDxyz*)	$$VARxyz = \frac{1}{3*T}*\left(SD_{x}^{2}*T + SD_{y}^{2}*T + SD_{z}^{2}*T\right)$$ $$SDxyz = \sqrt{VARxyz}$$
Movement Intensity (MI)	$$MI = \frac{1}{T}\sum_{t=1}^{T}\sqrt{x\left(t\right)^{2}+y\left(t\right)^{2}+z\left(t\right)^{2}}$$
Signal Magnitude Area (SMA)	$$SMA = \frac{1}{T}\sum_{t=1}^{T}\left(\left\|x\left(t\right)\right\| + \left\|y\left(t\right)\right\| + \left\|z\left(t\right)\right\|\right)$$
Energy (Energy)	$$Energy = \frac{1}{T}\sum_{t=1}^{T}\left(x\left(t\right)^{2}+y\left(t\right)^{2}+z\left(t\right)^{2}\right)^{2}$$
Entropy (Entropy)	$$Entropy = \frac{1}{T}\sum_{t=1}^{T}\left(1 + \left(x\left(t\right)+y\left(t\right)+z\left(t\right)\right)^{2}\right)*ln\left(1 + \left(x\left(t\right)+y\left(t\right)+z\left(t\right)\right)^{2}\right)$$
Movement Variation (MV)	$$MV = \frac{1}{T}\sum_{t=2}^{T}\left(\left\|x\left(t-1\right)-x\left(t\right)\right\| + \left\|y\left(t-1\right)-y\left(t\right)\right\| + \left\|z\left(t-1\right)-z\left(t\right)\right\|\right)$$

Dataset 3 is much larger than datasets 1 and 2 because the instances are much shorter in dataset 3, and there are more classes **(Table 2)**.

Dataset 3 is much larger than datasets 1 and 2 because the instances are much shorter in dataset 3, and there are more classes **(Table 2)**.

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

Empty Cell	**The total number of instances**				
Dataset	**Class 4**	**Class 3**	**Class 2**	**Class 1**	**Class 0**

© 2025 The Author(s). | Published by Agro Intelligence Research Press

1				47	970
2			47	535	435
3	675	754	4386	11745	13115

Machine learning of accelerometer data

Assigning categories to perennial ryegrass staggers based on accelerometer readings is a machine learning (ML) task. In conventional classification setups, the 30-minute instances from datasets 1 and 2, or the 1-minute epochs of dataset 3 **(Table 2)**, serve as the core data instances or objects. Stagger scores, indicative of illness severity, serve as the class or target variable for each data instance or object. Metrics capturing sheep locomotion patterns **(Table 1)** serve as features or attributes. Individual data instances incorporate various such features. The present research employs ML techniques to assign each data instance to an appropriate class, indicating whether a sheep has staggers.

Beyond complications arising from missing data points, a primary obstacle to deploying ML for this categorization is the data's unbalanced nature — marked variation in instance counts across the stagger versus no-stagger groups, specifically in dataset 1 **(Table 2)**. Such imbalance frequently undermines the effectiveness of classification algorithms, resulting in notably lower accuracy for the underrepresented group (the minority class). To illustrate, dataset 1 contains far fewer instances of the stagger group than of the non-stagger group, making stagger the minority class. Greater emphasis lies on correctly recognizing stagger instances over no-stagger ones; accordingly, the stagger class carries higher priority than the non-stagger class for dataset 1.

Model selection and resampling method

To address unequal class sizes, the primary strategy adopted is stratified cross-validation [15]. During the construction of classification models, the data is split into training and test sets. The training portion trains the ML algorithms to build a model. Following model development, the testing phase assesses how well the ML algorithms perform by computing an evaluation metric (e.g., accuracy, F1 score). This dataset splitting is done at random. Still, situations can arise in which the minority class is missing from the training set but present in the test set. When that happens, the model cannot correctly predict outcomes on the test data, leading to a poor evaluation metric. The reason is that the training phase leaves the model without exposure to or understanding of the minority class patterns, preventing it from properly mirroring the real population in the dataset. Stratified sampling effectively addresses imbalance by ensuring that the proportion of instances in each class remains constant across all resulting subsets. To improve model stability, k-fold cross-validation is used, splitting the full dataset into k folds. In k-fold cross-validation, k-1 folds are used for training, and the remaining fold is used for testing. Here, the experiment sets k = 10. When combined with stratified sampling, each fold maintains class proportions identical to those found in the complete dataset.

As a further step to correct class imbalance in a dataset, both oversampling and under-sampling strategies have been used across multiple machine learning settings [16-18]. In the present work, the Synthetic Minority Over-sampling Technique (SMOTE) [19] was applied, demonstrating its effectiveness in handling imbalanced datasets. SMOTE is an oversampling technique that increases the minority class instance count until it reaches parity with the majority class. Instead of merely copying existing minority instances, SMOTE generates new minority instances by creating "synthetic" samples. Employing SMOTE produces a training set with improved balance and greater diversity.

Features

Two distinct feature types were created to help classify possible stagger scores. The initial kind involves domain features **(Table 1)** obtained immediately from the accelerometer readings. Moreover, well-known feature extraction approaches such as Principal Component Analysis and Linear Discriminant Analysis (LDA) can raise the standard of domain features [20]. This study used LDA [21] to generate the second type of feature. LDA is a supervised technique that reduces dimensionality by mapping data points from a high-dimensional space into a lower-dimensional feature space. LDA works to both widen the gaps separating the various classes and shrink the gaps among instances within a class. In LDA, the

© 2025 The Author(s). | Published by Agro Intelligence Research Press

maximum number of feature projections is c-1, where c is the number of classes in the dataset. For that reason, the feature quantities here were arranged as 1, 2, and 4, matching the upper limits suitable for datasets 1, 2, and 3, respectively (since the class counts are 2, 3, and 5 for datasets 1, 2, and 3, respectively). Applying LDA reduces the total number of features from 16 domain features to 1, 2, and 4 in datasets 1, 2, and 3, respectively.

Machine learning classifiers

A wide range of machine learning classification algorithms has emerged, including Decision Trees, Random Forests, Support Vector Machines (SVMs), and k-nearest neighbors. The current research uses the ensemble approach [22], which integrates several classifiers to create a stronger predictive classifier than any single classifier could achieve on its own (known as the non-ensemble approach). Among ensemble techniques, Random Forests [23] stand out as one of the most common and robust algorithms, having delivered exceptional predictive results on imbalanced datasets, which led to their selection in this case. Random Forests are built from numerous decision trees. Every single tree delivers a class prediction, and the class gathering the largest vote total from all the decision trees forms the ultimate model prediction. In addition, to contrast the performance of ensemble versus non-ensemble methods, SVM [24] was chosen as a representative non-ensemble method, thanks to its solid track record as a supervised ML algorithm for classification in published studies and its earlier application to sheep behavior classification [8]. SVMs calculate a hyperplane (decision boundary) to partition the data. The model picks the ideal hyperplane that maximizes the margin between the closest training samples. Besides that, when datasets are not linearly separable, SVM uses kernels to perform non-linear classification. Kernel SVM classification proceeds as follows: non-linear data are first mapped to a higher-dimensional space where linear separation is possible, and then a linear SVM performs classification in that space. To check if kernel SVM would be required for this research, least squares error calculations came from a linear model produced through linear regression. Findings showed an extremely low value, indicating the dataset possessed non-linear properties. Therefore, the kernel SVM received application. To reduce the heavy computational demands of feature transformation in the new space, the Radial Basis Function (RBF) kernel was chosen for the kernel SVM. Every algorithm mentioned above was implemented using the scikit-learn library in Python [21].

Machine Learning approach evaluation

Several basic evaluation metrics have been recommended for assessing classification model performance, including accuracy, precision, recall, and F1-score. Given the unequal class distribution in this research, the F1-score is an appropriate metric for evaluating the model [18]. F1-score gets calculated according to the following equation:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (1)$$

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

In the preceding equations, true positive (TP) means the count of instances that models correctly label as stagger classes; false positive (FP) means the count of instances that models wrongly label as stagger classes; false negative (FN) means the count of instances that models wrongly label as non-stagger classes.

Repeated measures statistical analyses of accelerometer data

The four most significant features identified by the Random Forests algorithm in scikit-learn, based on Gini impurity, are Minimum Z-axis (Minz), Maximum Z-axis (Maxz), Average of all axes (Axyz), and Movement Intensity (MI). To pursue a different collection of highly relevant features, PROC HPForest in SAS [25] was used, focusing on out-of-bag (OOB) error when selecting the most meaningful classification variables. The four most significant features generated by the HPForest algorithm in SAS, based on OOB Gini values, were Maximum Z-axis (Maxz), Movement Intensity (MI), Maximum X-axis (Maxx), and Energy (Energy).

© 2025 The Author(s). | Published by Agro Intelligence Research Press

Independent statistical examination focused on the two overlapping, important features derived from both classification processes (scikit-learn [21] and SAS): Maxz and MI. The maximum value of the Z axis, along with MI, served as dependent variables and were examined using repeated-measures PROC MIXED in SAS [14]. Hourly averages of Maxz and MI drawn from every sheep carrying accelerometers for the full study duration were entered into the examination. Fixed effects inside the model covered 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—monitored sheep served as the subject for the repeated-measures analyses. Modelling of repeated measures covariance relied on the compound symmetry structure, selected because it yielded the smallest AIC value among the AR(1), compound symmetry, and unstructured covariance structures reviewed [14].

Results and Discussion

Sheep health

The bulk of the sheep failed to show intense ryegrass staggers symptoms. Control sheep remained entirely free of symptoms. Average final stagger score reached 1.14 (range 0 to 4) inside the endophyte-infected paddock. Control sheep (-2.14 kg ± 0.66 SE) suffered larger ($P = 0.002$) weight reductions than those located in the endophyte paddock (0.78 kg ± 0.42 SE).

GPS data

Daily examination of distance travelled per day revealed that control sheep (4403 m/day ± 93 SE) displayed a trend ($P = 0.080$) toward greater daily coverage relative to sheep inside the endophyte-infected paddock (4089 m/day ± 76 SE). Control sheep (3.08 m/min ± 0.06 SE) also trended ($P = 0.075$) toward elevated movement speeds (higher velocity) compared with sheep in the endophyte paddock (2.88 m/min ± 0.05 SE). Both distance covered and velocity showed differences ($P < 0.05$) from one study day to another. Absence of any interaction between treatment (endophyte vs control) and days appeared ($P > 0.10$) regarding either distance or velocity.

Hourly examination found no distinctions in travel (m/hour) separating sheep from control versus endophyte paddocks ($P = 0.17$). Travel rates per hour varied by time of day ($P < 0.001$). Travel per hour showed no differences ($P = 0.22$) from day to day, and interactions linking day with treatment ($P = 0.57$) or hour with treatment ($P = 0.43$) stayed undetected. Control sheep velocity (3.12 m/min ± 0.05 SE) surpassed ($P = 0.049$) the velocity recorded for sheep in the endophyte paddock (2.91 m/min ± 0.04 SE). Velocity levels differed ($P = 0.044$) between days while also differing ($P < 0.001$) between hours. Interactions involving day and treatment ($P = 0.51$) or hour and treatment ($P = 0.31$) remained absent for velocity.

Accelerometer data

Machine Learning analyses

Effectiveness levels for SVM and Random Forests classification based on domain features are shown in **Table 3**. Domain features enabled Random Forests to deliver markedly superior results compared with SVM **(Table 3)**. Top achievement (F1-score of 0.965) emerged from dataset 1 paired with Random Forests.

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.

Empty cell	Random forests	SVM

© 2025 The Author(s). | Published by Agro Intelligence Research Press

Dataset 1	**0.965**	0.887
Dataset 2	0.814	0.640
Dataset 3	0.871	0.646

SVMs combined with Random Forests using LDA delivered weaker outcomes than those with domain features **(Tables 3 and 4)**. Random Forests produced the strongest overall outcomes **(Table 4)**. Datasets 1 and 3 displayed LDA result patterns that resembled domain features **(Table 3)**, where Random Forests generated higher F1-scores than SVM, yet dataset 2 showed Random Forests performing marginally worse than SVM. Primary causes relate to differences in feature quantities and instance volumes present inside these datasets. An extended discussion of the outcomes and discrepancies is in Section 6.1.

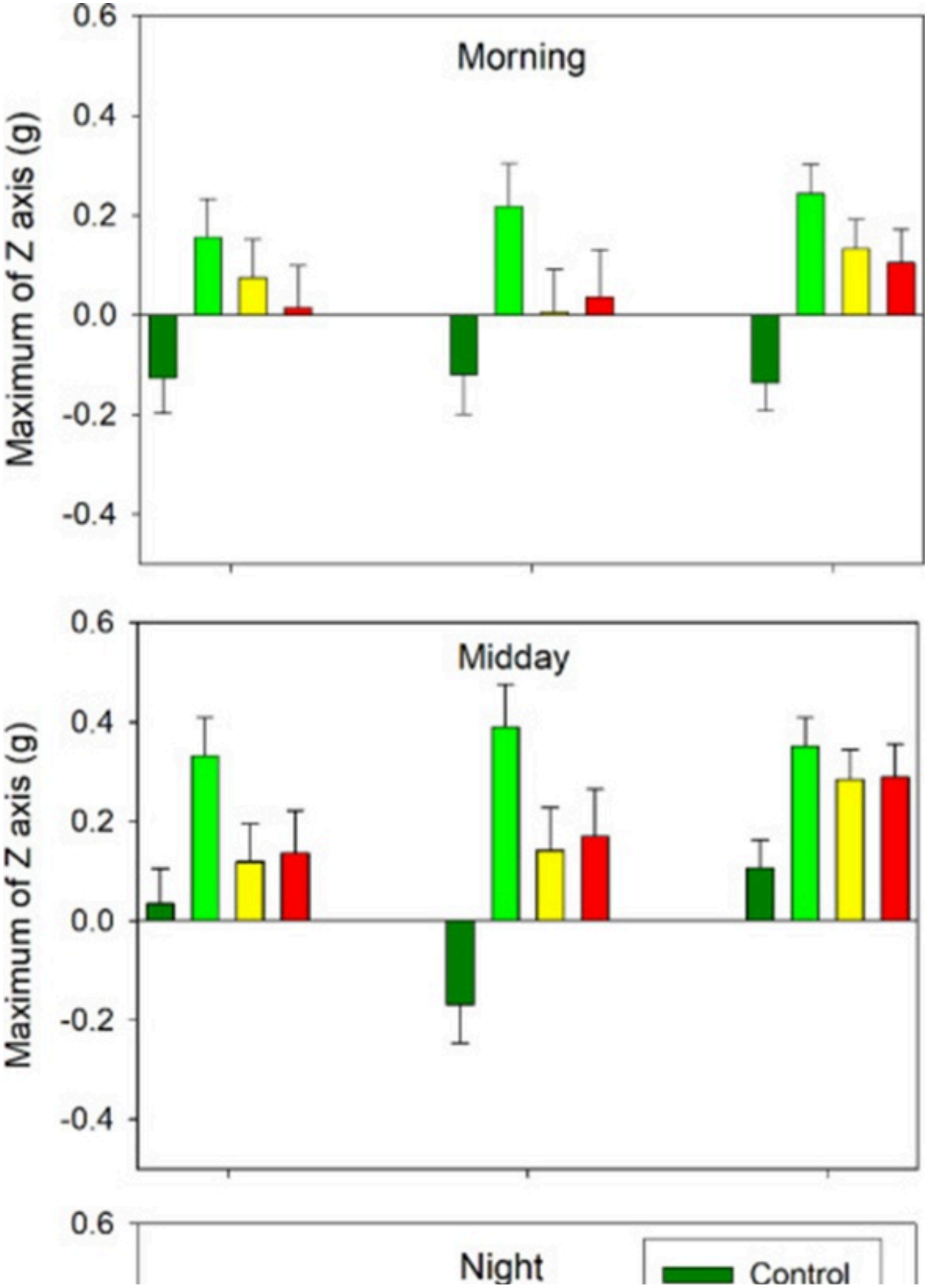
Table 4. F1-score for accuracy when predicting perennial ryegrass staggers scores when applying Single Vector Machines (SVM) and Random Forests with Linear Discriminant Analysis (LDA) in three datasets. Bold indicates the highest F1 score.

Empty cell	Random forests	SVM
Dataset 1	**0.836**	0.813
Dataset 2	0.615	0.630
Dataset 3	0.669	0.596

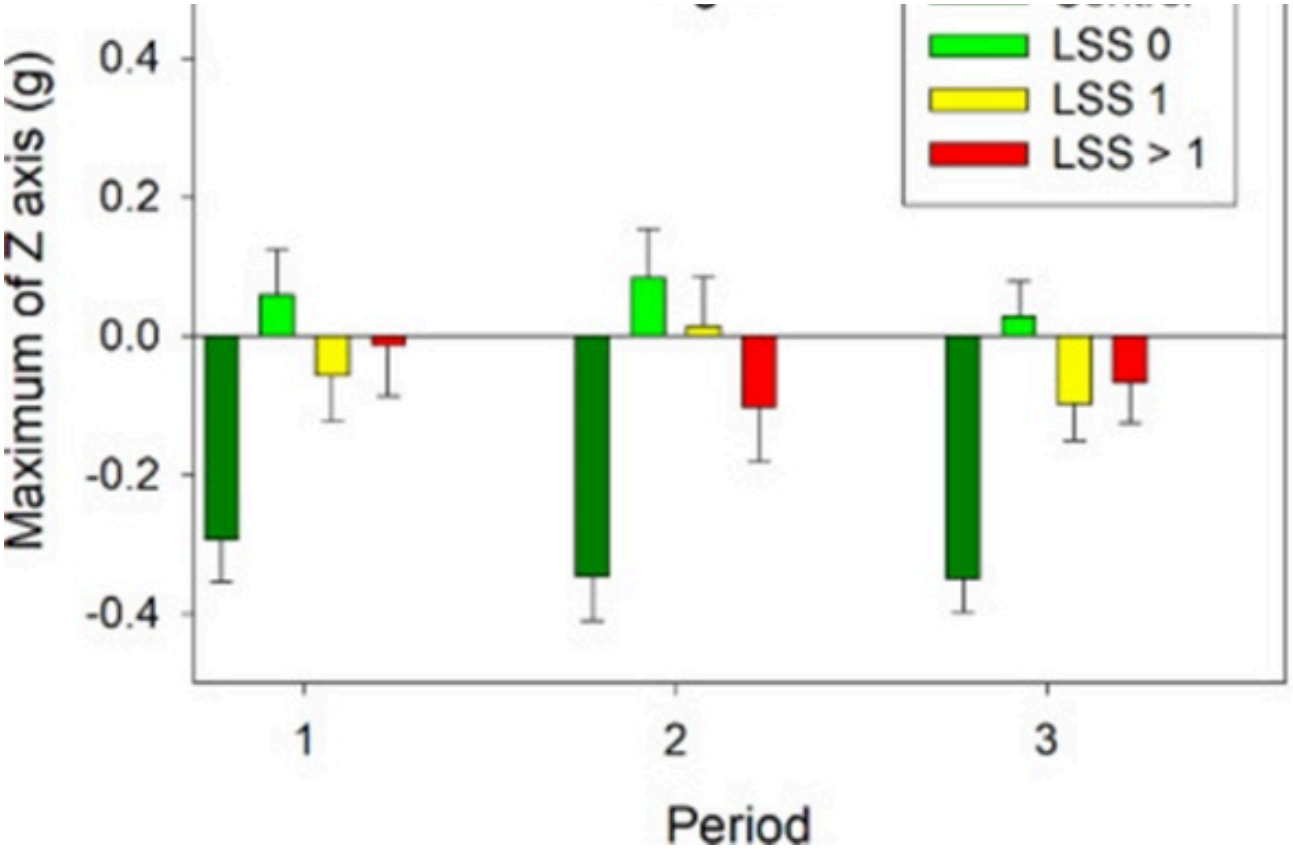
Repeated measures analyses of accelerometer data

Examination revealed no significant effects from LSS ($P = 0.66$) on Maxz. The Maximum Z-axis showed substantial variation across periods ($P < 0.0001$) and across days nested inside periods ($P = 0.0024$). Maximum Z-axis values also differed markedly ($P < 0.0001$) between the diurnal time classes (morning, midday, night) and showed clear differences ($P < 0.0001$) across hours within each class. A robust interaction emerged ($P < 0.0001$) connecting LSS to diurnal time class, along with an interaction between LSS and period ($P = 0.0092$). Furthermore, a highly significant 3-way interaction among LSS, period, and diurnal time class ($P < 0.0001$) was observed **(Figure 1)**.

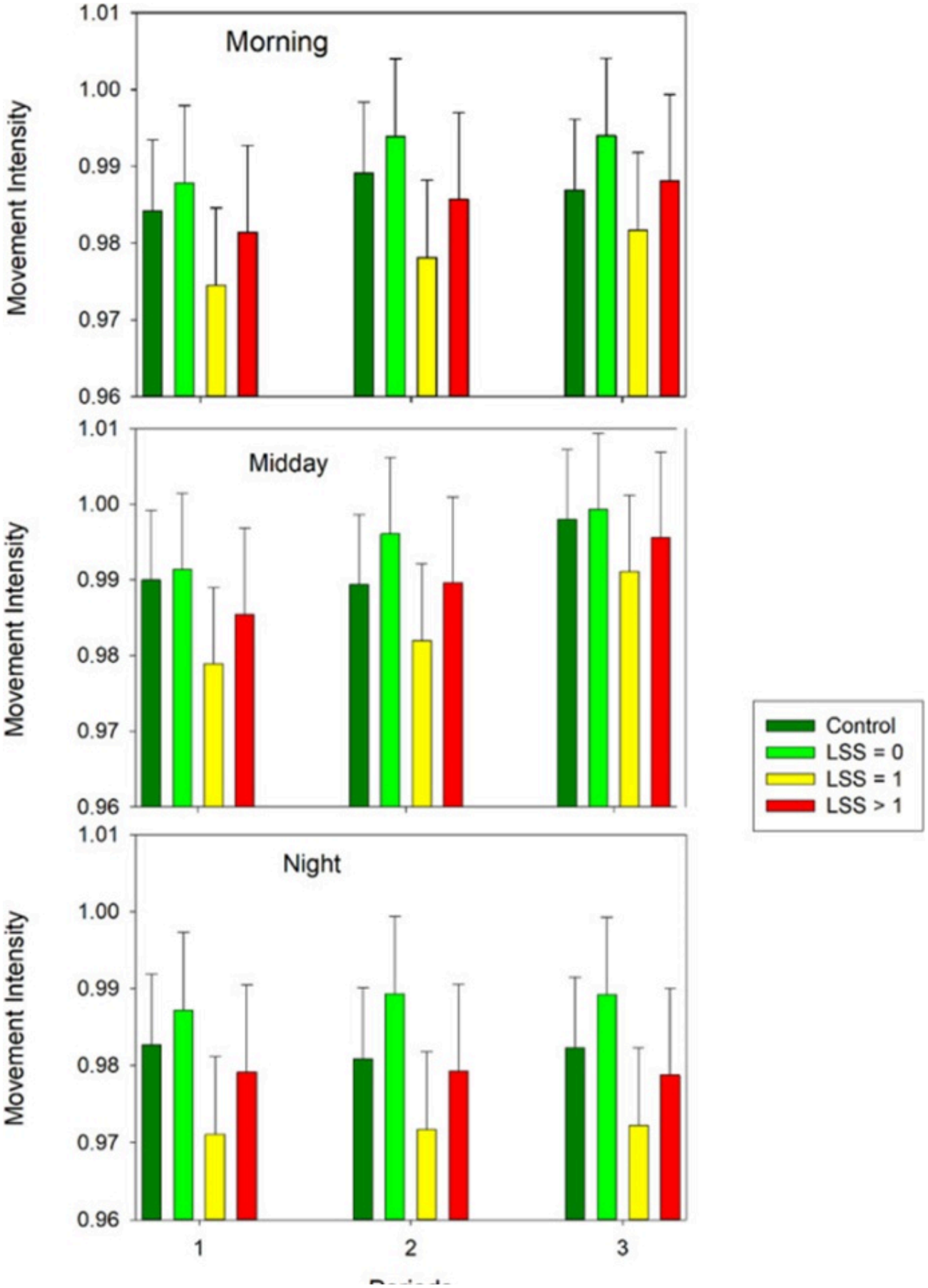
© 2025 The Author(s). | Published by Agro Intelligence Research Press



© 2025 The Author(s). | Published by Agro Intelligence Research Press



© 2025 The Author(s). | Published by Agro Intelligence Research Press



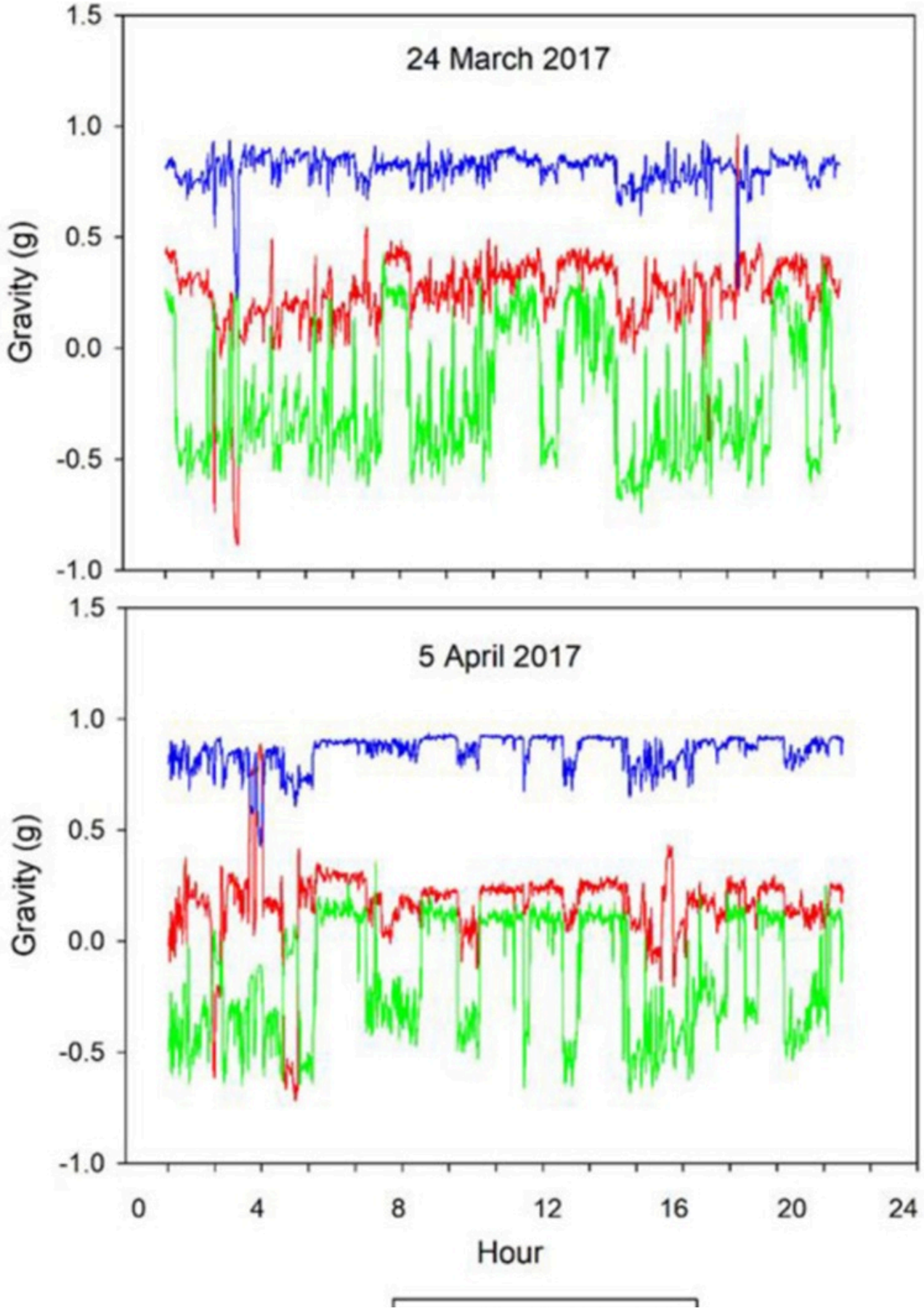
© 2025 The Author(s). | Published by Agro Intelligence Research Press



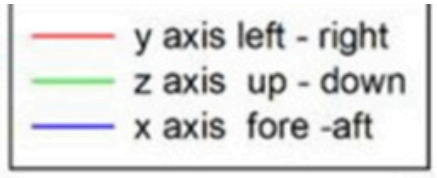
Figure 2. Mean movement intensity (MI) by period during the morning (0500 to 1000), midday (1000 to 1700), and night (1700 to 0500). Means and standard errors (error bars) are provided for the control pasture (C) and last stagger scores (LSS) 0, 1, and > 1 in the endophyte-infected pasture during period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

Daily activity rhythms for two sheep that eventually developed moderate to severe ryegrass stagger symptoms (LSS 3 and 4) by study completion differed noticeably from their patterns observed at the start of the trial, when stagger scores were 0 **(Figures 3 and 4)**. Intervals showing relatively stable readings near zero (most likely indicating periods of inactivity) on the y-axis and particularly on the z-axis lasted longer toward the conclusion of the study than they did initially. Although shifts in activity rhythms appeared across all 3 accelerometer axes, the y-axis (corresponding to side-to-side head motions) and, above all, the z-axis (corresponding to up-and-down head motions) exhibited the most prominent alterations.

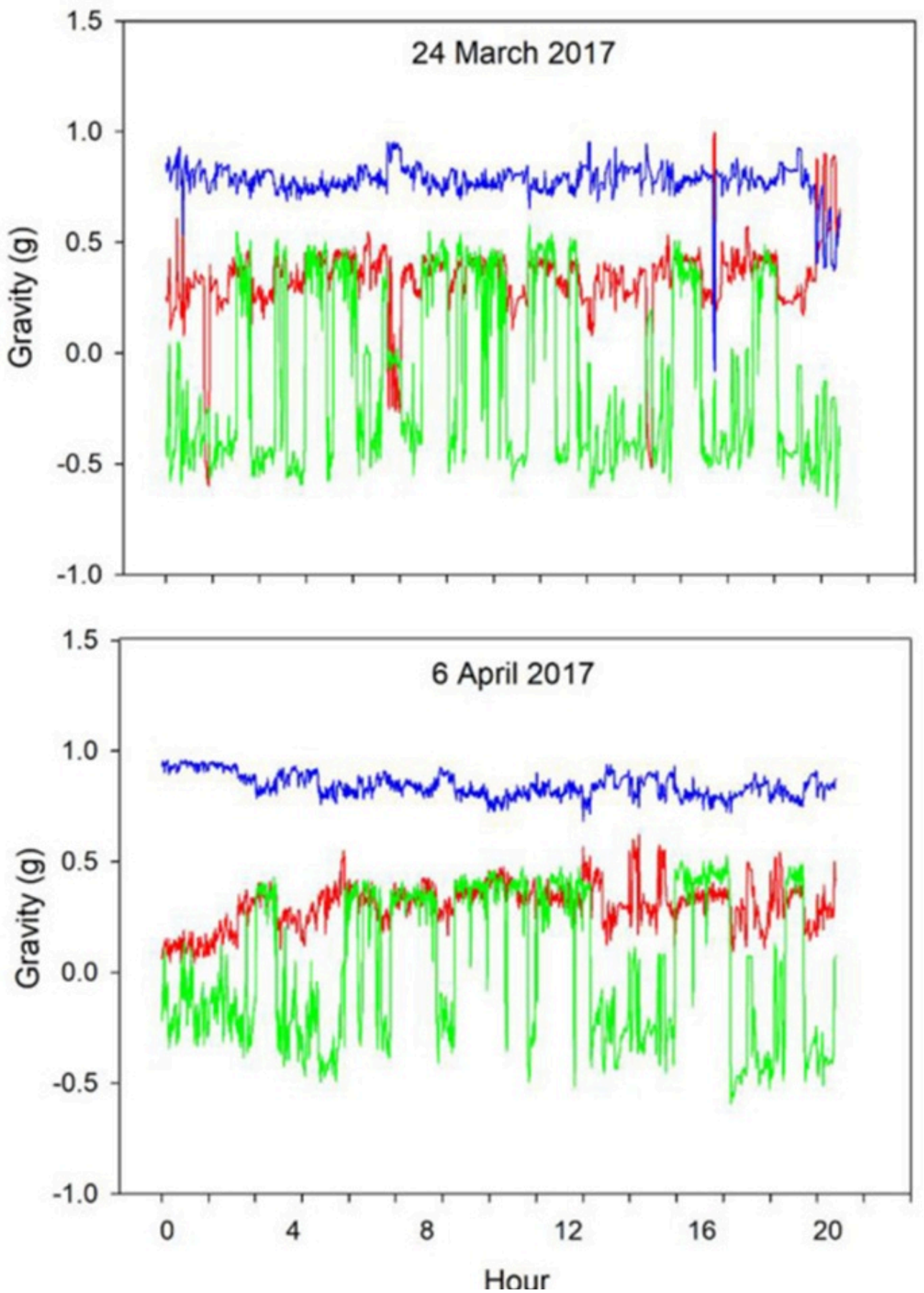
© 2025 The Author(s). | Published by Agro Intelligence Research Press



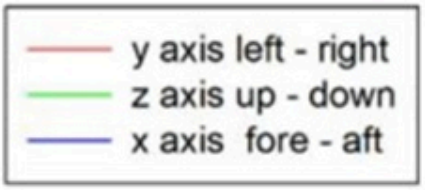
© 2025 The Author(s). | Published by Agro Intelligence Research Press



© 2025 The Author(s). | Published by Agro Intelligence Research Press



© 2025 The Author(s). | Published by Agro Intelligence Research Press



notion that ryegrass symptoms remained quite mild overall. Conditions stayed cool for much of the trial, a factor that could have decreased toxin concentrations in the herbage compared with typical late-summer peaks at this experimental site [12]. While some indications suggested that sheep in the endophyte paddock advanced at reduced speeds and covered modestly shorter daily total distances than control sheep, GPS monitoring does not emerge as a practical tool for stagger detection. No regular trends surfaced in the day-to-day velocity fluctuations. Neither distance travelled nor velocity demonstrated systematic upward or downward trends over the course of the trial, something that might have been anticipated if GPS data had flagged ryegrass staggers. Instead, day-to-day travel differences followed an unpredictable pattern. The paddocks used proved notably compact in this investigation (0.1625 ha), thereby constraining sheep from exhibiting their complete spectrum of spatial behaviors. Any observed travel discrepancies between the endophyte-infected and control paddocks most likely arose as a consequence of natural variation among individual sheep or from distinctions in forage attributes between the two areas—nevertheless, quantities of available forage registered as broadly equivalent in both paddocks.

Machine learning of accelerometer data explained a large amount of the stagger scores

LDA-created features fell short of the results achieved by domain features when tested with both Random Forests and SVM. Domain features involving accelerometer readings were modified and merged by Fogarty *et al.* [8] using formulas sourced from Alvarenga *et al.* [26], Barwick *et al.* [27], Barwick *et al.* [28], Campbell *et al.* [29], and Marais *et al.* [30] specifically for sheep behavior identification. These modifications yielded strong results in the current work. On the other hand, LDA here transformed the data points onto perpendicular axes, without aiding classification, given that the original features already had clear predictive value. The dimensionality reduction performed by LDA created drawbacks for datasets like ours that contain fewer instances, although it causes no issues for those with very large instance volumes. LDA tends to shine in scenarios with thousands of features, where redundancy or tight interrelations among some of them exist [31]. Because highly correlated variables were already excluded ahead of applying LDA within Random Forests and SVM in this project, the advantages normally expected from LDA were likely curtailed.

Overall, Random Forests deliver stronger results than SVM thanks to the advantages of ensemble learning. Ensemble strategies tend to produce superior outcomes compared to single non-ensemble models, as their error rates remain lower than those of any individual component classifier (when under 0.5) [18]. One exception, however, occurred with LDA applied to dataset 2, where the limited number of instances per class reduced classification accuracy. Although dataset 2 matches dataset 1 in overall instance count, it includes an extra class, since instances from class 0 in dataset 1 are split into classes 0 and 1 in dataset 2. As a result, dataset 2 provides less learning material per class to capture the patterns needed for classification than dataset 1 does. Dataset 1 achieved a higher F1-score than dataset 3, mainly because its longer instance duration likely provided richer details, thereby improving the F1-score. Dataset 3 has more instances than dataset 1, yet each instance in dataset 1 lasts 30 times longer, contributing to its better F1-score.

To verify that Random Forests was the top classifier in our experiments, the Waikato Environment for Knowledge Analysis (WEKA) [32], a popular ML platform, was run on the same datasets. WEKA offers algorithms for multiple ML tasks, including classification, clustering, and regression. We used the Auto-WEKA package [33] to automatically search for suitable algorithms and optimize their hyperparameters to achieve peak performance. Auto-WEKA likewise identified Random Forests as the leading classifier. Although it matched our earlier conclusion, Auto-WEKA is not suggested for the following reasons:

© 2025 The Author(s). | Published by Agro Intelligence Research Press

• Tree depth in Auto-WEKA has no upper restriction; therefore, the Random Forests implementation in WEKA carries a higher risk of overfitting.

• Auto-WEKA omits specialized resampling techniques for handling imbalanced data, and its K-folds cross-validation may not suit situations with uneven class distributions, particularly those found in our datasets.

In short, two ML classifiers and two feature categories extracted from accelerometer data were tested to determine the presence of staggers in sheep. In the end, Random Forests achieved the highest classification performance. These outcomes highlight the usefulness of ensemble methods for addressing problems with small sample sizes and class imbalance.

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

Patterns of daily activity captured through accelerometer metrics (Maxz and MI) showed apparent shifts in the final phase of the trial among sheep with visible symptoms (LSS > 1), as revealed by the significant interaction of period, diurnal time class, and LSS. MI-based activity rose across all sheep near the trial's conclusion, but those with symptoms registered a comparatively smaller rise during the closing days. This could reflect motor function decline associated with neurotoxins produced by the endophyte. Even with a few sheep displaying clear symptoms, accelerometer readings indicated behavioral modifications between the trial's early and late stages. The strongest modifications appeared during daylight. At night, periods of lower sheep activity produced no detectable differences in accelerometer metrics across animals with different LSS values. The clearest behavioral shift observed from start to finish centered on the z-axis, which tracks vertical head motions **(Figures 3 and 4)**. Vertical head motions (z-axis) happen mainly during grazing, with the other head motions (x- and y-axes) occurring to a lesser extent. Peak differences in activity levels during midday, combined with distinct z-axis contrasts between early and late trial phases, indicate that perennial ryegrass staggering affects grazing behavior.

Accelerometer metric values and trends showed noticeable differences among individual sheep even in the initial trial phase (e.g., **Figures 3 and 4**). These inter-animal differences probably arose from personal behavioral traits and slight inconsistencies in the accelerometer equipment [9]. Such animal-specific variation underscores the value of building algorithms centered on within-animal behavioral changes rather than on comparisons across groups of animals [34]. In addition to accelerometers, technologies such as noseband sensors could benefit livestock tracking on endophyte-affected pastures by measuring grazing duration and bite rates [2, 3, 35]. Advances such as Bluetooth® (Bluetooth SIG, Inc., Kirkland, WA, USA) support instant data transmission from monitoring devices. Individual sheep tracking, paired with machine learning and potential "change-point" detection methods, could prove highly useful for early identification of perennial ryegrass staggers and other health issues.

Remote monitoring tools, such as accelerometers, can assist producers by alerting them when sheep need to be relocated away from endophyte-affected paddocks toward areas with non-toxic endophyte ryegrass or endophyte-free forage. Store-on-board accelerometers used in this trial record movement data directly on the unit and release the information only after detachment and download. However, real-time livestock monitoring using accelerometers is now entering commercial use. Herddogg (https://herddogg.com, accessed 1 February 2022) ear tags generate an activity index from 3-axis accelerometers at six-minute intervals. The tags send this index to a "dog bone" receiver using Bluetooth® technology. The receiver then forwards the data to the internet through cellular networks. With ongoing improvements, near-real-time or real-time systems could upload data online for processing using machine learning techniques and/or change-point algorithms, followed by interpretation with artificial intelligence tools [36]. Producers could then receive alerts about possible endophyte-toxin effects in sheep, prompting further evaluation.

On-animal sensor monitoring of livestock should deepen insights into animal behavior within pastures containing endophyte-affected forages. Scasta *et al.* [4] observed that cattle alkaloid intake varied when tall grass prairies with

© 2025 The Author(s). | Published by Agro Intelligence Research Press

endophyte-infected tall fescue received patch burning. Tall fescue influences rangeland structural diversity by modifying outcomes from patch burning and grazing activities [2, 3]. Tracking grazing patterns and the possible health impacts of endophyte-infected forage could strengthen knowledge of these systems and improve management approaches.

Conclusion

This case study demonstrated the successful application of ML classification to accelerometer data to detect behavioral shifts associated with ryegrass staggers in sheep. Findings from the research highlight ML as an effective approach for building algorithms that detect staggers using accelerometer inputs, given its ability to integrate patterns across varied metrics. Although additional studies are required, merging ML with continuous accelerometer-based monitoring of sheep activity holds promise for identifying when endophyte toxins begin affecting animal welfare (ryegrass staggers) and when movement to alternative pastures is warranted. Furthermore, livestock health and grazing surveillance through accelerometers and similar on-animal devices can supply useful information for overseeing endophyte-affected pastures and supporting strategies such as patch-burn grazing.

Acknowledgements

None

Conflict of interest

None

Financial support

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.

Ethics statement

None

Received: 22 Feb 2024 Revised: 28 Apr 2024 Accepted: 14 May 2024
Published online: 25 November 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

© 2025 The Author(s). | Published by Agro Intelligence Research Press

1. di Menna ME, Finch SC, Popay AJ, Smith BL. A review of the Neotyphodium lolii/Lolium perenne symbiosis and its associated effects on animal and plant health, with particular emphasis on ryegrass staggers. N Z Vet J. 2012;60(6):315–28.

2. Raynor EJ, Derner JD, Soder KJ, Augustine DJ. Noseband sensor validation and behavioural indicators for assessing beef cattle grazing on extensive pastures. Appl Anim Behav Sci. 2021;242:105402.

3. Raynor EJ, Hillhouse HL, Debinski DM, Miller JR, Schacht WH. Recoupling cross-scale interactions in tall fescue-invaded tallgrass prairie. Landscape Ecol. 2021. https://doi.org/10.1007/s10980-021-01348-x.

4. Scasta JD, McCulley RL, Engle DM, Debinski D. Patch burning tall fescue invaded grasslands alters alkaloids and tiller defoliation with implications for cattle toxicosis. Rangeland Ecol Manag. 2021;75:130–40.

5. Bailey DW, Trotter MG, Knight CW, Thomas MG. Use of GPS tracking collars and accelerometers for rangeland livestock production research. Transl Anim Sci. 2018;2(1):81–8.

6. Dobos RC, Taylor DB, Trotter MG, McCorkell BE, Schneider DA, Hinch GN. Characterising activities of free-ranging Merino ewes before, during and after lambing from GNSS data. Small Rumin Res. 2015;131:12–6.

7. Fogarty ES, Swain DL, Cronin GM, Moraes LE, Bailey DW, Trotter MG. Potential for autonomous detection of lambing using global navigation satellite system technology. Anim Prod Sci. 2020;60(9):1217–26.

8. Fogarty ES, Swain DL, Cronin GM, Moraes LE, Trotter M. Behaviour classification of extensively grazed sheep using machine learning. Comput Electron Agric. 2020;169:105175.

9. Tobin C, Bailey DW, Trotter MG, O'Connor L. Sensor based disease detection: A case study using accelerometers to recognize symptoms of Bovine Ephemeral Fever. Comput Electron Agric. 2020;175:105605.

10. Suji Prasad SJ, Thangatamilan M, Suresh M, Panchal H, Rajan CA, Sagana C, et al. An efficient LoRa-based smart agriculture management and monitoring system using wireless sensor networks. Int J Ambient Energy. 2021:1–4.

11. Tzounis A, Katsoulas N, Bartzanas T, Kittas C. Internet of Things in agriculture, recent advances and future challenges. Biosyst Eng. 2017;164:31–48.

12. Fletcher LR, Finch SC, Sutherland BL, deNicolo G, Mace WJ, van Koten C, et al. The occurrence of ryegrass staggers and heat stress in sheep grazing ryegrass-endophyte associations with diverse alkaloid profiles. N Z Vet J. 2017;65(5):232–41.

13. Keogh RG. Induction and prevention of ryegrass staggers in grazing sheep. N Z J Exp Agric. 1973;1(1):55–7.

14. Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O. SAS for mixed models. Cary, NC: SAS Institute Inc. 2006. p. 814.

15. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proc 14th Int Joint Conf Artif Intell. 1995;2:1137–43.

16. Hernandez J, Carrasco-Ochoa JA, Martínez-Trinidad JF. An empirical study of oversampling and undersampling for instance selection methods on imbalance datasets. In: Iberoamerican Congress on Pattern Recognition. 2013. p. 262–9.

17. Mohammed R, Rawashdeh J, Abdullah M. Machine learning with oversampling and undersampling techniques: overview study and experimental results. In: 2020 11th International Conference on Information and Communication Systems (ICICS). 2020:243–8.

18. Tan P-N, Steinbach M, Kumar V. Introduction to data mining. Pearson Education India; 2016.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

19. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.

20. Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugen. 1936;7(2):179–88.

21. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.

22. Dietterich TG. Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Systems. 2000:1–15.

23. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.

24. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.

25. Nord C, Keeley J. An introduction to the HPFOREST procedure and its options. In: Midwest SAS Users Group Conference Proceedings; Cincinnati, OH. Paper AA20. 2016. https://www.mwsug.org/proceedings/2016/AA/MWSUG-2016-AA20.pdf.

26. Alvarenga FAP, Borges I, Palkovič L, Rodina J, Oddy VH, Dobos RC. Using a three-axis accelerometer to identify and classify sheep behaviour at pasture. Appl Anim Behav Sci. 2016;181:91–9.

27. Barwick J, Lamb D, Dobos R, Schneider D, Welch M, Trotter M. Predicting lameness in sheep activity using tri-axial acceleration signals. Animals. 2018;8(1):12.

28. Barwick J, Lamb DW, Dobos R, Welch M, Trotter M. Categorising sheep activity using a tri-axial accelerometer. Comput Electron Agric. 2018;145:289–97.

29. Campbell HA, Gao L, Bidder OR, Hunter J, Franklin CE. Creating a behavioural classification module for acceleration data: using a captive surrogate for difficult to observe species. J Exp Biol. 2013;216(24):4501–6.

30. Marais J, le Roux SP, Wolhuter R, Niesler T. 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). 2014:97–102.

31. Raschka S, Mirjalili V. Python machine learning: Machine learning and deep learning with python, scikit-learn, and tensorflow. 2nd Edition. Packt Publishing; 2017.

32. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The weka data mining software: An update. ACM SIGKDD Explor Newsl. 2009;11(1):10-8.

33. Kotthoff L, Thornton C, Hoos HH, Hutter F, Leyton-Brown K. Auto-weka 2.0: Automatic model selection and hyperparameter optimization in weka. J Mach Learn Res. 2017;18(1):826-30.

34. Bailey DW, Trotter MG, Tobin C, Thomas MG. Opportunities to apply precision livestock management on rangelands. Front Sustain Food Syst. 2021;5:93.

35. Weinert JR, Werner J, Williams CA. Validation and implementation of an automated chew sensor–based remote monitoring device as tool for equine grazing research. J Equine Vet Sci. 2020;88:102971.

36. Gurarie E, Andrews RD, Laidre KL. A novel method for identifying behavioural changes in animal movement data. Ecol Lett. 2009;12(5):395-408.

© 2025 The Author(s). | Published by Agro Intelligence Research Press