Accurate and interpretable prediction of poor health in small ruminants with accelerometers and machine learning

Axel X. Montout[a,b], Ranjeet S. Bhamber[a,b], Debbie S. Lange[c], Doreen Z. Ndlovu[d], Eric R. Morgan[e], Christos C. Ioannou[f], Thomas H. Terrill[g], Jan A. Van Wyk[h], Tilo Burghardt[i], and Andrew W. Dowsey[a,b,1]

[a]Bristol Veterinary School, University of Bristol, Bristol, UK; [b]Department of Population Health Sciences, Bristol Medical School, University of Bristol, UK; [c]13 Spey St., Extension 3, North Mead, Benoni 1501, Gauteng Province, South Africa; [d]KZN Department of Agriculture and Rural Affairs, P/B X9059, Pietermaritzburg, 3200, KwaZulu-Natal Province, South Africa; [e]School of Biological Sciences, Queen's University Belfast, Belfast, UK; [f]School of Biological Sciences, University of Bristol, Bristol, UK; [g]Department of Agricultural Sciences, Fort Valley State University, Fort Valley, Georgia, USA; [h]Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, South Africa; [i]Department of Computer Science, University of Bristol, Bristol, UK

This manuscript was compiled on August 4, 2020

Accurate assessment of the health status of individual animals is a key step in timely and targeted treatment of infections, which is critical in the fight against anthelmintic and antimicrobial resistance. The FAMACHA scoring system has been used successfully to detect levels of anaemia caused by infection with the parasitic nematode *Haemonchus contortus* in small ruminants and is an effective way to identify individuals in need of treatment. However, assessing FAMACHA is labour-intensive and costly as individuals must be manually examined at frequent intervals over the *Haemonchus* season. Here, we show that accelerometers can measure individual activity in extensively grazing small ruminants subject to natural *Haemonchus contortus* worm infection in southern Africa over long time-scales, and when combined with machine learning, can predict the smallest pre-clinical increases in FAMACHA score as well as those individuals that respond to treatment, all with high precision (>95%). We demonstrate that these classifiers remain robust over time, and remarkably, generalise without retraining across goats and sheep in different regions and types of farming enterprise. Interpretation of the trained classifiers reveal that as the effect of haemonchosis increases, both sheep and goats exhibit a similar reduction in the fine-grained variation of their activity levels. Our study thus reveals common behavioural patterns across small ruminant species, which low-cost biologgers can exploit to detect subtle changes in animal health and enable timely and targeted intervention. This has real potential to improve economic outcomes and animal welfare as well as limit the use of anthelmintic drugs and hence diminish pressures on anthelmintic resistance under conditions of both commercial and resource-poor communal farming.

FAMACHA | *Haemonchus contortus* | Anthelmintic Resistance | Precision Livestock Farming | Accelerometers | Machine Learning

Livestock farming in resource-poor (RP) communities presents multiple challenges. Sheep and goat farming in developing countries suffers from tremendous economic losses from a variety of diseases, including parasitic helminth infections (1). Optimal helminth management is imperative for a farmer to achieve, but is complex and especially difficult without access to expert help.

The gastro-intestinal nematode *Haemonchus contortus (H. contortus*) has a particularly heavy impact on small ruminants in tropical and subtropical regions, as these regions provide a favourable environment for its development. Each female *H. contortus* produces up to 10,000 eggs per day (2), and these develop to infective larvae in a few days under warm and moist conditions. Re-infection can result in high parasite burdens and acute disease outbreaks often leading to death, especially among young animals (3). Disease is primarily the result of blood-feeding by adult worms in the abomasum, leading to anaemia, protein loss, and associated consequences for health, growth and fertility (4). The economic loss due to helminth infection in sheep and goat production is substantial, for example an estimated $40 million per annum in the Kano area of northern Nigeria and $26 million per annum in Kenya (1).

Although multiple worm control strategies for RP farmers exist (5), including chemical dewormers, vaccination, anthelmintic drugs, grazing management, specific diets and ethnoveterinary remedies, they all require high manual labour and other expenses (6, 7). In addition, widespread use of anthelmintic drugs has led to high prevalence of anthelmintic resistance (AR) in countries such as South Africa (8–10). This is due to farmers relying on anthelmintics as the sole method of control against helminth infection, and poor practices such as treating the entire herd when only a few individuals are affected. Although helminth infections are curable, they are common and exact high ongoing costs relative to other diseases

Significance Statement

Increasing availability make biologgers and machine learning viable solutions to current challenges in global livestock farming. We demonstrate a pipeline that accurately predicts the earliest signs of parasitic disease in small ruminants. With *Haemonchus contortus* nematode infection in sheep and goats as the exemplar, we illustrate that the predictive model generalises across time and even species without retraining. We show that prediction is driven by a reduction in the variation of activity levels in animals with poor health. Our findings suggest that health of individual livestock can be monitored remotely, reducing labour costs, improving animal welfare, and allowing for targeted selective treatment under contrasting farming conditions. This will decrease animal loss, maximise economic outcomes, and reduce pressures on drug resistance.

A.W.D and T.B. designed and supervised machine learning research. J.A.V.W designed and supervised field trials and data collection. D.L. and D.Z.N. conducted field trials. J.A.V.W, E.R.M, T.H.T and C.C.I contributed parisitology, agricultural and behavioural science expertise. A.X.M and R.S.B. performed research and analyzed data. A.X.M wrote the paper with contributions from all other authors.

[1]To whom correspondence should be addressed. E-mail: andrew.dowsey@bristol.ac.uk

due to the complexity and operational difficulties of effective and sustainable management (11). Indeed, farmers in Sub-Saharan Africa rank helminths as the most important disease in small ruminants, in spite of more visible (*e.g.* ectoparasitic) and ostensibly more damaging (*e.g.* foot and mouth disease virus) pathogens (12).

A path to more sustainable and efficient control of *Haemonchus* infection is the effective clinical evaluation of individual animals, and selective treatment of those unable to cope (13), leaving the rest untreated. This is possible because helminths are aggregated among their hosts, such that a few individuals in a group tend to carry the majority of the disease burden, and disproportionally drive onward transmission (14). The untreated individuals provide refugia for anthelmintic-susceptible genotypes, slowing the development of AR (13, 15, 16). The FAMACHA clinical system was developed to deliver on this principle in sheep and goats, and consists of a calibrated colour chart against which the colour of the conjunctivae of sheep and goats is compared (17, 18). FAMACHA scores range from 1 to 5, with score 1 being the most healthy, equating to haematocrit $\geq 28\%$, and score 5 the most severely anaemic (haematocrit $\leq 12\%$). The scoring system requires minimal training, provides immediate results and does not rely on expensive equipment or laboratory analysis. However, the training is specific and the trainers are mostly in short supply, particularly in RP regions. Furthermore, the system relies on frequent close handling and examination of individual animals, which is laborious and costly, given that examination of entire herds is required weekly during high risk periods.

A promising alternative could be to use biologgers to remotely monitor behaviour linked to poor health (19–22). Thanks to advancements in on-animal sensor technology, effective tracking and monitoring of terrestrial animals can now give access to large quantities of data on their behaviour and their interactions between each other and with their environment (23). Accelerometry data has been successfully used to classify livestock behaviour, for example Moreau *et al.* (24) identified eating, resting and walking in goat activities by using tri-axial accelerometers together with moving averages and multiple thresholds. This approach is, however, unable to detect localised events in activity and is highly sensitive to the sensor position on the animal. Other studies used machine learning approaches. Vazquez *et al.* (25) used a combination of accelerometry and gyroscopic data with online learning to avoid sheep behaviour (walking, standing and lying) classification deterioration over time due to the intrinsic changing nature of the input data. Scoley *et al.* (26) studied the effect of milk and forage feeding in dairy calves with the IceQube® automatic activity sensors (IceRobotics Ltd., Edinburgh, Scotland, UK). The study revealed that calves had activity level which might be linked to hunger when they were fed conventional level of milk replacement after being fed high level of milk replacement in their early life. By combining GPS data and 3-axis accelerometry data González *et al.* (27) were able to accurately detect foraging and travelling behaviour in grazing cattle. Högberg *et al.* (28) measured the accuracy of two commercial sensors (CowScout(GEA Farm Technologies)) and the IceTag® (IceRobotics Ltd.) mounted on dairy cows. Both devices use accelerometers to determine lying, standing and walking. The study showed that lying and standing could accurately be detected but walking detection was inaccurate.

Although these studies focus on specific activities and all have the potential to be used for livestock management through the monitoring of behaviour, in all cases they did not consider the potentially rich raw data directly, but rather through summary statistics such as the frequency of detected behavioral events or lengths of behavioral periods, which are open to subjective judgement, detection errors and inaccuracies. With the aim of improving animal welfare and reducing drug use, labour and other costs, we have developed a novel automated approach for detecting changes in the FAMACHA score of individual animals based solely on accelerometry data (Fig. 1).

Our study group consists of a sheep flock in Delmas, Mpumalanga Province, South Africa, and a goat herd at Cedara Government Agricultural Animal Production Research Farm, Howick, KwaZulu-Natal, South Africa (Table 1). A transponder containing an accelerometer was suspended by

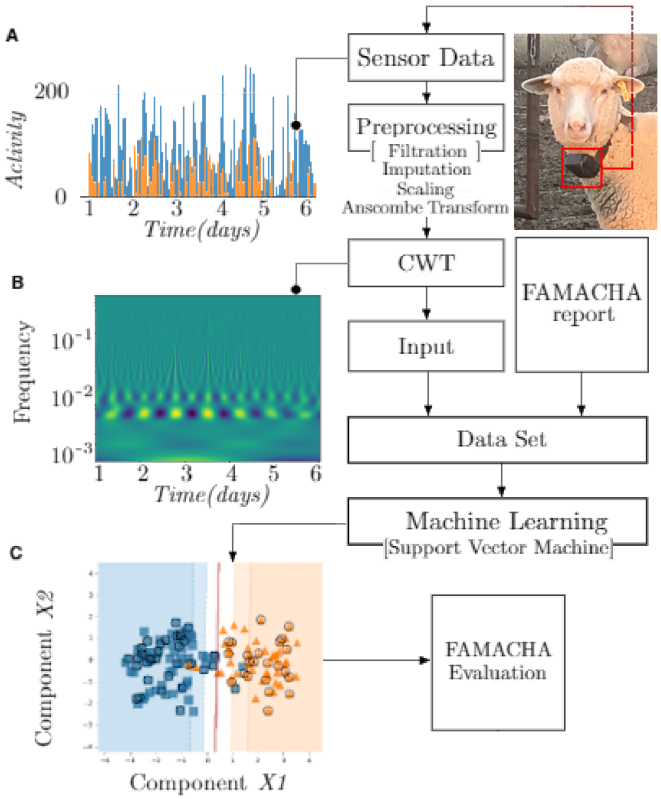

Fig. 1. Schematic of our machine learning pipeline. (A) The biologger outputs accelerometry data as a count of the instances the acceleration exceeded 2g over a short interval. Here we show derived activity time series for two animals (blue/orange) over a 5 days period with a bin size of 10 minutes. As can be seen, there are significant differences in signal magnitude which we account for through preprocessing. **(B)** After normalisation and variance stabilisation, we transform the data with a Continuous Wavelet Transform (CWT), which separates the activity levels into longitudinal dynamics (*x* axis) at a range of temporal scales (Frequency, *y* axis). Higher frequencies represent transient bursts of activity, while lower frequencies represent more consistent levels. For example, the day/night cycle can now clearly be seen. **(C)** The high-dimensional CWT data is then combined with the FAMACHA report for supervised machine learning. The scatter plot shows clustering of the animals in this space (blue and orange points represent healthy and unhealthy animals respectively; those not used for training are circled, and are used to test classification performance). A Support Vector Machine with 10-fold cross-validation repeated 100 times is then used to derive the classification boundary (red) and to derive predicted probabilities that each animal is healthy or unhealthy (shading from blue to orange).

Table 1. Characteristics of the study group

	Cedara	Delmas
Data collection period	April 2012 to July 2013	March 2015 to April 2016
Species	Goat	Sheep
Animals tagged	227	64
Tag type	Accitrack v1	Accitrack v2
Age range	2-6 years	2-6 years
Average weight	44.19 kg	72.79 kg
FAMACHA evaluation	fortnightly	weekly
Animals evaluated	64	31
FAMACHA 1 → 1	31.9%	34.9%
FAMACHA 1 → 2	12.8%	18.3%
FAMACHA 2 → 1	13.5%	18.4%
FAMACHA 2 → 2	13.9%	28.2%
FAMACHA 3+	28.8%	0.2%

a sturdy ribbon around the neck of each animal and used to measure activity levels continuously for more than a year (Table 1). During this time, the FAMACHA score was assessed every week for sheep, and every two weeks for goats. A supervised machine learning pipeline was developed (Fig. 1) to predict an increase in FAMACHA score from 1 ("optimal") to 2 ("acceptable"), as well as from 2 to 1 after individual anthelmintic treatment. Key to the performance of this pipeline is the use of the Continuous Wavelet Transform (CWT) on the binned accelerometry data so that the classifier can learn discriminative features based across the time, duration and intensity of fluctuations in activity levels. Classification drift including robustness to seasonality was assessed through temporal validation, while generalisability was assessed by using the sheep farm as an external validation dataset for the goats, and vice versa. As our model is linear, we were also able to perform inverse transformation of the trained classifiers for visual interpretation of the behavioral cues that distinguish healthy animals from those with a trajectory to poorer health.

Results.

Classifying health status. We first aimed at predicting which individuals would progress from a FAMACHA score of 1 to 2, based solely on the 7 days of accelerometry data immediately prior to the increase. For this we had 200 examples of FAMACHA $1 \rightarrow 1$ and 57 examples of FAMACHA $1 \rightarrow 2$ for the sheep at Delmas, respectively, and 176 examples of FAMACHA $1 \rightarrow 1$ and 67 examples of FAMACHA $1 \rightarrow 2$ for the goats at Cedara. With these annotations we trained and tested our supervised machine learning pipeline using 10-fold cross-validation repeated 100 times. As shown in Fig. 2, our dimensionality reduction revealed clear clusters in the datasets which could be accurately classified. The machine learning pipeline was hence able to predict an increase in FAMACHA score with mean precision of 96% and 98% for individual sheep and goats (respectively) declining in health, and a consistently low FAMACHA score with respective mean precision 95% and 98% for sheep and goats. In a Receiver Operating Characteristic (ROC) analysis, the mean area under the curve (AUC) was 97% for the sheep and 98% for the goats.

Classifying the response to treatment. Each animal which scored ≥ 2 during a FAMACHA evaluation was immediately treated with Levamisole (Ripercol-L, Bayer Animal Health) at $7.5\, mg.kg^{-1}$ as part of routine husbandry (29). We hence created a dataset to examine our ability to determine which animals responded well to treatment, comparing those that subsequently decreased from FAMACHA 2 to 1 with those that continued at FAMACHA 2. This was based on classifying the 7 days of accelerometer data immediately following treatment.

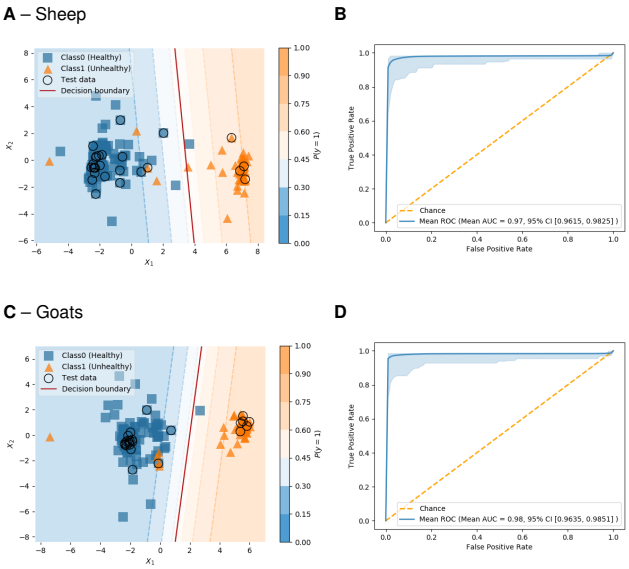

Fig. 2. Classifying health status. The machine learning was trained to discriminate FAMACHA rising to 2 against staying at 1 using 7 days of accelerometry data directly preceding FAMACHA evaluation. (a) Scatter plot and (b) Receiver Operating Characteristic (ROC) curve for training and testing on sheep. (c) Scatter plot and (b) ROC curve for training and testing on goats.

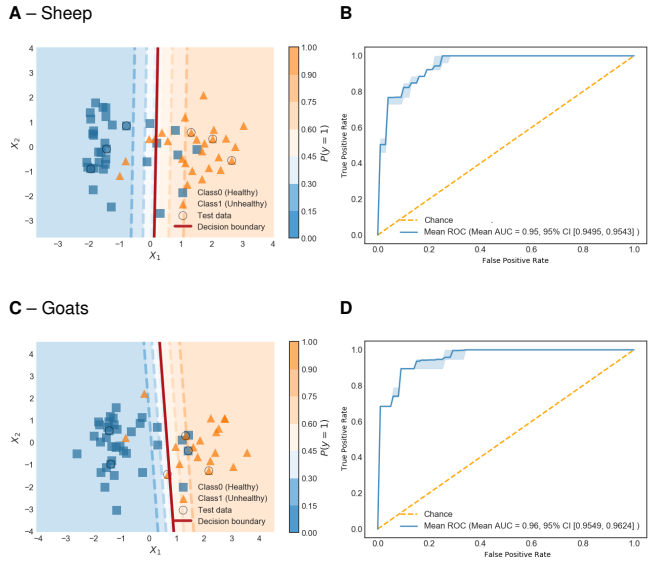

Fig. 3. Classifying the response to treatment. The machine learning was trained to discriminate FAMACHA falling to 1 against staying at 2 using 7 days of accelerometry data directly proceeding anthelmintic treatment. (A) Scatter plot and (B) Receiver Operating Characteristic (ROC) curve for training and testing on sheep. (C) Scatter plot and (D) ROC curve for training and testing on goats.

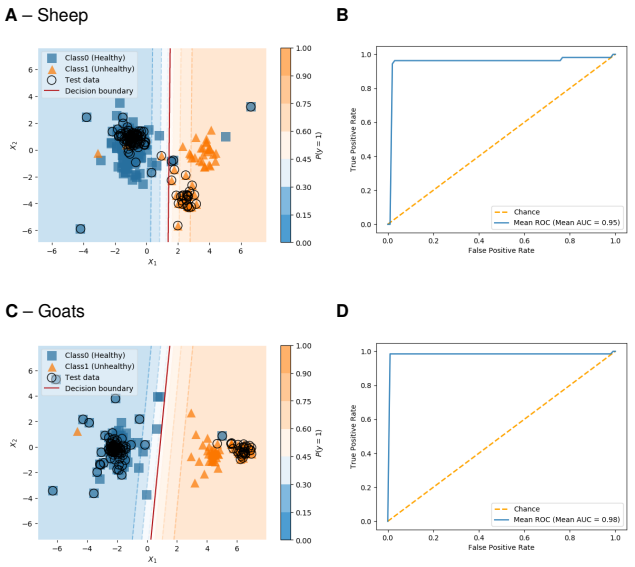

Fig. 4. Temporal validation. The machine learning was trained to discriminate FAMACHA rising to 2 against staying at 1 using 7 days of accelerometry data directly preceding FAMACHA evaluation. To assess concept drift, data from May to October was used for training while data from November to April was used for testing. (A) Scatter plot and (B) ROC curve for training and testing on sheep. (C) Scatter plot and (D) ROC curve for training and testing on goats.

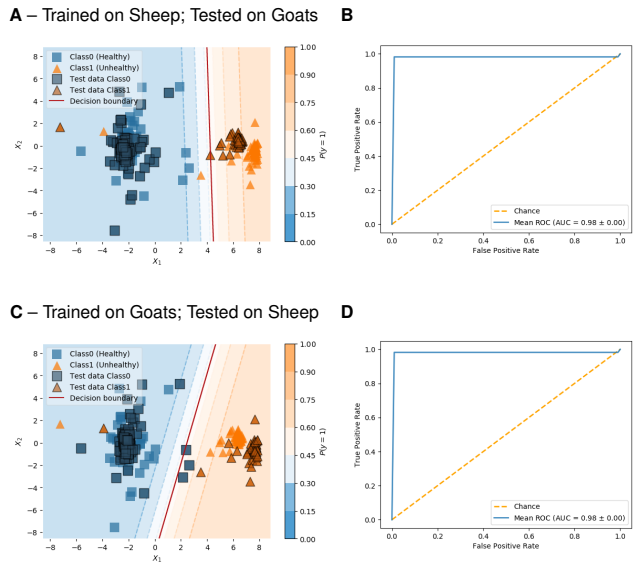

Fig. 5. External validation across farm, region and species. (A) Scatter plot and (B) ROC curve for the classifier trained on the sheep at Delmas and tested on the goats at Cedara. (C) Scatter plot and (D) ROC curve for the reciprocal.

For this we had 32 examples of FAMACHA $2 \rightarrow 1$ and 26 examples of FAMACHA $2 \rightarrow 2$ for the sheep, and 35 examples of FAMACHA $2 \rightarrow 1$ and 19 examples of FAMACHA $2 \rightarrow 2$ for the goats. This resulted in a modest training set size which may diminish the performance of the machine learning. Nevertheless, the classifier was able to predict a drop in FAMACHA score indicating a response to treatment with mean precision 80% and 76% for the sheep and goats, respectively, and no change in FAMACHA score with mean precision of 86% and 91%. As shown in Fig. 3, the mean AUC was 95% for the sheep and 96% for the goats.

Temporal validation to assess concept drift. The notion of "concept drift" describes the decrease in performance of a given classifier due to changing environmental or sensing conditions over a period of time. In other words, training data collected at the start of a given period becomes less representative of future data. This is a common issue in long duration supervised classification problems that use real life data which is, in most scenarios, changing intrinsically (25). An analysis was devised to test how much concept drift affects our findings. Two 6-month periods of data were extracted from the sheep and goat datasets to maximise seasonal differences: (i) the first 6 month period from May to October; (ii) the subsequent 6-month period from November to April. It is clear from observing the resulting scatter plots of the trained classifiers (Fig. 4) that while there is little observable drift in the activity of the healthy animals, the less healthy animals cluster differently depending on time period. Nevertheless, these clusters do not interfere with the decision boundary and hence the overall precision of prediction did not noticeably decrease, yielding a mean precision of 96% and 96% for the less healthy sheep and goats, respectively, a mean precision 98% and 100% for the healthy sheep and goats, and a mean AUC of 95% for the sheep and 98% for the goats.

External validation to assess generalisability across farm, region, and species. While a high degree of predictive power was achieved by training the model on each farm independently, a crucial requirement for the practical application of our technique is the ability of the classifier to generalise across farms without retraining. To demonstrate this, we went a significant step further by examining the generalisability of our approach from the goats to the sheep and vice-versa, hence assessing not only the robustness of our approach across farms, regions and a 3-year time gap, but also across species. In addition, version 1 of the Accitrack accelerometer tag was used on the goats, whereas version 2 was used on the sheep, and the FAMACHA evaluators were different on each farm. Since in this study analysis we pooled the data from both farms, we used 257 training samples for sheep, and 243 testing samples for goats (and vice-versa). The result revealed that despite the marked differences, it was still possible to accurately predict the increase of FAMACHA score highly robustly (Fig. 5), with a mean precision of 100% and 94% for the less healthy sheep and goats respectively, a mean precision 99% and 99% for the healthy sheep and goats, and a mean AUC of 98% for the sheep and 98% for the goats. Similar to the temporal validation above, the less healthy sheep and goats form separate clusters in the classifier space. A modest drift between healthy sheep and goats is also noticeable. Nevertheless, as we maintain strong predictive power, this is compelling evidence that the underlying behaviours captured by the accelerometers are remarkably similar between the sheep and goats, and that this is a robust and reproducible effect.

Interpreting the classifier. In order to understand how our model discriminates between animals on healthy and less healthy trajectories, we analysed the output of the classifiers in Fig. 2 by multiplying the derived feature weights (which define the classification boundary) by the CWT of the mean accelerometer

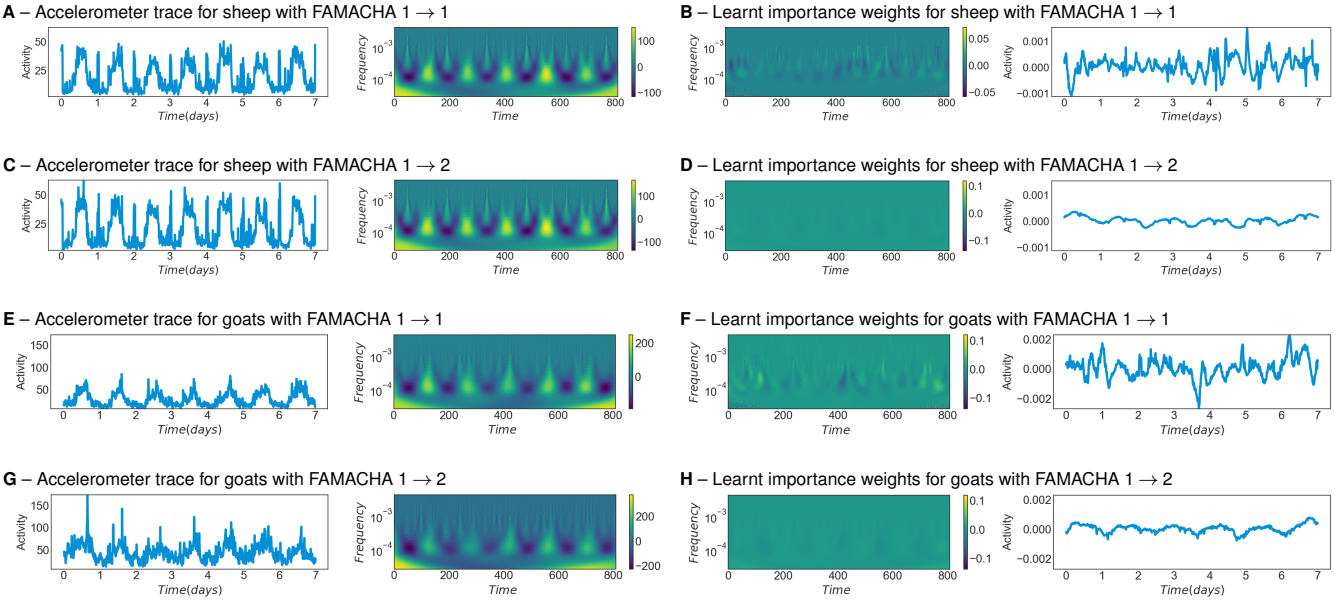

Fig. 6. Interpretation of the classifiers from Fig. 2. (A,C,E,G) Left: Mean accelerometer traces for the classifiers in Fig.2 illustrating noticeable differences between daily activity patterns of the sheep and goats, but no clear differences between healthy and unhealthy animals. Right: CWT transformed mean accelerometer traces. (B,D,F,H) Left: Mean accelerometer CWT multiplied by the learnt feature weights reveal CWT features that are important to classification in the healthy and unhealthy animals. Right: View of the feature importance as differential accelerometer traces after inverse CWT.

traces for each class. As illustrated in Fig. 6, the result shows that the trained model ascribes importance to high frequency activity in the healthy group and low frequency activity in the less healthy group. By performing an inverse CWT on this result, we can also illustrate this effect as a differential accelerometer trace, which shows that, compared to average activity levels across the day and night, the healthy animals appear to have frequent bursts of activity and rest throughout, whereas the less healthy animals are more uniform in their activity levels.

Investigating evolution of activity as parasitic burden increases. With the aim of estimating when the impact of *H. contortus* burden becomes detectable in the animal's activity profile, and hence potentially provide even earlier detection of ill health, for the sheep farm at Delmas we sampled a series of input datasets for our model over a 3 week period before a FAMACHA 2 test result, each with a 7 day sliding window, starting from the first week and moving to the third week by a fixed increment of half a day. In order to ensure a consistent sample size across the 3 week period, in this analysis we compared animals that remained at FAMACHA 1 across the 3 weeks with those that increased to FAMACHA 2 only at the end of the 3 week period. This led to 63 examples of FAMACHA $1 \rightarrow 1 \rightarrow 1 \rightarrow 1$ and 29 examples of FAMACHA $1 \rightarrow 1 \rightarrow 1 \rightarrow 2$. Even though this reduced the sample size considerably compared to Fig. 2, Fig. 7 shows that we still see modest predictive power two weeks before FAMACHA increased to 2 (mean precision 70% & 24%, mean AUC 68%), rising to strong predictive power a week before (mean precision 75% & 54%, mean AUC 83%) and excellent performance just before (mean precision 90% & 85%, mean AUC 98%; equivalent classifiers to Fig. 2, but with a smaller training set).

Discussion. Our analyses reveal that a subtle increase in FAMACHA score from 1 ("optimal") to 2 ("acceptable"), which is considered sub-clinical disease, can be predicted to a high degree of accuracy and precision from behaviour measured using low-cost biologgers. We discovered that the discriminative ability of our classifier was based on higher bursts of short-duration activity levels in the healthy animals. Moreover, the phenomenon appears to be shared between goats and sheep with such strong discriminative strength that a classifier trained on goats in 2012/13 transferred without retraining to a sheep farm 3 years later and 350 miles away, with no loss in predictive power.

It is important to note that due to significant calibration and mounting variation between transponders, including loosening of the transponder over time, it was necessary to perform normalisation of each activity trace to the herd/flock mean. This meant that uniform reductions in activity level from week to week are likely to be normalised out of our data. Nevertheless, a completely uniform reduction in activity level is biologically implausible; instead, intensities of some daily activities are likely to be impacted more than others. Indeed, we have shown that changes in the variation of activity levels constitute a very strong predictor of early changes in health status, as regards haemonchosis and are robust to technical variation. Hence, while efforts in developed countries have been focused on building high-precision, securely-mounted and precisely fitted sensors, we demonstrate that robust results can be gained from much simpler, low-cost systems with rudimentary maintenance requirements suitable for both commercial and RP farmers in developing countries.

In our work we have focused on accelerometry data, but exogenous covariates such as temperature, rainfall, body weight, and production data could all be beneficial to the prediction, particularly as *H. contortus* is well known to hatch after

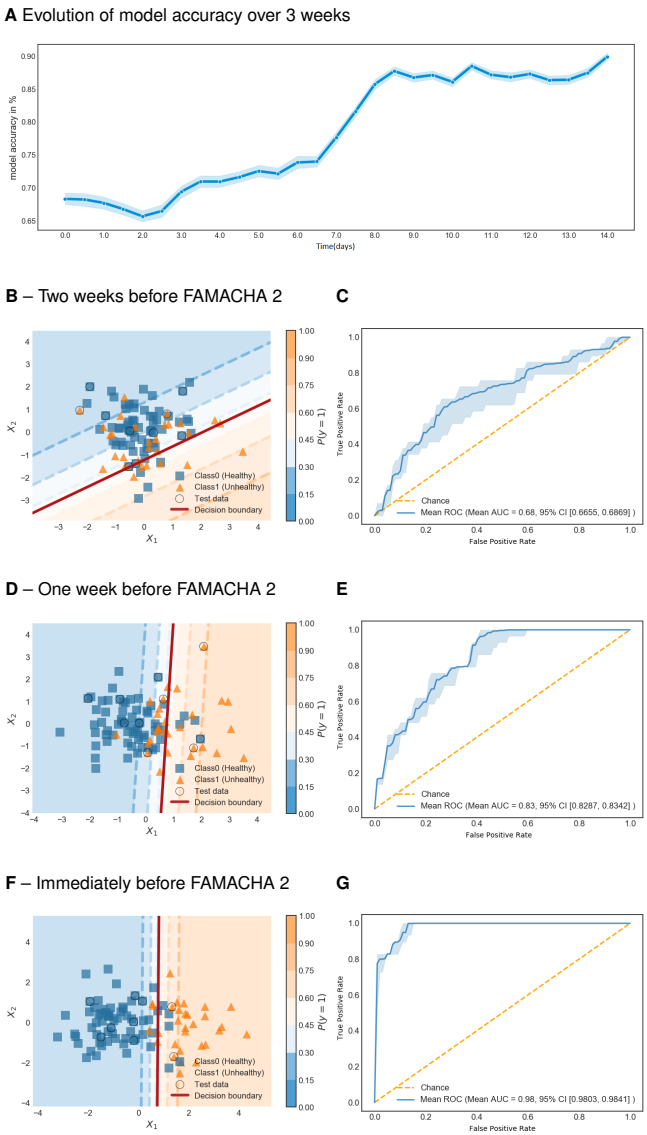

Fig. 7. Evolution of model accuracy for classifying health status of the sheep on a 7 day sliding window over a 3 week period prior to FAMACHA increasing to 2. (A) Classifier accuracy improves as animal activity becomes more discriminative closer to a FAMACHA assessment of 2. (B) Scatter plot and (C) ROC curve for classifier trained on 7 days of data 14 to 21 days prior to FAMACHA increase. (D) Scatter plot and (E) ROC curve for classifier trained on 7 days of data 7 to 14 days prior to FAMACHA increase. (F) Scatter plot and (G) ROC curve for classifier trained on 7 days of data directly preceding FAMACHA increase.

humid, hot weather, and to require rainfall for movement onto pasture (30, 31). Nevertheless, optimal incorporation of these data types is challenging because of their potential non-linear and/or lagged or cumulative effect on health status. Conversely, the fundamental advantage of high-dimensional longitudinal data from accelerometers is that the end effect of these covariates is intrinsically contained within the data directly, which machine learning approaches have the potential to deconvolute. The ability of machine learning to detect health-relevant changes in behaviour under variable climatic conditions could make it especially useful as climate change drives increasingly unpredictable transmission patterns among helminths (32); and, hence, to support adaptation to climate change by RP farmers (33).

Collecting robust annotated datasets is especially challenging in RP farming systems where farming practices are generally less consistent, regulated and well-funded. Because of this and the need for intensive manual labour over a prolonged period, our datasets are highly valuable. Although the training data obtained is dependent on farm topology, location and management, we have shown that a basic machine learning pipeline can discriminate on behavioural cues dominated by fluctuations in activity levels, which translates even across species. Some concept drift was found, particularly among animals with increasing parasitic burden. This suggests two broad avenues for future research: (a) Starting from deployments of our pre-trained model, use of online reinforcement learning techniques to create a 'life-long learning' decision support system which identifies animals for FAMACHA evaluation and feeds back the results to dynamically update model calibration and improve future predictions; (b) Multivariate time-course statistical modelling to further characterise the nature of sheep and goat behaviour in health and disease.

Notably, in this study we have focused on FAMACHA evaluation of *H. contortus* infection; whether multi-label classification of a range of different disease states and transient events is possible is currently unknown but would require a greatly expanded study cohort to attain a suitable predictive power for assessment. In addition, as our accelerometry data is based on a simple activity count paradigm, we hypothesise that activity traces could be derived from other sensor types, such as video, for direct input into our prediction model.

Helminths negatively impact livestock productivity worldwide, and in RP settings are considered 'neglected cold spot' diseases, in that they are preventable in principle, but farmers continue to struggle to manage their effects (34). Technical improvements in helminth control consequently have especially high potential to positively impact farmer livelihoods, with knock-on benefits for human nutrition and health (35). The FAMACHA system has been successfully adopted by smallholder farmers in Africa, but sustained use is difficult because of high training and labour requirements (6, 36). Our results show that it is feasible to apply machine learning approaches to data streams that are attainable on smallholder farms in Africa, to detect early changes in health status and support timely and targeted intervention. In support of the practical use of our platform, we also present evidence for an ability to detect a response to treatment within 7 days, as well as demonstrating that some predictive power exists up to two weeks prior to a FAMACHA assessment score of 2. Both of these findings may be conservative estimates given the substantively smaller training sets available for these analyses.

Importantly, less healthy animals were characterised not by lower activity levels, but by changes in behavioural variability. This pattern is not easily detectable by statistical analysis of activity metrics, but can be successfully discovered by machine learning of data transformed by the CWT. Similar approaches on animals under natural disease challenge could greatly enhance ability to detect and selectively intervene against a range of disease states, in support of animal welfare, food security, and sustainable antimicrobial use, including use in extensively grazed systems.

Methods.

Study group. Data were collected from a farm of 108 acres close to Delmas in Mpumalanga Province, South Africa, and at Cedara, a government research farm and agricultural college, the pastures of which comprise of 25 acres adjacent to Hilton, KwaZulu-Natal. As described in Table 1, 31 female adult *Ile de France* sheep ewes at Delmas and 64 goats at Cedara were individually FAMACHA-evaluated (17) at weekly or fortnightly intervals, respectively, and had associated longitudinal accelerometry data recorded within the study period. On both farms there were multiple improved pastures, which were irrigated, and utilisation occurred by alternation at intervals according to visual assessment of amounts of available water and forage. At both farms young adult ewes/does were randomly selected for the trials, without attention to reproductive class, but remained with their flocks/herds of origin for the duration of each trial. The animals were kraaled at night and let out at a standard time in the mornings for herding to pasture, where they remained until collected and returned to the kraals in late afternoon. Adjustments were made according to season and special management events such as vaccination, hoof inspection, which were conducted first thing in the morning.

Telemetric monitoring. Telemetric monitoring systems were supplied by Accitrack Ltd., Paarl, South Africa. Tagged animals on both farms were equipped with low cost RFID transponders suspended by a sturdy ribbon around the neck (Fig. 1). A single solar-powered base station was installed on each farm, mounted at the top of a five-meter wooden pole. Each transponder contains an active RFID transceiver operating at 868Mhz as well as a battery and an A1 type accelerometer for activity level measurement. The accelerometers had a set acceleration threshold of 2g so that every time an acceleration $\geq$ 2g is sensed, a stored integer is incremented by +1. Version 2 transponders had an increased range (10km versus 1km) and a larger battery, at the expense of significantly increased weight. In addition, version 2 transponders also output minimum and maximum acceleration for the three orthogonal axes at each time interval, but this data was not utilised. All tags were set to transmit the data every minute to the base station, at which time the accelerometer count was reset to zero. In order to extend battery life, the tag only transmits data to the base station once a minute, with data transmitted including the identifier of the transponder, the battery level, the signal strength, a timestamp, and the activity level. Data transmission is not performed if the signal-to-noise ratio drops below 10dB, which can occur when there are significant occlusions between the animal and base station. In these cases, the data for that time interval is lost. Through mobile connectivity with General Package Radio Services (GPRS), the base station then regularly forwards the received data to the Accitrack cloud repository.

Data management and visualisation. All raw data is stored on the cloud repository for two weeks only due to storage limitations, hence it was manually downloaded regularly by a researcher at the University of Pretoria for archival (29). The exported data took the form of Excel spreadsheets containing the sensor outputs in the desired time frame. In cases where the data was not retrieved from the cloud infrastructure, datasets for that time period were lost. For this work, we parsed the raw excel data into an SQL database. The table storing the raw data recorded at the original minute resolution contains 40,659,086 records. The data was re-binned into multiple time resolutions (Δt) for efficient interactive visualisation with zooming capability. This tool was developed for exploratory analysis to allow us to determine whether transponders were faulty and failed to transmit, and provided visual verification of expected behaviours such as decreased activity levels at night. The visualisation also revealed the need for data pre-processing, as changes in average activity signal amplitude across the herd was observed due to varying calibration of the sensors. In addition, the mounted position of the sensor on each animal influenced the sensor's measurement of acceleration, *e.g.* a looser collar would allow broader movement of the sensor thus higher activity values. In contrast, however, long wool tended greatly to reduce the numbers of occasions on which activity registered.

Pre-processing. We consider a set of animals $r = 1, 2, \ldots n$ with associated FAMACHA scores $f_{r,d}$ evaluated on days $d = 1, 2, \ldots T$, where T is the total number of consecutive days being sampled. Each FAMACHA evaluation is associated with a trace of activity counts $a_{r,d}(t_i)$ over time intervals t_i where $i = 0, 1, \ldots \tau$, where τ is the total number of samples within a day (d with a temporal resolution of Δt) either preceding or proceeding the evaluation depending on the anaylsis. Only traces containing a low percentage of missing data and zeros (thus indicating the collar was attached to the animal and functioning correctly) were retained. If the traces exhibited either 50% of the values to be zeros or 20% of the values to be nonexistent, then these were dismissed. For retained traces, a simple padding of the missing data points with a constant negative value was found not to impact the model performance negatively.

To correct for scaling differences due to device calibration and mounting differences between animals, and the mounting loosening over time, we first define a mean "herd-level" activity trace across all animals with FAMACHA evaluations on the same day:

$$h_d(t_i) \quad = \quad \left[\frac{\sum_{r=1}^{n} a_{r,d}(t_i)}{n} \right] \qquad [1]$$

A scaling coefficient for each trace $s_{r,d}$ is then derived as the median of the trace after having divided it by the herd-level trace:

$$s_{r,d} = median \left(\left[\frac{h(t_0)}{a_r(t_0)}, \ldots, \frac{h(t_\tau)}{a_r(t_\tau)} \right] \right) \qquad [2]$$

Normalised activity traces b_r are then derived by multiplying each original trace by its respective scaling coefficient:

$$b_{r,d}(t_i) = s_{r,d} \times a_{r,d}(t_i) \qquad [3]$$

Because our activity data are counts and hence more closely follow Poisson statistics than the Gaussian distribution expected by conventional machine learning methodology, after normalisation we perform an Anscombe transform (37) for approximate variance stabilisation to a Gaussian distribution:

$$b'_{r,d}(t_i) = 2 \times \sqrt{b_{r,d}(t_i) + \frac{3}{8}} \qquad [4]$$

Continuous wavelet transform. We apply a Continuous Wavelet Transform (CWT) (38) to the variance-stabilized activity traces described above. The CWT is a transformation of a temporal signal to a representation in the frequency domain localised by time, accomplished by performing a convolution of a series of mother wavelets with different bandwidths over the signal in the time domain using the following wavelet integral:

$$W\left\{b(t,s)\right\} = \int_{-\infty}^{\infty} b(u)\frac{1}{\sqrt{s}}\psi^{*}\left(\frac{u-t}{s}\right)du \qquad [5]$$

where ψ^{*} is a complex analytic wavelet, in this case we employed a Mexican hat wavelet with a wave length range of 60 minutes to 7 days. The above integral measures the variation of the activity $b(t)$ in the neighbourhood of t proportional to the bandwidth of the wavelet s. The CWT gives an intuitive understanding of our activity data and allows us to separate out localised fluctuations in an animal's activity. Fig. 1A shows the accelerometry data of an animal over a 6 day period in the time domain, while Fig. 1B shows the corresponding CWT.

Machine learning with internal validation. Classification analyses utilised a Support Vector Machine (SVM) (39). Training/testing datasets for this supervised machine learning technique were constructed using the FAMACHA scores and the standardized CWT-transformed activity traces. Repeated nested k-fold cross validation was used to optimize the hyperparameters and evaluate the model. Kim *et al.* (40) showed that the repeated cross validation estimator outperforms the non-repeated version by reducing the variability of the estimator and providing lower bias. Hence we choose to use 100-times repeated 10-fold cross validation to assess a realistic estimate of the performance of the model predictions while making the most of the dataset available. For the Area Under the Curve (AUC) statistic, confidence intervals were computed using the method of LeDell *et al.* (41).

ACKNOWLEDGMENTS. We would like to thank the farmers who so diligently assisted with data collection, and Rachel Coetzee and Jan van Rensburg for trial and data management and support. This work was supported by UK Research and Innovation BBSRC grants BB/S014748/1 and BB/H00940X/1, and The Alan Turing Institute under EPSRC grant EP/N510129/1. We also acknowledge support through research grants from Red Meat Research Development South Africa (RMRDSA) and the John Oldacre Foundation through the John Oldacre Centre for Sustainability and Welfare in Dairy Production, Bristol Veterinary School. The Fulbright Specialist Program supported THT on an assignment in South Africa related to this work.

1. JP Fabiyi, Production losses and control of helminths in ruminants of tropical regions. *Int. J. for Parasitol.* **17**, 435 – 442 (1987).
2. M Saccareau, et al., Meta-analysis of the parasitic phase traits of *Haemonchus contortus* infection in sheep. *Parasites Vectors* **10** (2017).
3. HM Gordon, The epidemiology of parasitic diseases, with special reference to studies with nematode parasites of sheep. *Aust. Vet. J.* **24**, 17–45 (1948).
4. R Besier, L Kahn, N Sargison, J van Wyk, The pathophysiology, ecology and epidemiology of *Haemonchus contortus* infection in small ruminants. *Adv. Parasitol.* **93**, 95–143 (2016).
5. GD Gray, JG Connell, V Phimphachanhvongsod, Worms in smallholder livestock systems: Technologies and practices that make a difference. *Vet. Parasitol.* **186**, 124–31 (2011).
6. JA van Wyk, DP Reynecke, Blueprint for an automated specific decision support system for countering anthelmintic resistance in Haemonchus spp. at farm level. *Vet. Parasitol.* **177**, 212–23 (2009).
7. I Maqbool, Z Wani, RA Shahardar, I Allaie, M Shah, Integrated parasite management with special reference to gastro-intestinal nematodes. *J. Parasit. Dis.* **41** (2016).
8. JA van Wyk, M Stenson, J van der Merwe, R Vorster, P Viljoen, Anthelmintic resistance in South Africa: Surveys indicate an extremely serious situation in sheep and goat farming. *Onderstepoort J. Vet. Res.* **66**, 273–84 (1999).
9. A Vatta, A Lindberg, Managing anthelmintic resistance in small ruminant livestock of resource-poor farmers in South Africa. *J. S. Afr. Vet. Assoc.* **77**, 2–8 (2006).
10. AM Tsotetsi, et al., Prevalence of gastrointestinal helminths and anthelmintic resistance on small-scale farms in Gauteng Province, South Africa. *Trop. Animal Heal. Prod.* **45**, 751–761 (2013).
11. BD Perry, D Grace, K Sones, Current drivers and future directions of global livestock disease dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20871–20877 (2013).
12. B Perry, T Randolph, JJ McDermott, KR Sones, P Thornton, Investing in animal health research to alleviate poverty. *(ILRI, Nairobi, Kenya)* (2002).
13. JA van Wyk, Refugia - Overlooked as perhaps the most potent factor concerning the development of anthelmintic resistance. *Onderstepoort J. Vet. Res.* **68**, 55–67 (2001).
14. RC Tinsley, HR Vineer, R Grainger-Wood, ER Morgan, Heterogeneity in helminth infections: factors influencing aggregation in a simple host-parasite system. *Parasitology* **147**, 65–77 (2020).
15. JE Hodgkinson, et al., Refugia and anthelmintic resistance: concepts and challenges. *Int. J. Parasitol. Drugs Drug Resist.* **10**, 51–57 (2019).
16. AW Greer, JA van Wyk, JC Hamie, C Byaruhanga, F Kenyon, Refugia-based strategies for parasite control in livestock. *Vet. Clin. North Am. Small Anim. Pract.* **36**, 31–43 (2020).
17. JA van Wyk, GF Bath, The FAMACHA system for managing haemonchosis in sheep and goats by clinically identifying individual animals for treatment. *Vet. Res.* **33**, 509–529 (2002).
18. R Besier, L Kahn, N Sargison, J van Wyk, Diagnosis, treatment and management of *Haemonchus contortus* in small ruminants. *Adv. Parasitol.* **93**, 181–238 (2016).
19. C Liddell, E Morgan, K Bull, C Ioannou, Response to resources and parasites depends on health status in extensively grazed sheep. *Proc. R. Soc. B.* **287**, 20192905 (2020).
20. G Falzon, D Schneider, M Trotter, DW Lamb, A relationship between faecal egg counts and the distance travelled by sheep. *Small Rumin. Res.* **111**, 171–174 (2013).
21. JA Martos-Sitcha, et al., Ultra-low power sensor devices for monitoring physical activity and respiratory frequency in farmed fish. *Front. Physiol.* **10** (2019).
22. MN Velasco-Garcia, T Mottram, Biosensor technology addressing agricultural problems. *Biosyst. Eng.* **84**, 1–12 (2003).
23. R Kays, MC Crofoot, W Jetz, M Wikelski, Terrestrial animal tracking as an eye on life and planet. *Science* **348** (2015).
24. M Moreau, S Siebert, A Buerkert, E Schlecht, Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behavior. *Appl. Animal Behav. Sci.* **119**, 158–170 (2009).
25. J Vazquez Diosdado, et al., A combined offline and online algorithm for real-time and long-term classification of sheep behaviour: Novel approach for precision livestock farming. *Sensors* **19**, 3201 (2019).
26. G Scoley, A Gordon, S Morrison, Using non-invasive monitoring technologies to capture behavioural, physiological and health responses of dairy calves to different nutritional regimes during the first ten weeks of life. *Animals* **9**, 760 (2019).
27. LA González, GJ Bishop-Hurley, RN Handcock, C Crossman, Behavioral classification of data from collars containing motion sensors in grazing cattle. *Comput. Electron. Agric.* **110**, 91–102 (2015).
28. N Högberg, J Höglund, A Carlsson, M Saint-Jeveint, L Lidfors, Validation of accelerometers to automatically record postures and number of steps in growing lambs. *Appl. Anim. Behav. Sci.*, 105014 (2020).
29. ND Babayani, Novel approaches to an automated decision support system for on-farm management of internal parasites of small ruminants. *Doctoral dissertation, University of Pretoria, South Africa* (2016).
30. S Walkden-Brown, L Kahn, , et al., Ecology of the free-living stages of major trichostorngylid parasites of sheep. *Vet. Parasitol.* **14**, 1–15 (2006).
31. T Wang, J van Wyk, A Morrison, E Morgan, Moisture requirements for the migration of *Haemonchus contortus* third stage larvae out of faeces. *Vet. Parasitol.* **204**, 258–264 (2014).
32. H Rose, et al., Climate-driven changes to the spatio-temporal distribution of the parasitic nematode, *Haemonchus contortus*, in sheep in Europe. *Glob. Chang. Biol.* **22**, 1271–1285 (2016).
33. PK Thornton, PJ Ericksen, M Herrero, AJ Challinor, Climate variability and vulnerability to climate change: a review. *Glob. Chang. Biol.* **20**, 3313–3328 (2014).
34. TF Randolph, et al., Invited review: Role of livestock in human nutrition and health for poverty reduction in developing countries. *J. Anim. Sci.* **85**, 2788–2800 (2007).
35. B Perry, D Grace, The impacts of livestock diseases and their control on growth and development processes that are pro-poor. *Philos. Trans. R. Soc. B* **364**, 2643–2655 (2009).
36. JG Walker, et al., Mixed methods evaluation of targeted selective anthelmintic treatment by resource-poor smallholder goat farmers in Botswana. *Vet. Parasitol.* **214**, 80–88 (2015).
37. FJ Anscombe, The transformation of poisson, binomial and negative-binomial data. *Biometrika* **35**, 246–254 (1948).
38. A Grossmann, J Morlet, Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.* **15**, 723–736 (1984).
39. C Cortes, V Vapnik, Support vector network. *Mach. Learn.* **20**, 273–297 (1995).
40. JH Kim, Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap.". *Comput. Stat. Data Anal.* **53**, 3735–3745 (2009).
41. E LeDell, M Petersen, M van der Laan, Computationally efficient confidence intervals for cross-validated area under the roc curve estimates. *Electron. J. Stat.* **9**, 1583 (2015).