DEVELOPING SERIALLY DEPENDENT CLASSIFIERS FOR ANIMAL BEHAVIORAL ANALYSIS

by

Thomas Lipinski

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

In

Industrial and Management Systems Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

August 2025

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The increasing global demand for meat production necessitates more sustainable and data-driven livestock management practices. Precision Livestock Farming (PLF) leverages sensor technologies and computational methods to monitor individual animals, enhancing productivity and welfare. In this study, we investigate the use of supervised serially dependent classifiers for predicting free-range cattle behavior using time-series data from GPS and tri-axial accelerometers. Over a nine-week winter period, 22 Angus-based cows were monitored at the Red Bluff Research Ranch located in Montana, USA. Four features—step size, turning angle, and the mean and variance of Signal Vector Magnitude (SVM)—are engineered and used to classify grazing, resting, and traveling behaviors via Hidden Markov Models (HMMs) and Long Short-Term Memory (LSTM) recurrent neural networks. Model development was conducted using leave-one-out cross-validation to assess generalizability across individuals. The HMM achieved an average classification accuracy of 92.7%, and the two types of LSTMs achieved accuracies of 94.3%, and 95.4%. Each of the three types of models revealed ecologically meaningful patterns, including diurnal activity cycles, temperature-dependent behavioral shifts, and increased nighttime grazing during full moon phases. Generalized Linear Mixed Models (GLMMs) were used to expose the interaction effects of time of day and temperature on behavior. The findings support the utility of HMMs and LSTMs as effective tools for cattle behavior classification in extensive rangeland settings. The approach holds promise for real-time PLF applications such as early anomaly detection, resource allocation, and adaptive grazing management.

INTRODUCTION

Precision Livestock Management

According to the Food and Agriculture Organization (FAO), global meat consumption is projected to increase by more than 14% by 2030 and 50% by 2050 driven by population growth and rising incomes, particularly in developing countries (OECD/FAO, 2021). Meeting this demand with traditional livestock farming methods is challenging, as it often results in inefficiencies, higher environmental impact, and increased pressure on natural resources (Chang et al., 2022; Grossi et al., 2018).

Precision Livestock Farming (PLF) refers to the use of advanced technologies and data analytics to monitor, manage, and optimize livestock production. Documented benefits of PLF include: improved animal welfare through continuous monitoring and automated health alerts; enhanced productivity and efficiency through optimized feeding and better reproductive management; cost reductions through labor efficiency and reduced feed and medication costs; environmental sustainability through lower waste and emissions and water and energy savings; data-driven decision making through informed management; traceability and food safety through real-time monitoring and improved biosecurity; and ultimately, increased profitability (Berckmans, 2017; Jiang et al., 2023). Overall, precision livestock farming integrates technology into livestock management, leading to smarter farming practices that benefit animals, ranchers, and the environment (Wyffels et al., 2020).

Behavioral Classification

Traditional livestock management focuses on optimizing production through the sustainable management of forage resources, typically targeted at the average needs of the herd.

While PLF offers a wide range of benefits, its true potential is realized through precise, real-time monitoring of individual livestock behavior (da Silva Santos et al., 2023). Monitoring behaviors such as grazing, resting, and traveling allows for the detection of changes in animal health and nutrition before any changes are detectable to the eye, even by trained individuals (Sundman et al., 2024). Additionally, tracking the movement and grazing activity of each animal allows for real time monitoring of fine-scale rangeland use (Parsons et al., 2023). Understanding individual cattle behavior and health not only contributes to animal welfare but also directly impacts productivity and resource allocation, making PLF essential for sustainable and profitable ranching (Terrell et al., 2017). Accurate monitoring of cattle behavior provides insights that extend across animal sciences, including health assessment, movement pattern prediction, and reproductive management (Diskin and Kenny, 2014; Nyamuryekung'e, 2024; Rautiainen et al., 2022). Each of these fields utilize animal behavior data to obtain actionable information about the animals and resources they use.

By analyzing the patterns in animal movement and behavior, ranchers can optimize grazing schedules, promoting the sustainable use of pastureland while reducing environmental impact. Furthermore, detecting anomalies in behavioral patterns enables the early detection of health-related issues, such as lameness or injuries, which could impact productivity if left untreated. Such insights also assist in selecting animals for breeding programs based on observed health and activity patterns, potentially improving reproductive yield (Diskin and Kenny, 2014). These behavioral classifications can provide a more thorough way for ranchers to make decisions regarding resource allocation. Ranchers will be more informed when choosing animals for breeding that show favorable traits such as a tendency to roam, and more capable of proactively addressing health issues, thereby improving herd productivity and welfare (Mahmud et al., 2021).

While GPS data alone provides valuable information for livestock monitoring, such as pinpointing the location of missing animals or identifying potential theft events (Mancuso et al., 2023), combining GPS data with accelerometer data allows researchers and ranchers to go beyond simple tracking, enabling a deeper understanding of cattle behavior in real-time. This combination not only supports more informed decision-making but also fosters data-driven approaches to livestock management (Aquilani et al., 2022).

To enable the benefits of behavioral predictions, real labeled data must first be collected to train the models which generate behavioral predictions. Collecting extensive labeled data on animal behavior is particularly difficult for animals grazing vast rangelands and animals in cold climates (Beumer et al., 2020) as human observers may struggle to locate and follow the animals without suffering the effects of extreme weather. Additionally, trail cameras (Cusack et al., 2015), UAVs (Koger et al., 2023), and collar-mounted video devices (Thompson et al., 2012) are all appealing ways to circumvent this issue which have been implemented successfully, but come with unique challenges not limited to video quality, camera positioning, and data storage. Human observations are more common due to the simplicity, reliability, and affordability for obtaining activity labels. Once labeled data is collected, creating predictions of cattle behavior can seem straightforward, but validating and verifying these predictions is a substantial challenge. Evaluating prediction accuracy against a limited subset of labeled data provides only a narrow glimpse of overall behavior patterns. To construct a comprehensive understanding, it is essential to examine how patterns in unlabeled data align with established findings in existing literature.

In this study, we use two serially dependent models—Hidden Markov Models (HMMs) (Awad and Khanna, 2015; Rabiner, 1989; Zucchini et al., 2017) and Long Short-Term Memory (LSTM) neural networks (Graves, 2012; LeCun et al., 2015)—to classify cattle behavior. Serial

dependence refers to the characteristic that observations are inherently ordered in time, meaning that each observation depends on preceding (and potentially following) observations. Animal behavior is fundamentally sequential (Kilgour, 1984), and thus, models capable of capturing temporal dependencies are well suited for the task of behavioral classification.

HMMs handle serial dependence through a transition matrix, where the predicted behavior at each timestep is influenced by the behavior predicted at the immediately previous timestep. LSTMs use recurrent connections and gating mechanisms to learn temporal relationships through the entire timeseries.

HMMs can be trained on partially labeled data with Newton-type methods (Nocedal and Wright, 2006). The probabilistic nature of HMMs makes them robust to noise, uncertainty, and missing data, which are common in real-world applications. Additionally, HMMs are interpretable, allowing clear insights into the relationship between behaviors and observed data.

LSTMs can be trained on partially labeled datasets by directing the learning algorithm to focus solely on fitting labeled data. LSTMs are known for their ability to learn complex temporal relationships and produce highly performant predictions in classification tasks (LeCun et al., 2015) but are black box models which cannot be interpreted like HMMs can.

As noted by (Jiang et al., 2023) in their recent scientometric review of the literature, the volume of research in this area is rapidly increasing. Numerous studies have investigated cattle behavior using GPS and accelerometer data, employing a range of modeling techniques, labeling methods, and environmental variables such as snow cover, terrain, and climate. These studies have been reviewed in several recent surveys (Aquilani et al., 2022; Hlimi et al., 2024; Jiang et al., 2023; Mahmud et al., 2021; Monteiro et al., 2021; Riaboff et al., 2022). Additionally, there is a rising emphasis on machine learning (ML) approaches. Despite the growth of the field and the

advantages HMMs and LSTMs offer, serial models are perhaps underexplored in the cattle behavior classification literature. Our objective in this work is to demonstrate the effectiveness of serially dependent models for classifying free-range cattle behavior using GPS and accelerometer data, evaluate their classification performance, and assess their predictions' alignment to behavioral patterns known in existing literature.

Dataset Description

Study Environment: From 2022-01-15 00:00 to 2022-03-23 00:00, 22 multiparous Angus-based cows at the Red Bluff Research Ranch in Norris, MT, USA (long = -111.62°, lat = 45.57◦) were equipped with Lotek LiteTrack 800 collars featuring GPS trackers and tri-axial accelerometers. At the start of the study, the cows grazed the First Feeder pasture (336 ha). After the first weigh day, January 26th, the cows were allowed to graze both First Feeder and Sixteen Mile pastures (225 ha). During the study period, all cows were corralled inside the pasture each Monday, Wednesday, and Friday to be supplemented. On 1/26, 2/23, and 3/23 the cows were taken to another location overnight to be weighed. Figure 1 shows GPS fixes for the cows during the study overlayed on the two pastures. The First Feeder pasture is shown with blue dots, and the orange dots show the Sixteen Mile pasture. Point A shows the corral where cows were taken to be weighed, and Point B is where the cows were corralled inside the pasture to be supplemented.

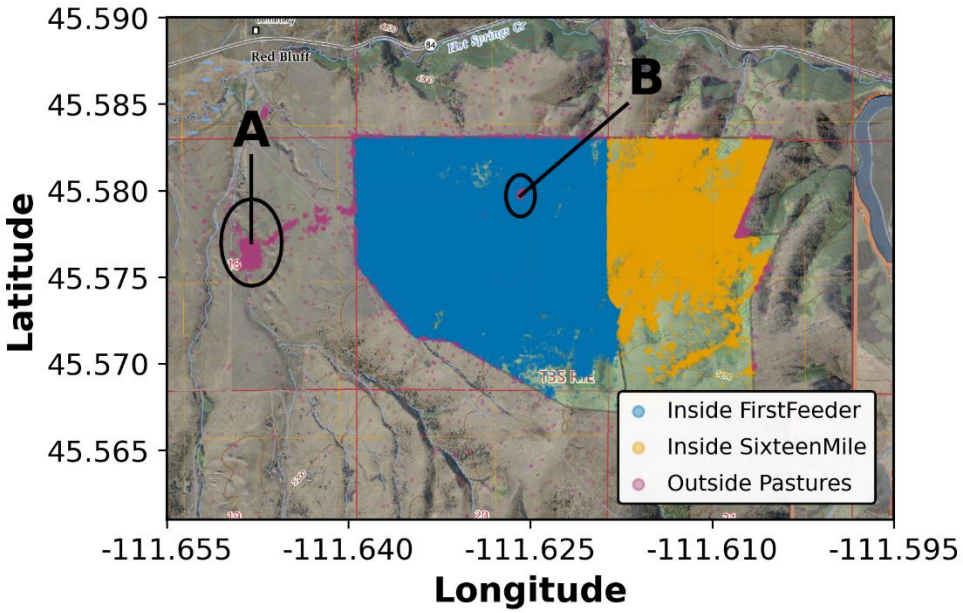

Figure 1. GPS fixes for the cows during the study overlayed on the two pastures.

GPS Data: The LoTek collars were programmed to transmit GPS fixes every 5 minutes and each transmission included the data described in Table 1. A total of 424,146 GPS records were collected during the experiment.

Table 1. GPS data structure.

Variable	Data Type	Sample	Description
GMT Time	Timestamp	1/16/2022 22:20	GMT Time of fix transmission recorded approximately every 5 minutes
Latitude	Float	45.5774483	WGS 84
Longitude	Float	-111.62914	WGS 84
Altitude	Float	1622.35	Reported in meters
Duration	Float	2	Seconds between signal to acquire a fix and the fix being acquired
Temperature	Float	8	Within 0.5°C accuracy

Table 1. Continued.

Variable	Data Type	Sample	Description
Dilution of Precision (DOP)	Float	2.1	Indicator of fix quality error (0 to 50), lower is better.
Satellites	Integer	8	The number of satellites used to acquire the fix (0 - 12)
Cause of Fix	String	GPS Schedule	What signal prompted the fix to be collected

Accelerometer Data:The tri-axial accelerometers were programmed to record raw readings every minute as described in Table 2. During the study period, a total of 2,121,138 accelerometry observations were collected. The accelerometers were attached to collars on the underside of each animal's neck. Positive X (surge) points towards the cows' rear, Positive Y (sway) is to the port side of the cow, and Positive Z (heave) is upward towards the sky.

Table 2. Accelerometer data structure.

Variable	Data Type	Sample	Comment
GMT Time	Timestamp	1/16/2022 22:20	GMT Time recorded every minute
X	Integer	-6	Measured in increments of 32 milli g's
Y	Integer	12	Measured in increments of 32 milli g's
Z	Integer	4	Measured in increments of 32 milli g's
Temperature	Float	8	0.5 C accuracy

Observation Data: Four human observers on March 18th from 10:00 to 12:05 and on March 21st from 09:40 to 11:20 recorded the dominant behavior of the observed cows every minute as one of Grazing, Fighting, Resting, Traveling, Scratching, Drinking, or Consuming Mineral following the structure shown in Table 3. The observations resulted in a total of 42.4 hours of

labeled data corresponding to 2,544 observations. The distribution of dominant behaviors recorded are shown in

Figure 2.

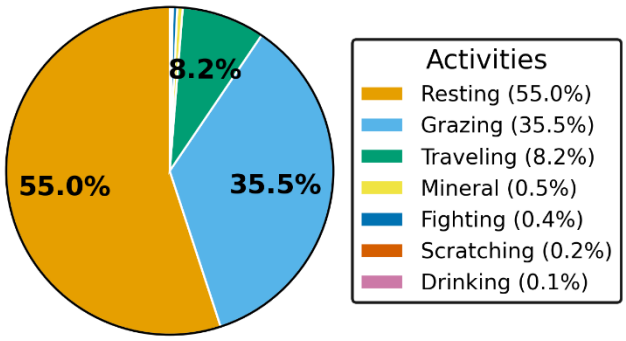

Figure 2. Observed cattle behavior frequency.

Table 3. Activity observation data structure.

Variable	Data Type	Sample	Description
date	Date	3/21/2022	Date of Observation
time	Time	9:44:00	Mountain Standard Time, recorded every minute
cow id	integer	8132	Unique identifier for cows
observer	string	John Doe	Human observers name
activity	string	Grazing	The dominant behavior during the minute
collar	integer	1008	The collar index is related to the sensors. This is a unique key to match cows and collars.

Data Preparation

Missing, Erroneous and Irrelevant Data: Over the 67 full days in this study, with 22 collars generating GPS fixes every 5 minutes, we expect to get 424,512 GPS fixes. We observed 424,146 GPS fixes in the study. Of these fixes, 1,251 records had “0” recorded for each of the numeric fields including longitude and latitude. These fixes were recast as NA.

GPS fixes were not acquired precisely every 5 minutes. To align the timestamps to the observational and accelerometer data, each record was rounded to the nearest 5^{th} minute. Occasionally, multiple values are rounded to the same 5^{th} minute. These collisions were resolved by averaging numeric values. Once GPS data was standardized to 5-minute intervals, there were 178 fewer rows because of collision handling.

Finally, 965 of the standardized fixes had DOP values above 10, and were removed as per the recommendations provided in (Lewis et al., 2007). The remaining 421,752 GPS fixes were kept for further analysis. The variables Altitude, Duration, DOP, Satellites and Cause of Fix were considered irrelevant and not used for classification. Temperature was also not used for classification but was used for studying the patterns of the unlabeled predictions.

The accelerometer data had only 102 missing values which appeared randomly distributed across cows and time. These missing observations all had values directly before and after, so they were imputed as the average between the two neighboring records. After this imputation, 2,121,240 accelerometer values were included in the study.

Regarding activity observation data, 32 observations (1.2%) with labels fighting, mineral, scratching and drinking in were re-designated as unlabeled for model training data due to insignificant representation, resulting in 2332 activity labels. The data column ‘Observer’ was considered irrelevant and was removed from the data set. 51 observations for cow 1021 were duplicated, and the second instances of these observations were dropped.

Feature Engineering: Two features derived from GPS data and two features derived from the accelerometer data were constructed. The GPS based features are the step size and turning angle. Step size is the displacement between two consecutive GPS fixes, and the turning angle

measures the change of heading between the previous GPS fix and the next GPS fix. Figure 3, adopted from (Michelot et al., 2016), illustrates the step size and turning angle features.

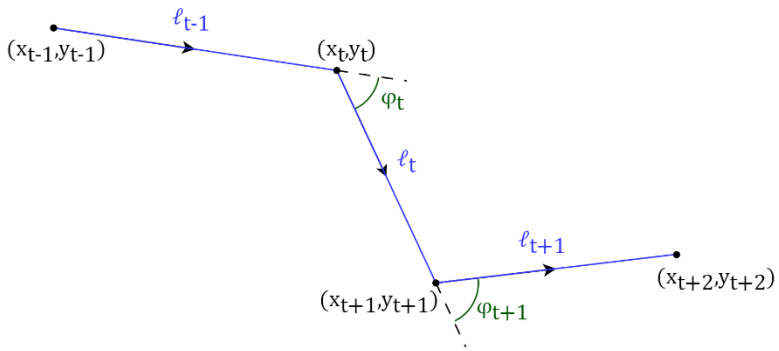

Figure 3. Representation of step size (ℓ_t) and turning angle (φ_t) used in momentuHMM.

Features derived from accelerometer data traditionally fall into two categories; orientation-dependent and orientation-independent (Versluijs et al., 2023). In this work we focused on orientation-independent features and used Signal Vector Magnitude (SVM) (Karantonis et al., 2006) in the classification models. Orientation-independent features ensure that the data remains valid regardless of positional shifts of the collars during the study, making behavior classification models more robust and reliable. Each accelerometer reading is first converted to its corresponding SVM using $\sqrt{x_i{}^2 + y_i{}^2 + z_i{}^2}$ where x_i, y_i, and z_i are the instantaneous static accelerometer readings of surge, sway, and heave. Since accelerometer data is recorded every minute while GPS fixes are recorded every five minutes, the five accelerometer readings preceding each GPS fix are used to compute two aggregate statistics; mean and variance. The mean SVM captures the average acceleration intensity over the past five minutes, while the variance of SVM reflects the consistency of acceleration intensity.

Table 4 summarizes the four features used in the classification models.

Table 4. Summary of features used in this study.

Feature	Formula	Description
Step Size	$l_t = \sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2}$	The displacement between the current and the next GPS fix where x is the longitude and y is the latitude recorded at time t.
Turning Angle	$\phi_t = \arctan2(y_{t+1} - y_t, x_{t+1} - x_t) - \arctan2(y_t - y_{t-1}, x_t - x_{t-1})$	The change in heading from the previous heading to the next heading.
Mean SVM	$\text{MeanSVM}_t = \frac{1}{5}\sum_{i=t-5}^{t} \text{SVM}_i$	The mean SVM of the previous 5 SVM values.
Var SVM	$\text{VarSVM}_t = \frac{1}{5}\sum_{i=t-5}^{t} (\text{SVM}_i - \text{MeanSVM}_t)^2$	The variance of the previous 5 SVM values.

Label data aggregation: To align activity labels with GPS readings, we aggregated the 2332 minute-by-minute labels into 487 5-minute windows, assigning the last recorded activity label to represent the entire window. We chose the last-label approach because it preserves temporal proximity between the behavioral state and the corresponding GPS reading, offering a more direct representation of the animal's activity at the time of location recording and it maintained similar proportions of the three target behaviors compared to the original distribution, at 42.2% grazing, 48.2% resting and 9.6% traveling.

The remainder of this thesis is organized as follows. Chapter 2 presents the implementation of Hidden Markov Models (HMMs), including classification performance on labeled data and exploratory analysis of predictions on unlabeled data. Chapter 3 follows a similar structure but focuses on Long Short-Term Memory (LSTM) networks. Chapter 4 offers a comparative discussion of the two serially dependent models, evaluating their predictions and assessing how

well the unlabeled data align with established behavioral patterns in the literature. Finally, Chapter 5 outlines the conclusions, limitations, and potential directions for future research.

HIDDEN MARKOV MODELS

Hidden Markov Models, or HMMs, are tools for modelling stochastic processes. HMMs have been used since the late 1960s in various fields such as speech recognition (Rabiner, 1989), gene sequence modelling (Soruri et al., 2013), pattern matching (Yamato et al., 1992), and more (Mor et al., 2021). The reason HMMs are used in such a wide range of applications is their flexibility and explainability. HMMs can be used to model just about any system which has patterns of change, and are governed by, in the most general cases, a small number of parameters which can have very tangible meanings. In the field of PLF, HMMs have been used for behavior classification, but are not used with the same ubiquity as other tools. (Riaboff et al., 2022) provides a systematic review of 66 papers in cattle behavior classification using accelerometer data. The field is dominated by Supervised Machine Learning models (56%) such as Support Vector Machines (Tatler et al., 2018), Linear Discriminant Analysis (Giovanetti et al., 2017) and Quadratic Discriminant Analysis (Barwick et al., 2018). The second most common approach uses Supervised Ensemble ML techniques (18%) like Random Forests (Brennan et al., 2021) and XGBoost (Riaboff et al., 2020). Few of the papers in this field (3%) apply statistical models such as Logistic Regression (Roux et al., 2019) or HMMs (Leos-Barajas et al., 2017). In the next section, the most closely related works where HMMs are used to classify animal behavior based on GPS and accelerometer data are reviewed.

Literature Review

The search query ("cattle behavior" OR "cow behavior" OR "bovine behavior" OR "cattle behaviour" OR "cow behaviour" OR "bovine behaviour" OR "cattle activity" OR "cow activity" OR "bovine activity") AND ("classification" OR "recognition" OR "detection") AND ("Hidden

Markov Model" OR "HMM") on Google Scholar yielded 244 results, 10 of which were duplicate listings. After filtering to exclude conference papers, dissertations, and unpublished works, duplicate listings, datasets, and papers not available in English, there are 136 results. Many of the papers did not apply HMMs in their methodology. Once filtering out papers which did not apply HMMs, there were 24 results. Three of those papers were in completely different finds such as machine tool wear or atmospheric sciences. Of the 21 candidate papers, four were not based on cattle, 12 of the studies were conducted indoors, and nine were focused on data from fundamentally different sensors such as microphones, cameras, and pedometers. Only three papers (Guo et al., 2009; Scriban et al., 2024; Williams et al., 2017) were left after filtering. All three of these papers used unsupervised HMMs whereas we used supervised HMMs. Nevertheless, these three works are considered highly applicable to our study, meaning they apply HMMs for behavioral classification on GPS or accelerometer data from outdoor cattle. The first three studies listed are the highly related works, and the next three studies apply HMMs for behavioral classification of other animals.

(Guo et al., 2009) studied six cows equipped with GPS sensors and accelerometers over four days. The authors use linear and angular velocity features to predict behavioral labels. Two models are used in tandem. The authors identify six “stay” regions (places where cows frequent), and HMMs are used to classify behavior inside of the stay regions. Slow moving intervals are labeled as 1, slow moving intervals with high angular velocities were labeled 2, and fast-moving intervals were labeled 3. The other model, a Long-Term-Prediction model, is used to forecast cow paths between stay regions. No labeled behavior data was collected, but the pasture was studied to identify stay regions. The authors report that using HMMs for behavior detection proved feasible and effective.

In (Williams et al., 2017), grazing, resting, and walking behavior of 24 dairy cows was studied using GPS sensors that transmitted coordinate fixes every 5 seconds. From this data, the authors computed velocity, heading, acceleration, and 40 other features as described in (Williams et al., 2016). The data was grouped into 160-second sequences for each cow. Additionally, a discrete emission matrix was used where emissions were categorized by a set of rules into one of three emission levels. The authors employed a "one-per-sequence" HMM, where a behavioral label was assigned to each 160-second sequence. The model was trained and tested on 51 hours of observation data and achieved a 94% accuracy rate in predicting the dominant behavior within each sequence.

(Scriban et al., 2024) conducted a study on GPS data from cattle collected in West Africa over 2.5 years. A total of 21 animals from 19 different herds were equipped with GPS collars which each collected fixes every 30 minutes. Data from some collars was excluded due to hardware issues or the swapping of collars. In total, there were 560,490 GPS fixes in the cleaned dataset. The moveHMM R package was used to compute step size and turning angle of the animals. A normal distribution was used for the step size, and the Von Mises distribution was used for the turning angles, and temperature recorded by the GPS collars was used as a covariate for the transition matrix. As the HMM was unsupervised, the authors tested the number of states the model should have and found three states to be optimal. The predicted unsupervised states were labeled as resting, foraging, and traveling. The authors find that on average the cows spend 47% of the day resting, 37% of the day foraging, and 16% of the day traveling. A diurnal grazing pattern is clearly observed in the cold season and dry season but not observed during the rainy season. The animals tend to travel much further from their home village during the rainy season than the other two identified seasons. The authors note that a limitation of their research is that the three states

are not actually based on real behavior, but are inferred from the patterns; i.e. state 2 was labeled as traveling because the step distance tended to be longer. If labeled data had been collected and applied, these states could be more representative of the behaviors they are trying to capture.

(Leos-Barajas et al., 2017) highlight that most analyses of accelerometer data in ecological studies tend to ignore the temporal dependence inherent in such measurements. HMMs are presented as a common remedy to address this limitation. A key contribution of this work is the application of a joint modeling approach, in which uncertainty from the HMM is carried forward into a logistic regression model which allows for evaluating the relationship between covariates and the predicted states. The study demonstrates that HMMs can capture and explain temporally dependent behavioral patterns, offering insights that may be missed by models that assume independence across observations.

In (Rautiainen et al., 2022) the authors implemented a methodology for studying captive reindeer in northern Sweden. The researchers used collar-mounted accelerometers (sampling at 10 Hz) to collect movement data and video recordings to label behaviors during observation periods. To classify reindeer behavior, the authors evaluated Random Forests, Support Vector Machines, and HMMs. When compared to the other models, HMMs performed best for classifying animal behavior using a leave-one-out cross validation (LOOCV) scheme with an accuracy of 82%.

In (Beumer et al., 2020) the authors studied the behavior of muskoxen in northeast Greenland using hourly GPS fixes from 19 tracked animals. The authors used third party data sources such as terrain ruggedness and snow depth, alongside features calculated from the GPS data. The study applied an unsupervised HMM to classify muskox behavior into three states: resting, foraging, and relocating. Since no labeled data was available, domain experts estimated initial HMM parameters. The primary objective was to examine seasonal variation in behavior,

and the animals were tracked over multiple years to support longitudinal analysis. The authors noted that some of the covariate data such as wind speed or snow depth, which they expected to have significant impact on the animal's behavior, was not as influential as high-fidelity data such as ruggedness and land cover. This was attributed to muskox being a well-adapted arctic animal, or more likely the low fidelity and low frequency of the weather data which was based on weather patterns in the region was not necessarily representative of the local conditions experienced by the muskox.

Methods

Table 5 defines the key variables and parameters of HMMs. An HMM has two levels. The first level consists of the hidden states. The hidden states resemble a classic Markov Chain which satisfies the Markov Property. These hidden states exist in a set of size N. Each X_t must exist in said set, and transitions between hidden states are governed by the state transition matrix A. The initial state distribution is denoted by π, which represents the likelihood of starting in each hidden state. The second level consists of observations. At each time step t, an observation y_t is emitted according to the emission probabilities B, which in our case is a collection of continuous distributions, depending on the hidden state x_t. In Figure 4, the anatomy of an HMM is shown. Each y variable represents the emitted feature values at a given time t. Each y_t depends on the current hidden state, x_t. The current hidden state x_t is dependent on the previous state x_{t-1}.

Table 5. Definitions of variables used in classical HMMs.

Notation	Description
N	Number of unique hidden states
T	Length of the sequence (number of time steps)
$X = x_1, \dots, x_T$	Sequence of T hidden states.

Table 5. Continued.

Notation	Description
$Y = y_1, \ldots, y_T$	Sequence of T emitted observations where each y_t stores values of the four features.
$A = [a_{ij}]$	State transition probability matrix: a_{ij} is the probability of transitioning from state i to state j.
$B = b_i(y_t)$	Emission probability vector: $b_i(y_t)$ is the probability of observation y_t being emitted by state i. As y_t is multivariate, each b_i is defined as a product of univariate distributions across features.
π	π Represents the probability that a chain starts in state i. $\pi_i = P(X_1 = i)$
$\lambda = (A, B, \pi)$	Complete parameter set of the HMM.

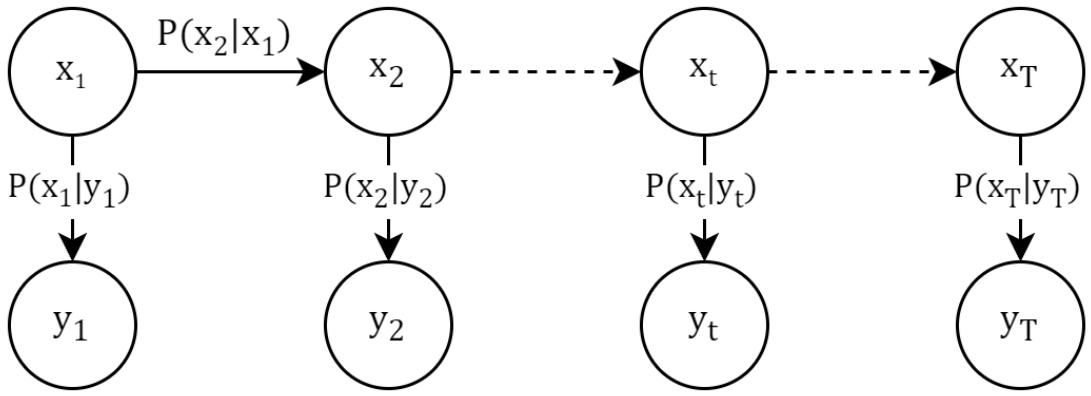

Figure 4. A diagram of states (X's) and emissions (Y's) in an HMM.

Model Objective, Implementation and Evaluation

The objective of the behavior classification model is to classify each 5-minute period as one of grazing, resting or traveling behaviors. We used the R package Maximum Likelihood Analysis of Animal Movement Behavior Using Multivariate Hidden Markov Models (momentuHMM) (McClintock and Michelot, 2018) to implement the HMM.

Leave-one-out Cross Validation (LOOCV) (Geroldinger et al., 2023) is used to evaluate the classification model's performance. LOOCV is particularly suitable for animal sequence

analysis as it evaluates the model's ability to generalize to an individual animal excluded from the training data. In this approach, the model classifies the entire sequence for the left-out animal, providing insight into the model's performance on unseen subjects. We evaluated the model's performance using accuracy, specificity, sensitivity, and the F1 score (Sokolova and Lapalme, 2009).

Model Inputs: The momentuHMM package requires several key inputs to fit HMMs to animal movement data. The primary input is a data frame containing location data, including animal identifiers, timestamps, and GPS coordinates. Based on this data, step length and turning angle features are calculated. Additionally, two features, MeanSVM and VarSVM, are computed using windowed accelerometer data. Each of these features requires specified emission probability distribution when generating the HMM model. Due to the heteroskedastic nature of VarSVM, a log transformation was performed to improve the quality of fits. Finally, the hidden behavioral states are specified as Grazing, Resting, and Traveling.

Figure 5 shows ridgeline plots of the raw input data categorized by behavioral states and by cow. Missing ridges correspond to behaviors which were not observed for an individual cow. For instance, Cow ID 996 was never observed grazing, and hence has no labeled data for the grazing state. These plots show clear patterns for some feature and behavioral state combinations as well as deviations from these patterns in the training data. For example, Cow ID 1006 shows atypical variation in Step Size, MeanSVM, VarSVM, and turning angle compared to the other cows for the resting state in the labeled dataset.

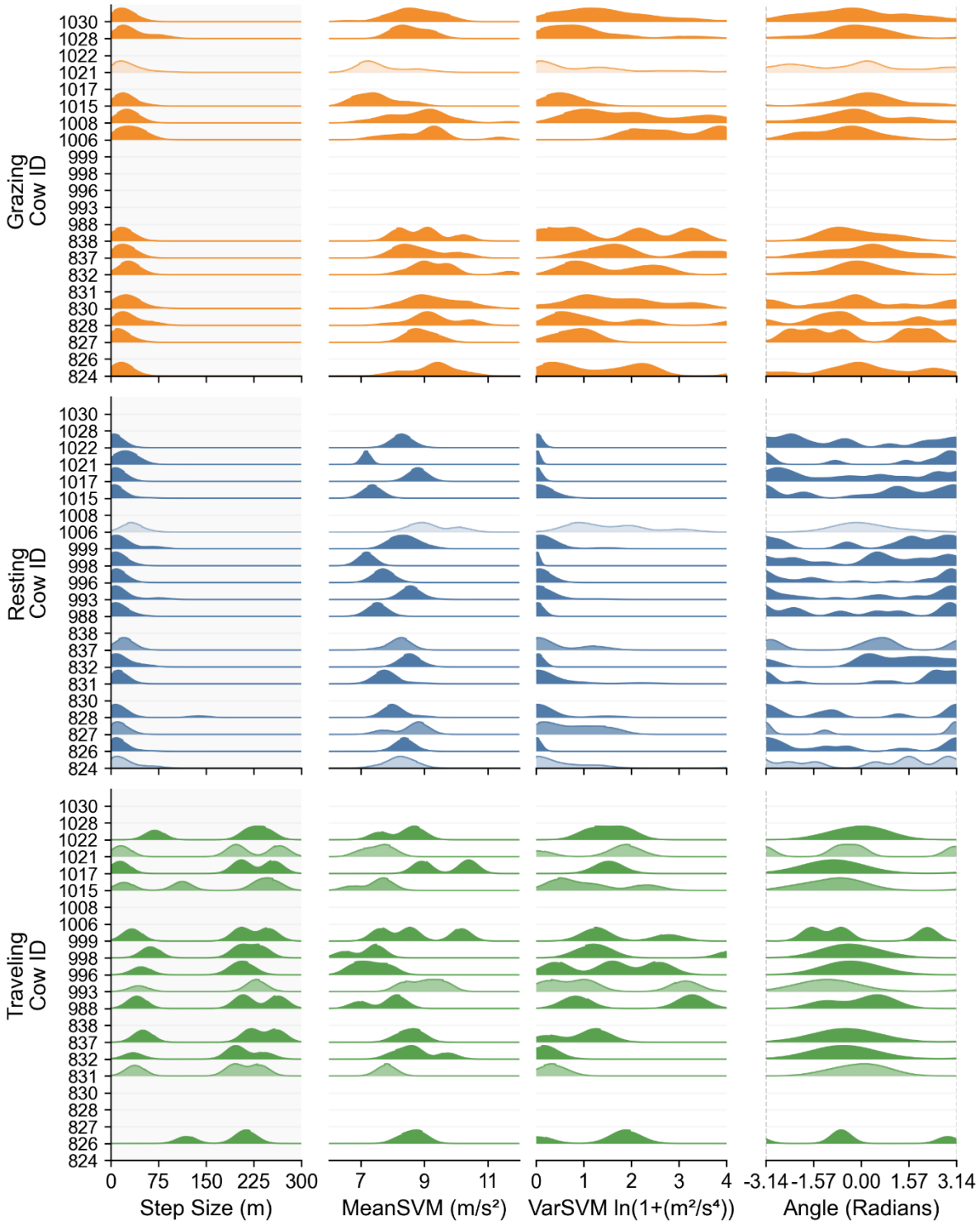

Figure 5. Ridgeline plot of the labeled data used to train the HMM by behavior and by cow.

The R package fitdistrplus (Delignette-Muller and Dutang, 2015) was used to evaluate the quality of fits for emission probability distributions for each feature. For the Step Size, MeanSVM

and VarSVM features, Exponential, Gamma, Normal, and Lognormal distributions were considered. For the circular feature Turning Angle, the Von Mises distribution and the Wrapped Cauchy distribution were evaluated. The quality of fit was evaluated using the log-likelihood statistic (Zucchini et al., 2017), with higher (less negative) values indicating a better model fit, as shown in Table 6. The best fitting distribution, namely Lognormal for Step Size and MeanSVM, Gamma for VarSVM, and Wrapped Cauchy for the Turning Angle feature were used as inputs into the model.

Table 6. Quality of distribution fits by feature.

Feature	**Lognormal**	**Gamma**	**Exponential**	**Normal**	**Von Mises**	**Wrapped Cauchy**
Step Size	-2107.54	-2241.7	-2177.57	-2620.1	NA	NA
MeanSVM	-621.52	-625.99	-1524.57	-638.73	NA	NA
VarSVM	-1038	-161.56	-494.66	-231.38	NA	NA
Turning Angle	NA	NA	NA	NA	-892.4	-891.72

Figure 6 displays the cumulative distribution functions (CDFs) of the fitted distributions for each behavioral state and feature. The slope of the CDF is the probability density function (PDF). Figure 6a shows the step length feature with lognormal distribution fitted to the labeled data by state. Grazing and resting have steep slopes for short distances, indicating that it is likely for those states to emit short step lengths. Conversely, traveling has a more gradual slope, indicating that traveling states are more likely to emit longer step lengths. In Figure 6b, we see that resting has the sharpest ascent for MeanSVM. This indicates that the lowest SVM values typically correspond to the resting state, while higher SVM values are more likely to be emitted by grazing and traveling states. In Figure 6c, we see that resting typically corresponds to very low VarSVM values. Traveling and grazing behaviors have similar CDFs and we see that grazing cows tend to

have the largest variability in magnitude measured by the accelerometers. The turning angle feature shown in Figure 6d is represented as a PDF rather than a CDF due to its circular nature. Visual inspection reveals distinctive patterns across behavioral states. Traveling cows predominantly maintain forward movement, while resting cows exhibit a higher probability of 180-degree turns, potentially attributable to GPS measurement error around stationary positions. Grazing cows display intermediate turning behavior, with fewer sharp turns than resting cows but more directional changes than traveling cows. These differential distributions provide a robust basis for state classification based on movement parameters.

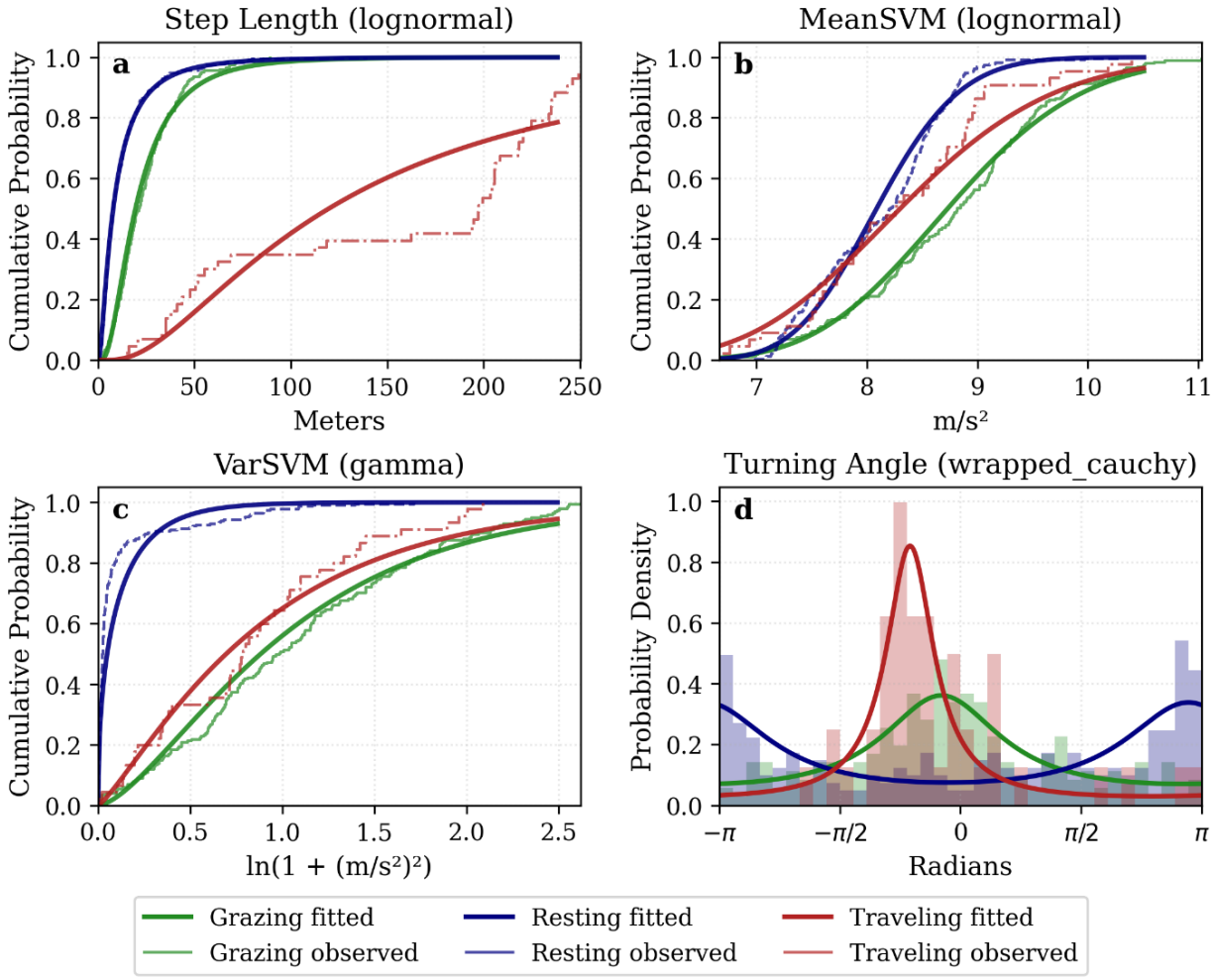

Figure 6. CDFs of non-circular distributions, and the PDF of the circular distributions overlayed on raw training data.

Model Development: The first step in model development is solving the learning problem. Using LOOCV, data for one cow is held out and the data from the remaining 21 cows is used to calculate the initial distribution parameters for each feature and state combination. The fitHMM function in momentuHMM uses the base R function nlm—a Newton-type nonlinear minimization algorithm—to determine the parameters of the HMM, including the initial probabilities of starting in each state, the transition probabilities, and the emission distribution parameters to maximize the likelihood that each sequence of cow observations are emitted by the sequences of labeled states.

The second step involves solving the decoding problem where the Viterbi Algorithm (Zucchini et al., 2017) and the trained HMM are used to determine the sequence of hidden states (behaviors) which maximize the likelihood of the left-out cow's emissions. These two steps are repeated 22 times, each time holding out a different cow. The predicted and actual observed labels are then compared to determine the accuracy of the 22 HMMs. A final HMM is trained on all 22 cows' labeled data and used to predict the most likely sequences of cow activities for the entire 67 day study period.

Results

Classification Accuracy: The 22 HMM models created using LOOCV achieved an average of 92.72% classification accuracy over the three behaviors of interest. The standard deviation for accuracy across cows was 0.057, indicating a relatively tight clustering around the mean. In fact, only two cows' classification accuracy was lower than 85%, at 61.11% and 64.52% showing a significant difference from the rest. Table 7 shows the confusion matrix, and Table 8 presents classification accuracy metrics, including overall accuracy, precision, specificity, sensitivity, and F1-score for each class. The majority class, Resting, was identified with an accuracy of 94.4%, while Grazing and Traveling achieved accuracy of 93.6% and 97.5%, respectively. Traveling exhibited the highest specificity (0.991), indicating that non-traveling behaviors were rarely misclassified as Traveling. However, its sensitivity (0.826) was lower than that of the other classes, suggesting that true instances of Traveling were sometimes misclassified as either Grazing or Resting. Resting showed balanced performance with precision (0.952) and sensitivity (0.931), resulting in the highest F1-score (0.941) among all classes. Grazing had slightly lower precision (0.906) and specificity (0.928) but maintained strong sensitivity (0.946) and an F1-score of 0.925, indicating consistent model performance.

Table 7. Confusion matrix of LOOCV results from the HMM.

Predicted	Reference		
	Grazing	**Resting**	**Traveling**
Grazing	192	13	7
Resting	10	216	1
Traveling	1	3	38

Table 8. HMM Classification performance metrics by class.

State	Actual	Predicted	Accuracy	Precision	Specificity	Sensitivity	F1
Grazing	203	212	0.936	0.906	0.928	0.946	0.925
Resting	232	227	0.944	0.952	0.956	0.931	0.941
Traveling	46	42	0.975	0.905	0.991	0.826	0.864

Cattle Behavior: The final HMM model applied to the data collected throughout the study period produced insightful results that further strengthened the confidence in both the model's effectiveness and the overall modeling approach. The following sections present these findings.

Daily Grazing Behavior: Figure 7 presents a heat map showing the percentage of each day spent grazing for each cow as classified by the HMM model. Days when the cows were corralled for weighing—January 25–26 and February 22–23—are excluded from the figure, as they do not reflect typical grazing behavior. On average, cows spent 8.5 hours per day grazing with a standard deviation of 1.8 hours. The minimum and maximum time any cow was predicted to have spent grazing in a day was approximately 1.3 and 16.3 hours respectively. The heatmap highlights considerable variability in daily grazing time. For instance, overall grazing appears reduced between March 8th and 11th, while days such as March 19th and 21st show increased grazing activity. Additionally, the average time spent grazing per day was not uniform across the animals. Some animals grazed approximately 4% less than average, while others grazed 3% more than

average. The variability across days was more extreme than the variability across cows for the average time spent grazing. The difference in the most intense grazing day and the least intense grazing day was 24%.

When the final model is applied to the entire dataset, distinct diurnal patterns emerge, as illustrated in Figure 8. The figure presents two polar plots depicting the predicted daily activities of two representative cows throughout the study. In these plots, the study days are represented as concentric circles, with the innermost circle corresponding to the first day and the outermost circle to the last day. The model reveals that grazing behavior predominantly occurs before sunset and after sunrise, while resting behavior is more frequent during midday and nighttime.

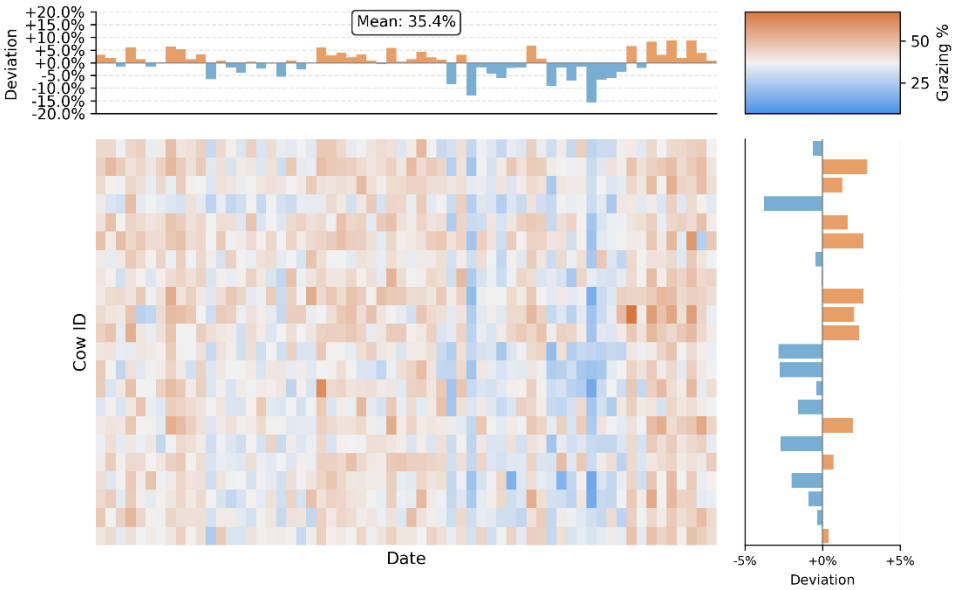

Figure 7. Heatmap of percent of day grazing by cow and by date through the study.

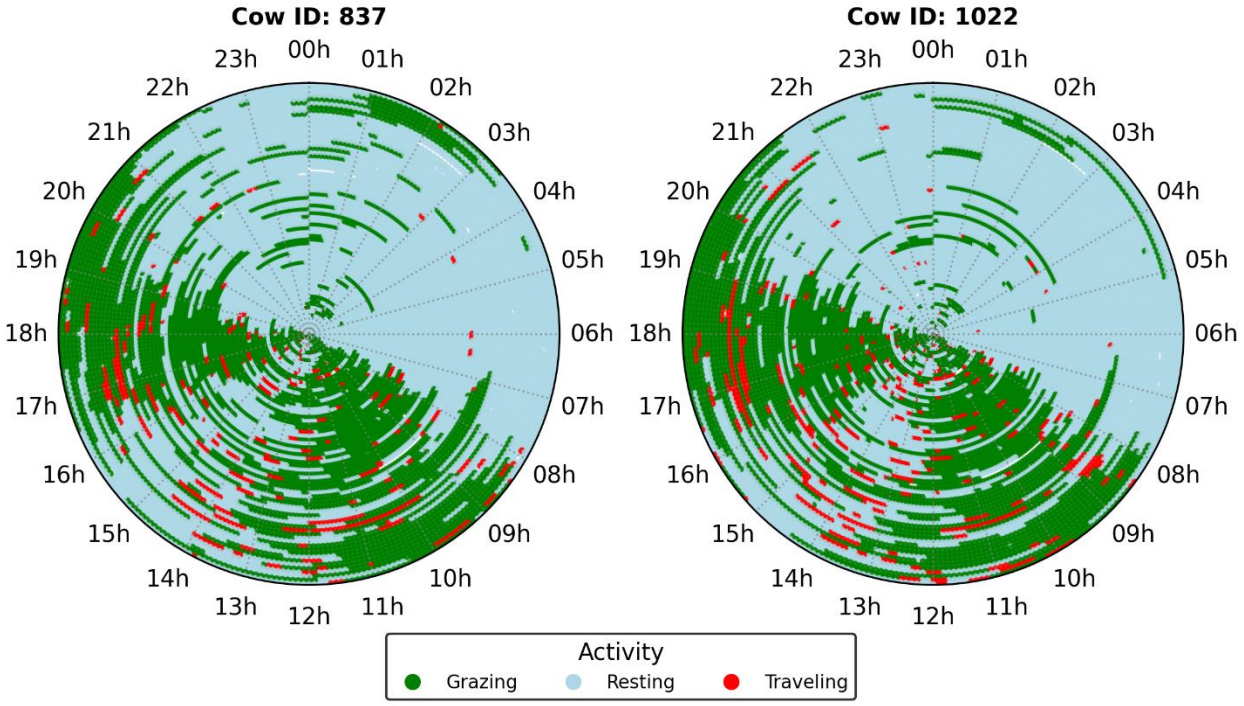

Figure 8. Two sample polar plots of cow activity over the study.

To better understand the herd's grazing behavior over time, Figure 9 displays a 3D surface plot illustrating how the proportion of time spent grazing varies with date, time of day, and temperature. This smoothed surface highlights key interaction effects across these dimensions, offering insight into temporal and environmental influences on grazing patterns. The previously observed diurnal pattern is evident, with peak grazing occurring around 10:00 and 18:00. Between these peaks, there is a clear dip in grazing activity, presumably when cows travel to water sources to cool down before resuming grazing. Another set of grazing peaks occurs around 02:00 on January 17th, February 16th, and March 18th. Interestingly, these dates align with full moon phases, suggesting that nocturnal grazing activity may increase under full moonlight. Temperature also appears to influence grazing behavior. Diurnal peaks colored blue (indicating colder conditions) are noticeably lower, suggesting reduced grazing in cooler temperatures. In contrast,

red-colored peaks (warmer conditions) are generally higher, implying increased grazing when it's warm.

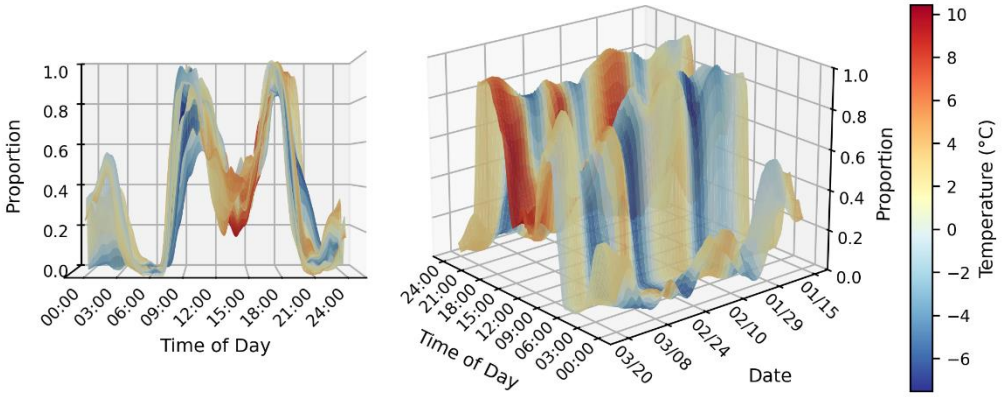

Figure 9. 3D Surface plot of the proportion of cows grazing at each date and minute combination colored by temperature smoothed with Gaussian smoothing.

Nighttime grazing behavior:To explore whether moon phases—and consequently, nighttime light levels—affected nighttime grazing behavior, we analyzed grazing patterns based on the moon phase using the Python package PyEphem (Rhodes, 2011) which provides precise data on the moon's illumination percentage. Nights where illumination was greater than or equal to 95% were considered full moon, and illumination less than 95% were considered as non-full moon. Nighttime was set to start 90 minutes after sun set and end 90 minutes before sunrise. Figure 10 shows box-whisker plot of full vs non-full moon predicted nighttime grazing percentages for each individual cow. When the overall predicted nighttime grazing percentage during full moon nights (n=286, mean=17.1%) is compared with non-full moon nights (n=1210, mean 9.9%) we found a statistically significant ($p<0.001$) difference using the independent samples paired t-test (Zimmerman, 1997). This suggests that increased moonlight is associated with elevated nighttime grazing activity.

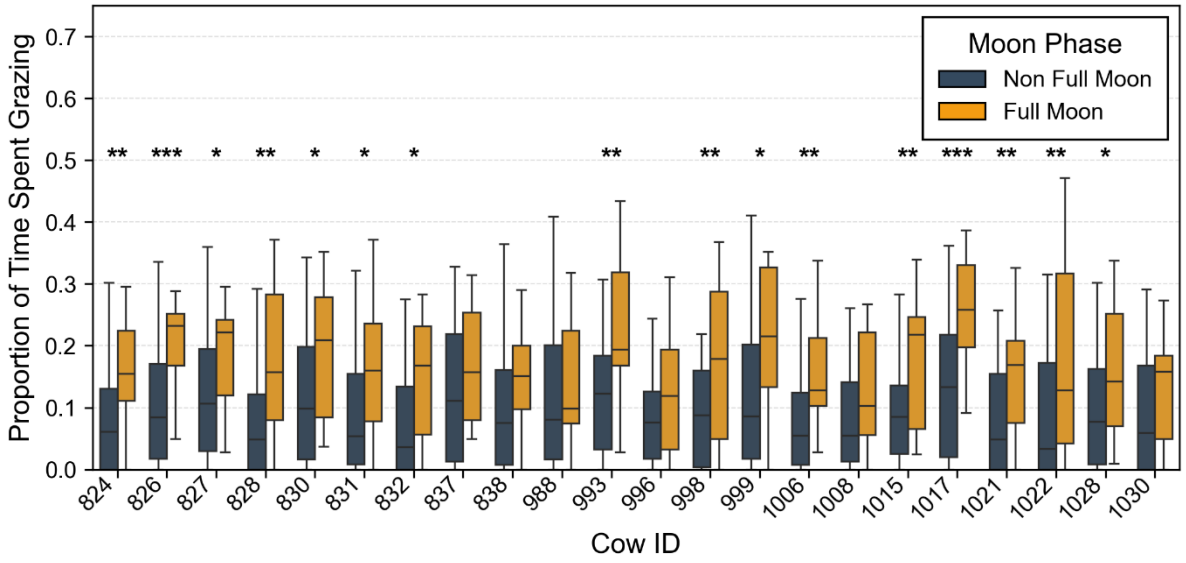

Figure 10. Comparison of nighttime grazing patterns during full moon and non-full moon nights for individual cows. (*** : $p<0.001$, ** : $p<0.01$, * : $p<0.05$).

Effect of Time of Day and Temperature on Cattle Behavior: To better understand the interaction of diurnal patterns and temperature during the day, we used General Linear Mixed Models (GLMMs) with binomial link functions (Bolker et al., 2009). We constructed three GLMMs, one corresponding to each of three behavioral states: grazing, resting, and walking. Time of Day and Temperature and their interactions were considered fixed effects, while each cow was considered a random effect. Time variable represents the minute in the day, and ranges from 315 to 1270 minutes (approximately between 5am and 9pm). Temperature readings from the GPS sensors ranged between -29 and 26.5 °C, with an average of 0.79 °C.

To capture the bimodal daily activity patterns of cows, orthogonal polynomial terms up to the fourth degree for Time of Day are used. Interaction terms between Temperature and the first- and fourth-degree Time of Day polynomial term were included to capture the effects of temperature on the diurnal grazing patterns. All continuous predictor variables were standardized, and binary response variables were created using one-hot encoding. Variance Inflation Factors (VIFs) for fixed effects were all below 2, indicating negligible multicollinearity (Zuur et al., 2010).

Each model predicts the probability of a cow being in a given state, as defined by the linear predictor in Eqn. (1). The probability of the behavioral state is obtained using the inverse logit transformation shown in Eqn. (2). The GLMMs were implemented with the pymer4 library (Jolly, 2018) which is a Python wrapper of the R package lme4 (Bates et al., 2015).

$$Y_{state} = \beta_0 + \beta_1 temp + \sum_{i=1}^{4} \beta_{i+1} time^i + \beta_6 (temp \times time) + \beta_7 (temp \times time^4) + u_{ID} \quad (1)$$

where $u_{ID} \sim N(0, \sigma^2{}_{ID})$

$$P(Y_{state} = 1) = (1 + \exp(-Y_{state}))^{-1} \quad (2)$$

Table 9 presents GLMM model parameter estimates. While the effects of parameters are statistically significant—partially due to our large dataset (221,242 observations)—the ecological interpretation and practical implications are of primary importance and are best illustrated using response curves shown in Table 9. Continued.

Parameter	**Grazing**		**Resting**		**Traveling**	
Time^2	2.471	***	-2.504	***	-0.359	***
Time^3	1.506	***	-1.467	***	0.07	***
Time^4	-1.239	***	1.274	***	-0.121	***
Temperature	-0.2469	***	0.3397	***	-0.2098	***
Temperature x Time	-0.267	***	0.336	***	-0.039	**
Temperature x Time^4	0.171	***	-0.185	***	0.002	(NS)

Figure 11. The response curves show the probability of being in each behavioral state by Time of Day and Temperature. Temperature curves are evaluated at five equally spaced temperatures spanning $\mu \pm 2\sigma$ of the observed data. We observe that on colder days cows tend

to start and stop grazing later and are more likely to continue grazing through the middle of the day. Alternatively, on warmer days, cows start and stop grazing earlier with a more pronounced break from grazing in the middle of the day. The afternoon grazing peak is also slightly higher than the morning peak, indicating that cows may be more likely to graze in the afternoon than in the morning. We also see that cows appear to travel less as temperatures increase.

Table 9. Fitted GLMM parameter estimates and the significance of the measured effect.

Parameter	**Grazing**		**Resting**		**Traveling**	
Intercept	-0.282	***	-0.085	*	-2.428	***
Time	-0.3951	***	0.3456	***	0.1118	***
Time^2	2.471	***	-2.504	***	-0.359	***
Time^3	1.506	***	-1.467	***	0.07	***
Time^4	-1.239	***	1.274	***	-0.121	***
Temperature	-0.2469	***	0.3397	***	-0.2098	***
Temperature x Time	-0.267	***	0.336	***	-0.039	**
Temperature x Time^4	0.171	***	-0.185	***	0.002	(NS)

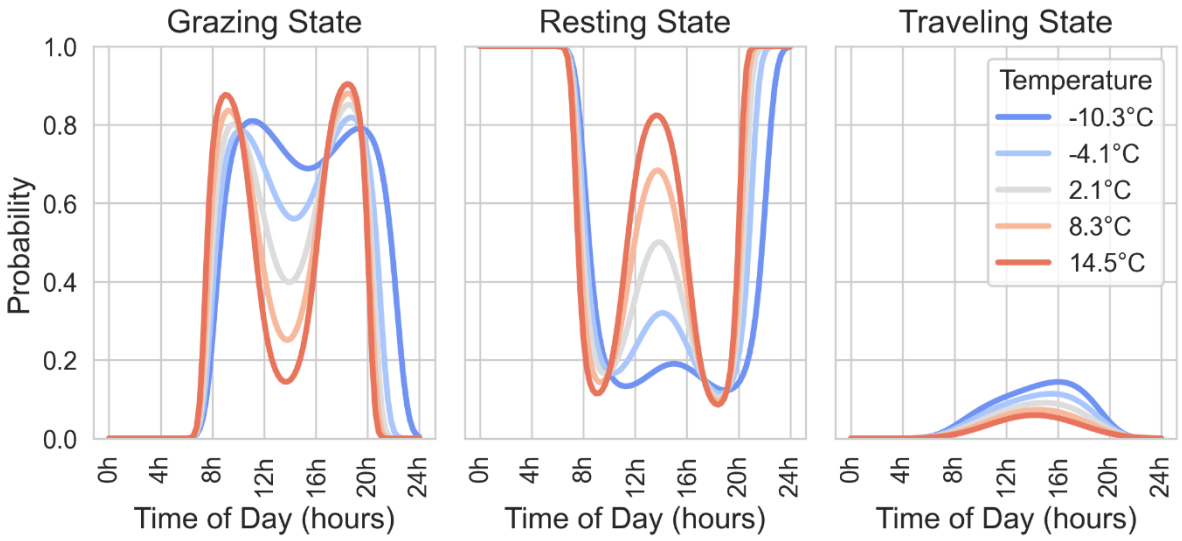

Figure 11. Probability of activities by time of day (x axis) and temperature (color).

In this chapter, we described the structure, cross-validation results, and validation of predictions from Hidden Markov Models. The HMM approach has very appealing characteristics, such as logical and interoperable transition matrices and emission distributions. We find parameters which fit the training data well and appear to generalize to the rest of the data in a realistic fashion. However, each state in our HMM only has dependencies on the immediately previous state and the current state's emissions. This may limit the model's ability to find long-term patterns in the data. To address this limitation, we next examine Long Short-Term Memory (LSTM) neural networks, which are known for their effectiveness in capturing both recent and distant temporal dependencies in sequential data.

LONG SHORT-TERM MEMORY NEURAL NETWORKS

Serially dependent neural networks such as recurrent neural networks (LeCun et al., 2015) and transformers (Zhang et al., 2024) have proven capable of learning patterns in classification problems (Ismail Fawaz et al., 2019). There is typically no easily interpretable meaning of the parameters in a neural network, but through abstraction they gain flexibility. A neural network will contort itself to learn patterns in data and become highly effective at predicting data similar to what it was trained on. One simple, common type of neural network is called a Multi-Layer Perceptron (MLP). MLPs are fundamental networks upon which LSTMs are built.

Before introducing LSTMs, a simple example scenario using an MLP is provided. Assume there are two features: temperature and humidity. If an MLP was constructed to do binary classification of whether or not it is raining, the network could look like Figure 12 which, in this case, happens to have one hidden layer of size three. This network is small for simplicities sake, but the network could plausibly have many layers, each with more or fewer nodes than the example. If the batch size is four, meaning four observations are processed simultaneously, then $\boldsymbol{X}$ contains the four independent measurements of temperature and humidity. Each row in $\boldsymbol{X}$ is an independent observational unit. Let n_1 correspond to the first column of $\boldsymbol{X}$, temperature, and n_2 correspond to the second column of $\boldsymbol{X}$, humidity. Additionally, the layers of the network are denoted with L. Let $L = 0$ correspond to the input layer, $L = 1$ correspond to the network's singular hidden layer, and $L = 2$ the output layer. The arrows between the nodes represent weights ($\boldsymbol{W}$) which scale the effect that each previous node has on the current node, and the biases ($\boldsymbol{b}$) provide a default behavior for each node. Lastly, activation functions ($\boldsymbol{\phi}$) need to be defined. Common activation functions include the sigmoid and ReLU functions. Refer to (Goodfellow et al., 2016) for information on

activation functions and much more on neural networks. These activation functions squash each value in $\boldsymbol{a}$ to some defined range, such as [0,1]. Then, using (3), the hidden layer of this network and the output layer can be computed. The prediction of whether it is raining or not can be extracted from $\boldsymbol{Y}$. For each row in $\boldsymbol{Y}$, If the first column, which corresponds to n_6, is greater than the second column, $\boldsymbol{n_7}$, then the model predicts it is raining.

$$\boldsymbol{a}^{(L+1)} = \boldsymbol{\phi}\left(\boldsymbol{W}^{(L)} \cdot \boldsymbol{a}^{(L)} + \boldsymbol{b}^{(L+1)}\right) \qquad \text{where } \mathbf{a}^{(0)} = \boldsymbol{X},\ \ \mathbf{a}^{(2)} = \boldsymbol{Y} \tag{3}$$

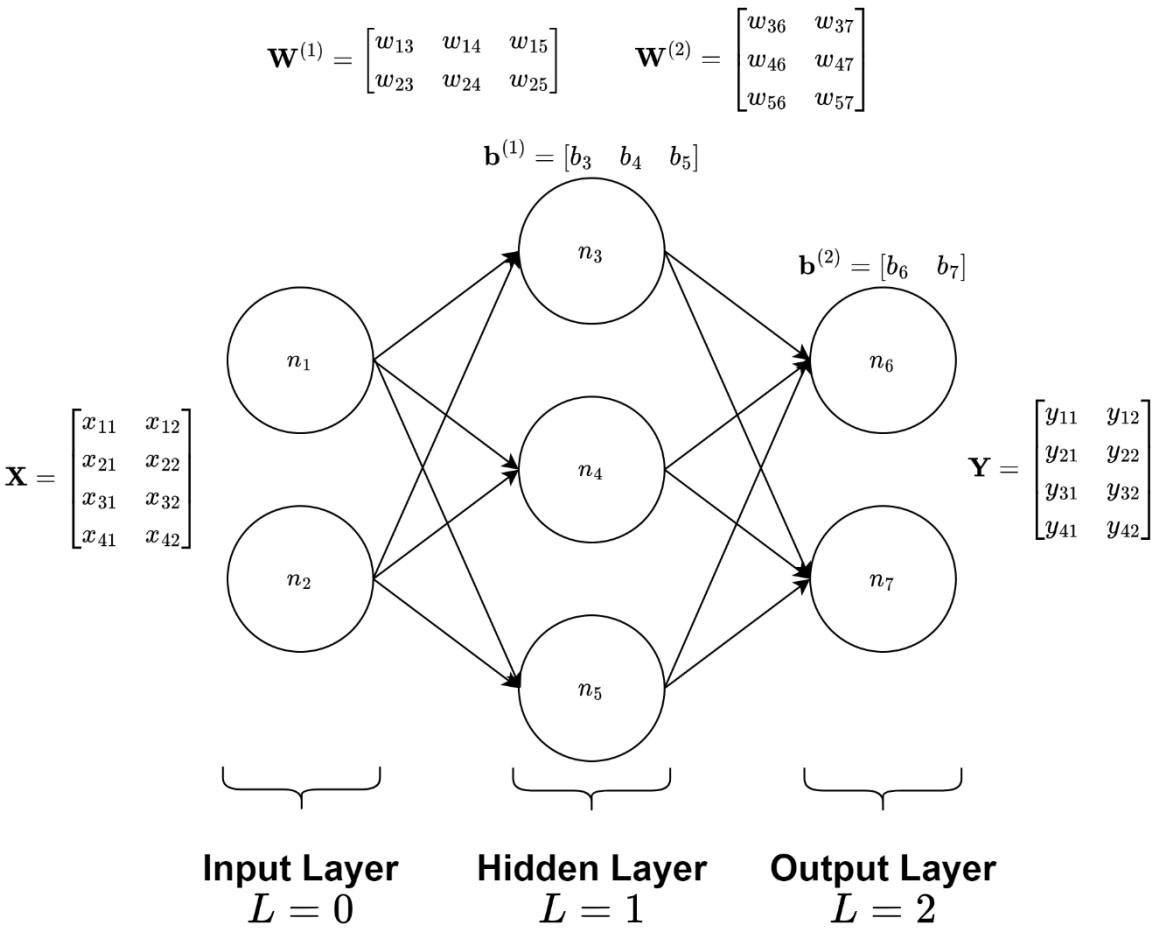

Figure 12. Example MLP to demonstrate the fundamentals of neural networks with inputs, weights, biases, and outputs shown.

While MLPs are powerful, they treat each observation independently. If the order of the rows of $\boldsymbol{X}$ were swapped around, there would be no difference in the predictions for each observation. In pursuit of serial classification, a serial dependence needs to be introduced. That is—if the model predicts rain in the first three observations, then the model should be more inclined to predict rain for the fourth observation. Conversely, if the model predicts no rain in the first three observations, then the model should be less inclined to predict rain on the fourth observation.

The standard layer in the neural network toolbox for handling serial dependence is the Recurrent Neural Network (RNN) layer. RNNs form the broader family of serially dependent neural networks, of which LSTMs are a part of. RNNs have special characteristics which enable this serial dependence. Instead of each row in $\boldsymbol{X}$ representing the independent observations of temperature and humidity, RNNs require each row in $\boldsymbol{X}$ to be a timeseries. In the MLP example above, $\boldsymbol{X} \in \mathbb{R}^{B\times F} \rightarrow \boldsymbol{X} \in \mathbb{R}^{4\times 2}$, where B is size of the batch and F is the number of features. Now for RNNs, $\boldsymbol{X} \in \mathbb{R}^{B\times T\times F} \rightarrow \boldsymbol{X} \in \mathbb{R}^{1\times 4\times 2}$ where the entire timeseries of four observations is now the independent observational unit instead of each observation. Many works refer to this size T as the window size. The values of the hidden state h_t, which can be thought of as the RNN analogous of $\boldsymbol{a}^{(1)}$ above, use iteratively compute the interaction of every measurement in the sequence.

Although recurrent neural networks (RNNs) address the limitations of MLPs by introducing temporal dependence through a hidden state, they suffer from a well-documented problem when modeling long sequences. During training, the contribution of early inputs to later outputs is gradually diminished due to repeated transformations of the hidden state. As a result, standard RNNs struggle to learn long-term dependencies because gradients either vanish to zero or explode uncontrollably. This limitation has led to the development of more advanced architectures, such as Gated Recurrent Units (GRUs) and LSTMs. LSTMs have memory cells with

their own parameters that are adjusted during training. Memory cells allow the LSTM to selectively remember what happened earlier in the timeseries and are an effective way to make the model focus on the important historical measurements instead of just the most recent measurements. In the following section, applications of LSTMs in cattle behavior classification are explored.

Literature Review

In the literature survey (Riaboff et al., 2022), only 8% of the identified behavioral prediction works on accelerometer data used deep learning. Of the 8% of papers applying artificial neural network (ANN) models, diverse types of ANNs including Multilayer Perceptrons, 1D+ Convolutional Networks, and Recurrent Neural Networks are used. Here we will look at works focused on RNNs and LSTMs.

In (Peng et al., 2019) studied LSTM classification of cattle behavior on inertial measurement unit (IMU) data. For this study, motion data from six Japanese steers were collected at a frequency of 20 Hz. Eight behaviors were classified in this study from raw 9-axis IMU data. The animals were held in a barn for the duration of the study, where cameras collected video data. The video data was used for two purposes—One for observers to label data, and two, for an input stream to the image classifier the LSTM was compared against. Three window sizes with 50% overlap were tested in this study. The authors found that the longer windows typically performed worse, especially for infrequent behaviors such as head-butting. Overall, the LSTM proved suitable for classifying the raw IMU data, as it outperformed the image classifier. The best LSTM model had a window size of 64 and a general accuracy of 88.7%, and a F1-score of 88.7%.

(Hosseininoorbin et al., 2021) published a work studying cattle behavior with LSTMs with time-frequency domain data. Ten cattle in Australia were equipped with tri-axial sensors and their

data was sampled at a frequency of 50 Hz. To filter the static effect of gravity and thermal noise from the data, a third order Butter-worth filter was applied. The data were then transformed to the frequency domain. Window sizes of 5, 10, and 15 seconds were tested. Nine behaviors were classified with the neural network. The authors evaluate their model performance with stratified k-fold cross validation (SCV) and LOOCV. Overall the TFD features seemed to outperform the standard time series data, with LOOCV F1-scores of 89.3% versus 73.55%.

(Wang et al., 2023) uses neural networks to classify cattle behavior from four different datasets. The number of animals monitored with ear or collar sensors per dataset varies between 8 and 19. The collar and ear tags sampled with a frequency of 50 and 62.5 Hz respectively. There were manually annotated behaviors which included subsets of grazing, ruminating, walking, resting, and other. Three of the datasets had four labels, the fourth dataset had three labels. Data was segmented into non-overlapping sequences of 256 records. The authors look at bidirectional and unidirectional gated recurrent units (GRUs) and LSTMs with 32, 64, or 128 nodes, and 1 or 2 layers. These models were compared against two CNNs which were used as baselines. In total, the authors had 26 models and four datasets for a total of 104 results. The authors find that GRUs typically perform the best, and more parameters typically corresponded to higher performance with the bidirectional, 2 layer, 128 node GRU having the highest performance across the datasets. Additionally, the models were evaluated in terms of computational cost. The authors conclude that the uni-directional GRU with a single layer and 64 nodes yields the best performance with respect to cost, which has comparable performance to the baseline CNNs, but is significantly less expensive to train and apply.

(Gao et al., 2023) published a work fusing Bidirectional Long Short-Term Memory (Bi-LSTMs) and Convolutional Neural Networks (CNNs) to classify eight behaviors from 50

Simmental cows in China. These cows were held in a barn with two pens. Cameras were set up around the area and video data was processed with four types of compound neural networks; MASK-RCNN, CNN-LSTM, EfficientNet-LSTM, and CNN-Bi-LSTM. The authors find that the CNN-Bi-LSTM had the best performance of day-time behavior with an accuracy of 94.3% and an F1-score of 94.1%, demonstrating a high level of capability for classification of a diverse set of behaviors.

(El Moutaouakil and Noureddine, 2024) examines the effectiveness of LSTMs for classifying cattle behavior from triaxial accelerometer data sampled at 25Hz. The dataset used is public (Ito et al., 2022). The authors filtered the dataset and constructed a multilayer LSTM based neural network and achieved high accuracies of 95.57%, and an F1-score of 94.87%.

Methods

Sequence Generation: For the LSTM network, each cow's timeseries is segmented into smaller sequences which helps mitigate extreme gradients during backpropagation. During preprocessing all the observations were already aligned to 5-minute intervals. This standardization is highly suitable for LSTMs which require fixed numbers of observations per sequence. On the occasion that records are missing, the sequences are padded and the neural network is set to ignore those padded values when updating.

Model Architecture: The LSTMs are implemented using the Python packages tensorflow (Abadi et al., 2015) and keras (Chollet, 2015). In general, there are many ways to tweak a neural network, and many of the design choices are vaguely arbitrary or hard to justify. As (Snoek et al., 2012) puts it, "tuning is often a 'black art' requiring expert experience, rules of thumb, or sometimes brute-force search". One key architectural design choice for sequence modelling is the

input-output (I/O) mapping strategy. In this analysis, two structures are compared: One-Per-Observation (OPO) and One-Per-Sequence (OPS).

In the OPO model (also known as many-to-many), every observation within a sequence is associated with a corresponding label, resulting in a one-to-one alignment between timesteps and predictions. Sequences are typically non-overlapping and of a fixed window size T. While computationally efficient, this OPO strategy results in early timesteps having limited historical context as the first timestep only has itself, the second has only two inputs, etc. while the last observations in each timeseries have more historical information to use as a basis for prediction.

For OPS networks (also referred to as many-to-one networks), each timestep is predicted by using a fixed number T of previous observations as the sequence. This means that all the timeseries are overlapping and each observation will be input into the model T times. This increases the computational resources required significantly but poses to allow greater yield from the history than an OPO network, as every prediction has a full history of size T as a basis. The difference of OPO and OPS architectures is shown in Figure 13. Additionally, the LSTM layer is shown. Each LSTM node in the diagram has T historical states which get turned on and off based on the logic learned in the memory cell during training.

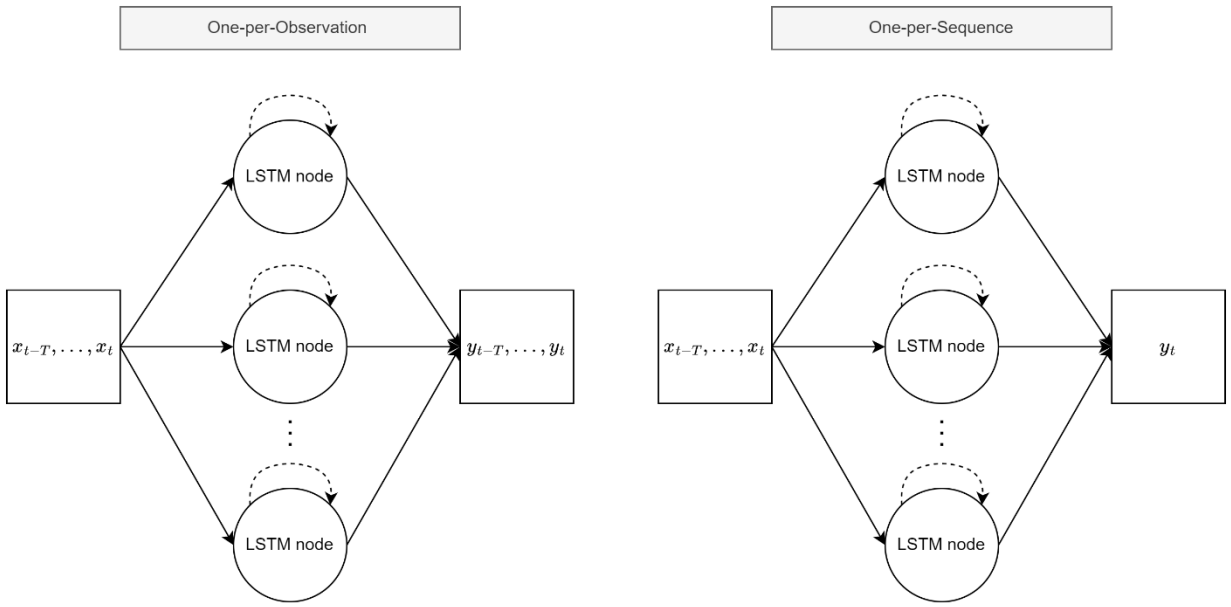

Figure 13. Diagram illustrating the two input-output strategies evaluated. Left: OPO, where each input in the sequence has a corresponding output. Right: OPS, where each output is generated from a full sequence of preceding inputs.

Additionally, the number of layers and the size of layers can be adjusted in the networks. (Reimers and Gurevych, 2017) studies these parameters and finds that, dependent on the problem, there are multiple solutions for the best number of layers. They tested one, two, and three layers for sentence prediction and classification tasks and recommend as a rule of thumb two LSTM layers for general performance. We use a single hidden LSTM layer in our models as increasing beyond a single layer did not markedly improve performance.

Parameter Selection: To select other parameters, Bayesian Optimization (BO) was used. This is a technique for optimizing expensive nonlinear functions. A search space S was defined with eight dimensions (Table 10). Because of the computational cost of LOOCV, we hold out two to three cows per cross validation fold instead of one which significantly reduces the time it takes for each iteration of BO. In the first ten iterations of BO, the performance is evaluated at random points in S. After the first ten iterations, an acquisition function is used to calculate points in S which are most likely to improve the objective value for the next 190 iterations. We use the

Expected Improvement (EI) acquisition function which guides BO to find parameters which maximize the improvement of the next objective value over the current best objective value (Snoek et al., 2012). The objective value is the negative micro F1 score (Lipton et al., 2014). The following paragraphs break down and explain each of the eight dimensions of S.

Sequence length is the number of 5-minute intervals which are used per input. If sequence length is 10, then 50 minutes of data are used as the predictor. If sequence length is 288, then a whole 24 hours is used as the predictor. Any integer between 10 and 288 was allowed as a valid sequence length during BO. Because the study intersected with a time zone shift, One day was "missing" an hour's data. To reconcile this with long sequence lengths which are capturing daily patterns, the day with missing records is padded with a masked value so each daily sequence starts at 00:05 and ends at 24:00.

Batch size is the number of input samples to process before committing to updating model parameters. As each sample in the batch goes through the model, a loss is computed. Loss is a continuous measure of how well the model is performing. Once the whole batch has been processed, all the batches' losses are aggregated into a mean loss, which gets used in backpropagation. Smaller batch sizes relate to more responsive training processes. Batch sizes were set to be any integer between 4 and 48 during BO.

The learning rate quantifies the magnitude of the update in BPTT. In (5), the learning rate is η. This is a multiplicative factor which scales the update. We use an exponential function (4) to change the learning rate η based on multiple factors; an initial learning rate η_0, the index of the current epoch i, the maximum number of epochs N, and the decay rate λ. Decay rate and initial learning rate are float values, their bounds are in Table 10. The number of decay steps is an integer value between 50 and 3000.

$$\eta_i = \eta_0 \lambda^{i/N} \quad (4)$$

The Clipnorm parameter sets a maximum value of the L2 vector norm of a weight tensor, which is the square root of the sum of the squared gradient values of a layer. If the L2 vector norm exceeds the Clipnorm value, then the gradient values are scaled so the maximum L2 vector norm is equal to the Clipnorm parameter value. This keeps the LSTM weights from exploding during the training process, which is commonly associated with an overfit model. Clipnorm was allowed to be any float value between 0.1 and 2.5, where 0.1 would be very defensive and 2.5 permits the parameters to update faster.

LSTM size and Dense size are the number of nodes in each respective layer. These are both categorical variables with 8, 16, 32, or 64 nodes allowed per layer. Larger layers typically correspond to more flexible networks which are capable of learning more about the relationship between inputs and outputs. As we have four input features and three output classes, our network does not benefit from extensive numbers of layers and nodes.

Table 10. Search space for the eight dimensions of BO

Dimension	**Data type**	**Prior**	**Minimum**	**Maximum**
Sequence Length	Integer	Uniform	10	288
Batch size	Integer	Uniform	4	48
Initial learning rate	Real	Uniform	1e-4	0.5
Decay steps	Integer	Uniform	50	3000
Decay rate	Real	Uniform	0.001	0.95
Clipnorm	Real	Uniform	0.1	2.5
LSTM Size	Integer	Uniform	8	64
Dense Size	Integer	Uniform	8	64

LSTM Training: With the search space, objective function, and acquisition function set, The model was set to run for 200 iterations. In each iteration, sequences were made for the labeled data based on the maximum sequence length. Then, sequences from two to three cows were set

aside as testing data. Sequences from the remaining 19 to 20 cows were used as training data. Approximately 30% of the training labels were separated and used as validation data.

A stratified splitting strategy was used to keep the labels within the training and validation data balanced. Not only did we aim to get 30% of the labels into the validation split, but the splitting methodology also tried to get a proportional number of each class into the validation split. The validation split is held constant throughout each fold of LOOCV and is used for early stopping and learning diagnostics. The ADAM optimizer (Ruder, 2017) uses validation loss to guide updates after each batch. Additionally, validation loss is used to enable early stopping of the model through a callback, which halts training if no improvement is seen after 15 epochs, restoring the best model weights.

Backpropagation is a two-phase optimization procedure used to train neural networks. For each time series input to the network, a forward pass is performed first, where the observations are transformed as they propagate through the network to produce outputs, and the loss is computed. Then, during the backward pass, the algorithm computes the gradients of the loss with respect to each network parameter using the chain rule of calculus, a process formally known as backpropagation through time (BPTT) in recurrent networks like LSTMs (Werbos, 1990). The update function for gradient decent in BPTT is:

$$\theta \leftarrow \theta \, \eta \frac{\partial \mathcal{L}}{\partial \theta} \tag{5}$$

Where θ consists of the weights and biases of the LSTM, η is the learning rate, and $\mathcal{L}$ is the loss function sparse categorical crossentropy (Jaén-Vargas et al., 2021) which computes loss for multiclass classification problems. This function guides the training process of minimizing loss. Loss is a differentiable measure of fit where low loss indicates high quality fit.

To ensure the model did not propagate gradients from missing labels, a masking strategy was applied both through the masking layer and by constructing a sample weight matrix that zeroed out contributions from unannotated time steps.

To mitigate overfitting, the model incorporated a fixed dropout rate of 0.1 on the kernel and recurrent weights. Dropout means that a random proportion of the weights in the network are deleted. The final model architecture included a masked input layer, an LSTM layer, a dense ReLU-activated hidden layer, and a softmax output layer for classification. In the OPO model, the LSTM layer returned the values for every index in the input sequence array, where as in the OPS model, the LSTM layer returned solely the final observations' values. The output layers were adjusted to account for these two possible shapes of input.

In each fold of LOOCV, a new network with new training, validation, and testing data is initialized based on the hyperparameters set in BO. Once all folds are evaluated, a final objective value is matched to the set of hyperparameters and BO finds a new set of hyperparameters to repeat the process with. Figure 14 shows the objective value after 200 calls of BO for OPS and OPO io types. For both architecture types, there is a steep descent in the first 20 iterations or so, then the improvements plateau and smaller improvements if any are made through the rest of the BO process. After 200 rounds of cross validation holding out 2-3 cows per fold, the OPS model has

attained a best micro F1 score of 0.8962, and after 200 rounds the OPO model has attained a best micro F1 score of 0.8961 with the parameters from Table 11.

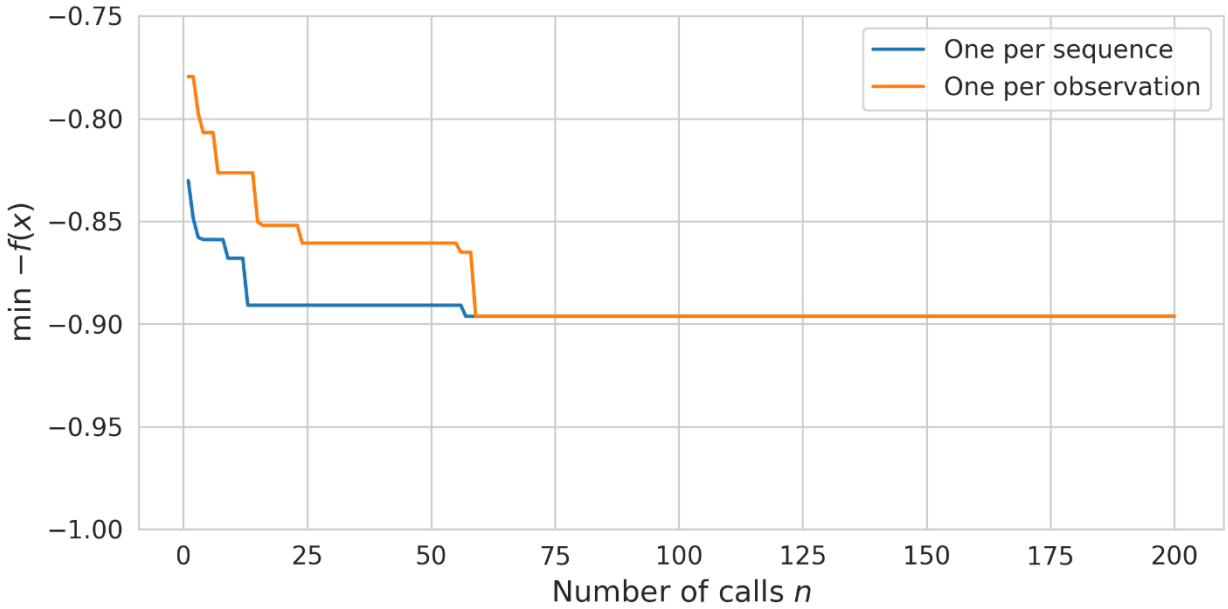

Figure 14. Convergence plots for BO on the two model input/output structures.

Table 11 shows the parameters used for the OPS and OPO structures. The LSTM classifiers were first evaluated with optimal parameters holding out a single cow per LOOCV fold for continuity with the HMM section. Then for the final dataset, the entire dataset was used for training and validation to make a model which predicts the full 67 days of data.

Table 11. Optimal parameters after running BO on OPO and OPS models.

	Optimal Value	
Parameter	**OPS**	**OPO**
Sequence Length	60	122
Batch size	23	32
Initial learning rate	0.023111	0.011718
Decay steps	3000	1667
Decay rate	0.333982	0.138811
Clipnorm	1.433801	1.5
LSTM Size	8	16
Dense Size	16	64

The OPO model had 26 labeled response sequences. Four of the 22 cows were observed on two days, so those cows each have two sequences containing labels. There were an average of 18.5 labels per sequence of length 122. The fewest number of labels in a sequence was five, and the most labels in a sequence was 25. Five labels corresponds to 25 minutes of observation, and 25 labels corresponds to 125 minutes of observation. As for the OPS data, every label has 60 observations (five hours) of history. Again, balanced by class, 30% of the labels in the OPS and OPO models were used as validation data for the full models.

Results

With the parameters of the models selected and the methodology laid out for determining results, the LSTM models were evaluated. These models are first evaluated via LOOCV to get class-wise metrics. Then, the model is rebuilt with all the labeled data as training and validation data. These full information models are used to create predictions over the 67 days of the study.

Cross Validation: In both the OPO and OPS case the LSTM models converged quickly. Figure 15 shows the OPO training history and Figure 16 shows the OPS training history for LOOCV folds. The models have high performance in the first epochs. This indicates that the input features are highly related to the outputs, and the models find weights and biases which predict the training data quickly. There were less than 40 epochs in both training cycles. With a patience of 15, the last 15 epochs of each fold did not exceed the best performance seen. So, for both models, in less than 25 epochs the models were well fit and performant on both training and validation data.

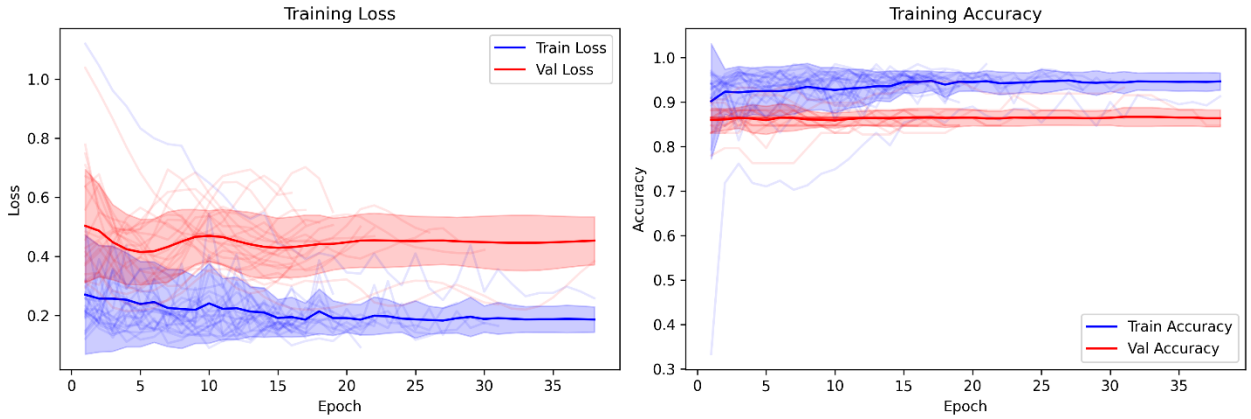

Figure 15. Training progress of the OPO LSTM through the epochs.

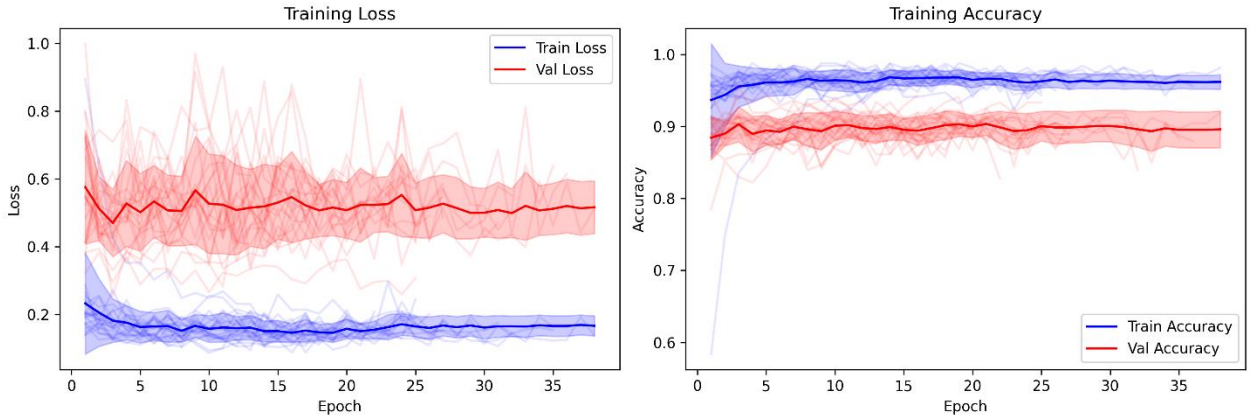

Figure 16. Training progress of the OPS LSTM through the epochs.

The confusion matrix and performance metrics of the OPO model are provided in Table 12 and

Table 13. We see that the grazing class was correctly identified in 197 out of 203 true grazing instances. Conversely, the model predicted 18 non-grazing records as grazing. This means the model was very sensitive (97.04%) but was less specific (93.53%) for the grazing class. For the resting class, 217 of the 232 actual resting cases were correctly identified, but the model incorrectly classified 6 non-resting cases as resting. The model was very specific (97.59%) but not as sensitive (93.53%) to resting cases. As for the traveling class, of the 46 true traveling labels, 41 were correctly identified. Two non traveling cases were labeled as traveling. This means the model

was extremely specific (99.54%) but less sensitive (89.13%) to the traveling case. The OPO model achieved an overall accuracy of 94.19%.

Table 12. Confusion matrix of LOOCV results from the OPO LSTM.

Predicted	**Reference**		
	Grazing	**Resting**	**Traveling**
Grazing	197	14	4
Resting	5	217	1
Traveling	1	1	41

Table 13. Classification performance of the OPO network using LOOCV.

State	**Actual**	**Predicted**	**Accuracy**	**Precision**	**Specificity**	**Sensitivity**	**F1**
Grazing	203	215	0.950	0.916	0.935	0.970	0.943
Resting	232	223	0.956	0.973	0.976	0.935	0.954
Traveling	46	43	0.985	0.954	0.995	0.891	0.921

The confusion matrix (Table 14) and class wise performance metrics (
Table 15) for the predictions of the OPS LSTM show that the model was very performant. Of the 203 actual grazing cases, 198 were correctly identified. The model incorrectly identified 14 non-grazing cases as grazing. The model was very sensitive (97.50%) and very specific (95.00%) when identifying grazing cases. As for the resting class, the model correctly identified 220 of the 232 resting cases. The model misclassified 6 non-resting cases as resting. The model was very specific (97.60%) with its applications of the resting class, and slightly less sensitive (94.80%) to resting cases. The traveling class was correctly applied 42 out of 46 times. Only one non-traveling case was incorrectly identified as traveling. This indicates the model was extremely specific

(99.80%) with its applications of the traveling case, but was less sensitive (91.30%) to traveling cases. Overall, the model achieved an accuracy of 94.59%.

Table 14. Confusion matrix of LOOCV results from the OPS LSTM.

Predicted	**Reference**		
	Grazing	**Resting**	**Traveling**
Grazing	198	11	3
Resting	5	220	1
Traveling	0	1	42

Table 15. Classification performance of the OPS network using LOOCV.

State	**Actual**	**Predicted**	**Accuracy**	**Precision**	**Specificity**	**Sensitivity**	**F1**
Grazing	203	212	0.960	0.934	0.950	0.975	0.954
Resting	232	226	0.963	0.973	0.976	0.948	0.961
Traveling	46	43	0.990	0.977	0.998	0.913	0.944

Cattle Behavior: The final LSTM models applied to the data collected throughout the study period produced insightful results that further strengthened the confidence in both the model's effectiveness and the overall modeling approach. The following sections present these findings.

Daily Grazing Behavior: Looking at the daily grazing activity, The OPO model predicts that cows spend an average of 36.2% of each day grazing, which is approximately 8.7 hours per day grazing. The OPS model predicts that cows spend an average of 37.4% of the day grazing, which is approximately 9 hours per day grazing. Figure 17 shows the heatmap of the grazing activity by cow and day from the OPO model, and Figure 18 shows the heatmap of the grazing activity by cow and day from the OPS model. The bar chart above the heatmap shows daily trends and the day's average deviation from the grand mean. The bar chart to the right of the heat map

shows animal trends and average deviations of each animal from the grand mean. The vertical strips of blue show days where the model predicts that all cows reduced their grazing activity. The horizontal strips in the heatmap show cows which have distinct grazing tendencies. The daily deviation aggregations have very similar patterns among almost all days. When the OPO model predicts a day as an above or below average grazing day, the OPS model does too. As for the cow aggregations, there is more variance between the two models. Generally speaking, the OPS model predicts higher variability between the cows, while the OPO model has a more homogenous grazing aggregation between cows.

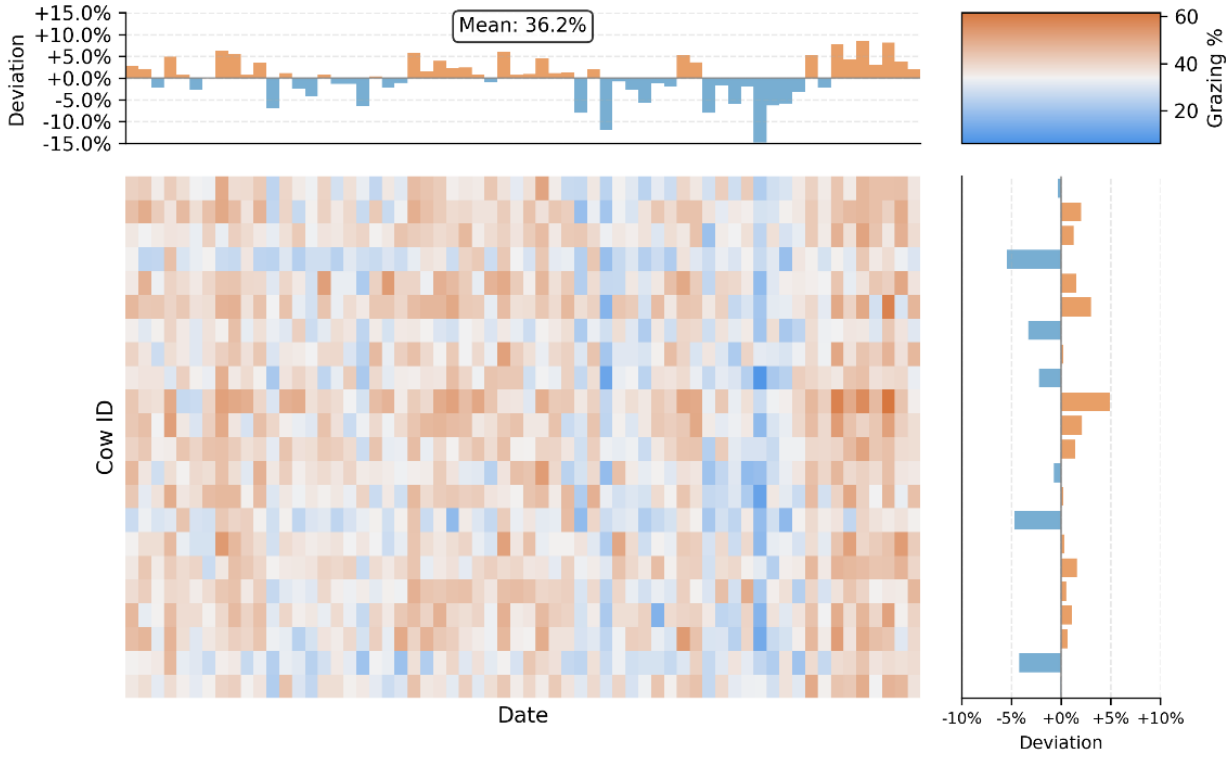

Figure 17. OPO heatmap of predicted grazing activity over the study.

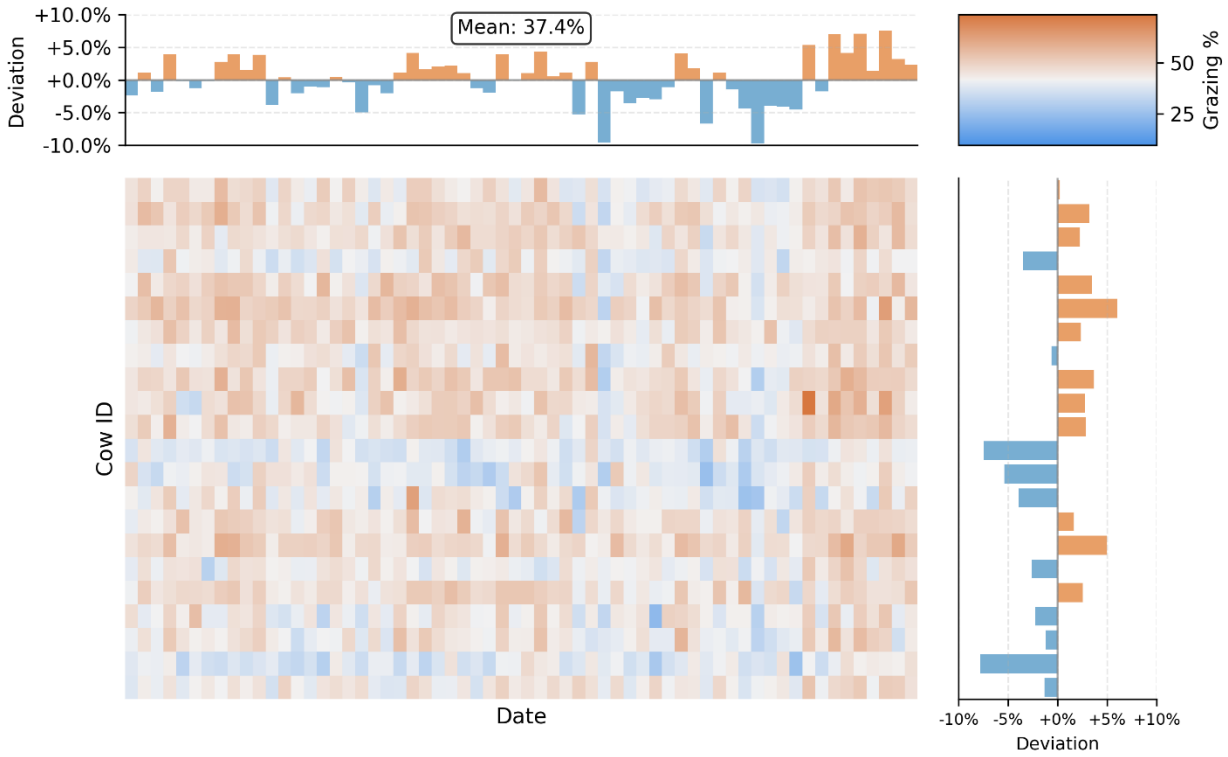

Figure 18. OPS heatmap of predicted grazing activity over the study.

These heatmaps show that our models predict variable grazing patterns across both time and animals, but do not show the daily time budget for these animals. To answer questions about what activity is happening when, radar plots were created (Figure 19,

Figure 20). These radar plots show the activity for one animal at a time. Two animals have been selected randomly to compare. Both models show strong diurnal patterns with grazing bouts in the morning and evening separated by a period of resting and traveling. The OPO model seems to predict more traveling behavior than the OPS model. The models both predict that cows tend to spend the nighttime resting, but both models predict similar patches of grazing in the night.

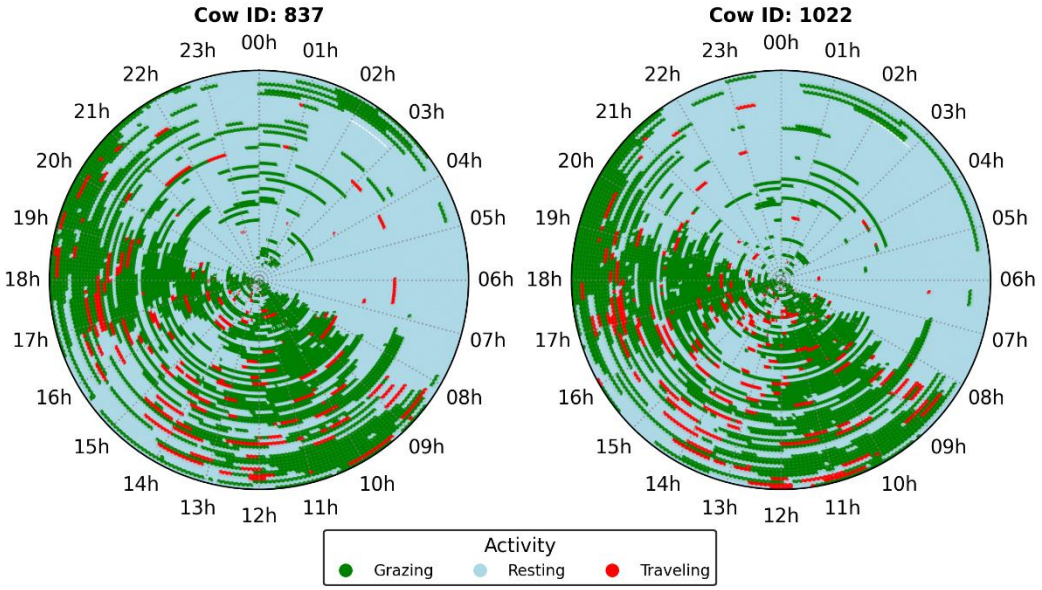

Figure 19. OPO radar plots of two cows activity over the study.

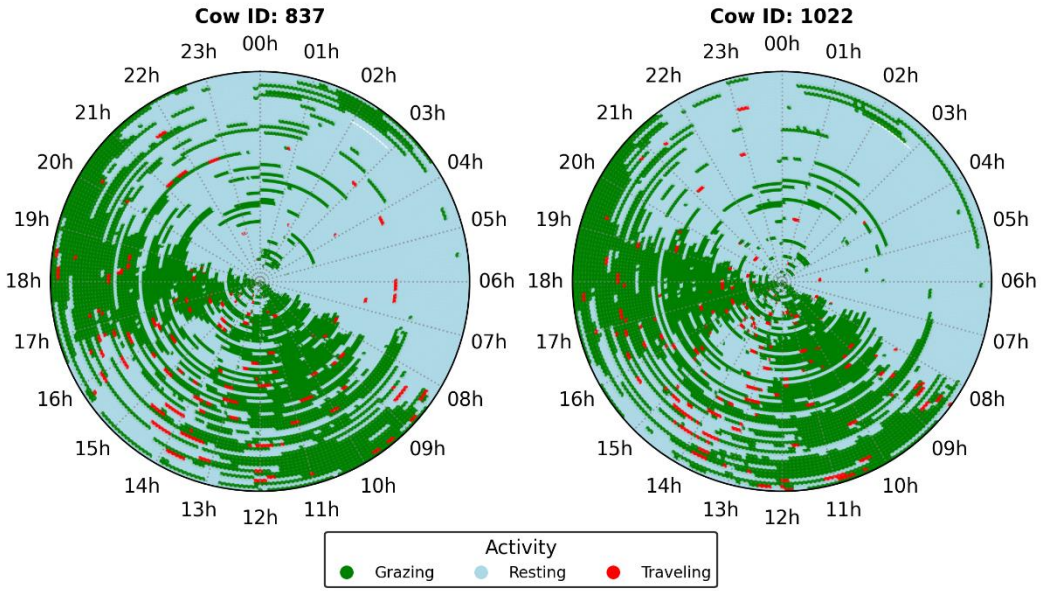

Figure 20. OPS radar plots of two cows activity over the study.

Another way we can look at the activity of the animals is by seeing how the grazing predictions interact with temperature. Figure 21 shows a gaussian smoothed surface where the x

and y axes are time of day and date, while the z axis represents the proportion of animals that were grazing for the OPO model, and Figure 22 shows the same for the OPS model. The color indicates the average temperature measured. Grazing activity seems to depress on the peaks slightly with colder temperatures, but in the "valley" between the morning and evening peak, grazing activity is more common with decreased temperatures, and less common with increased temperatures. Additionally, there are three grazing peaks in the early morning which correspond to full moons. The predictions from the OPO model show more variability in the afternoon peak and a deeper midday trench than the predictions from the OPS model.

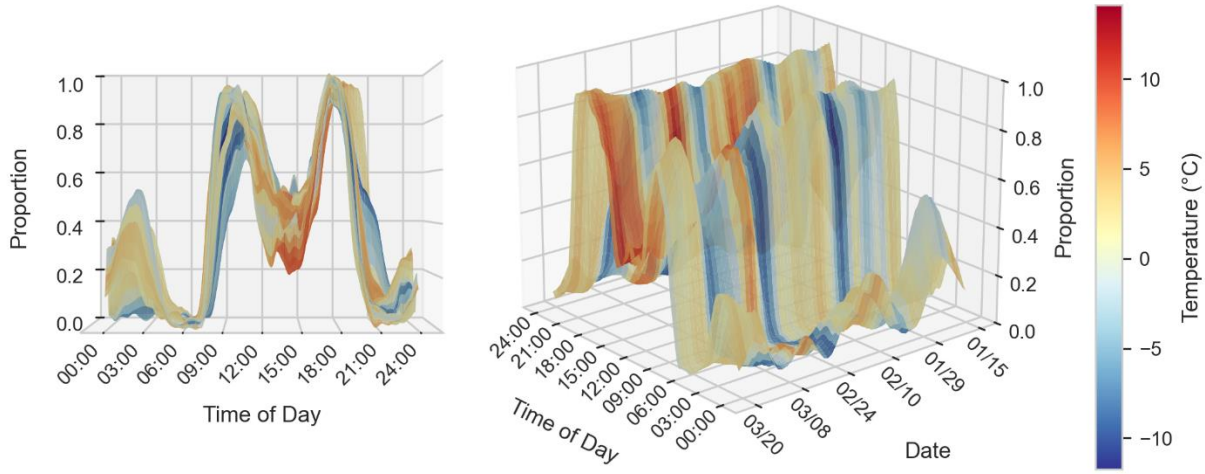

Figure 21. Smoothed surface plot of the predicted grazing activity from the OPO LSTM.

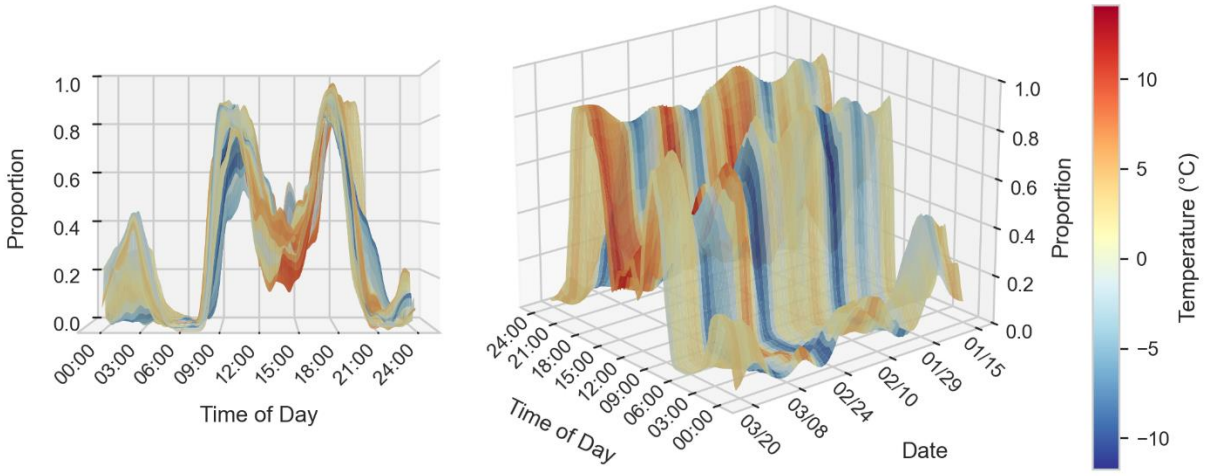

Figure 22. Smoothed surface plot of the predicted grazing activity from the OPS LSTM.

Nighttime grazing behavior: To look closer at the nighttime grazing activity and see if full moons had a measurable interaction with the amount of grazing activity our model predicts, a paired t-test was performed. Figure 23 shows nighttime behavior for the OPO model predictions, and Figure 24 shows the nighttime behavior for the OPS model predictions. Nighttime was set to be 1.5 hours after sunset until 1.5 hours before sunrise. A night has a full moon if the illumination of the moon is at least 95%. If the illumination is less than 95%, it is classified as a non-full moon. In these two figures, we observe that both models predict that most animals seem to have measurable increases in grazing activity during the night when there was a full moon versus when there was not a full moon. The models predict that more than half of the cows shows significantly different amounts of nighttime grazing when there was a full moon.

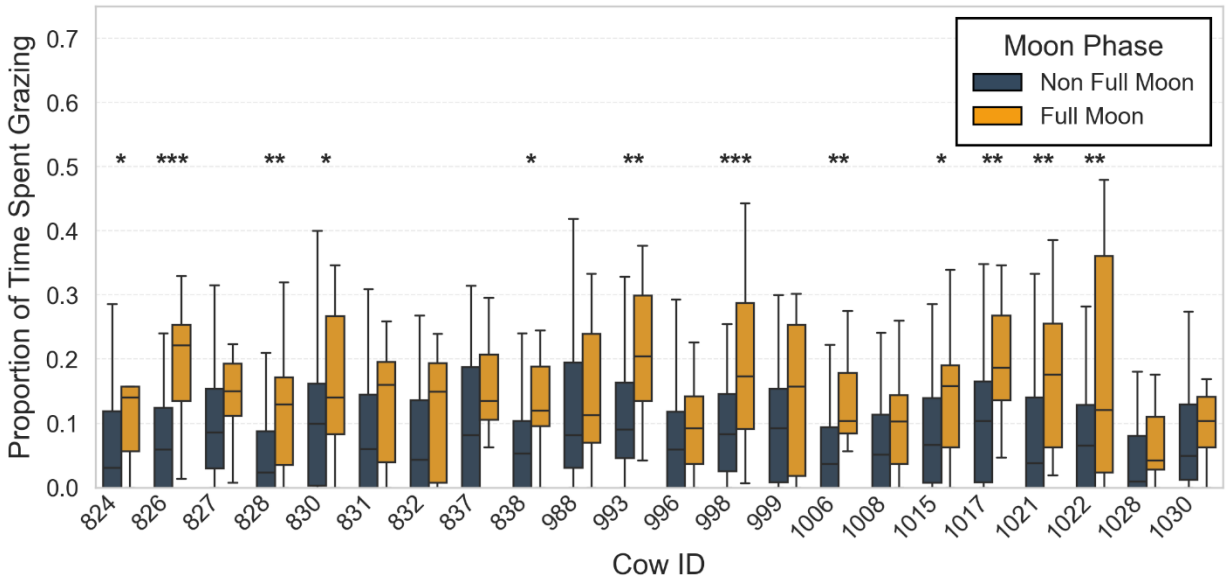

Figure 23. Nighttime grazing for cows as predicted by the OPO LSTM split by the presence of a full moon.

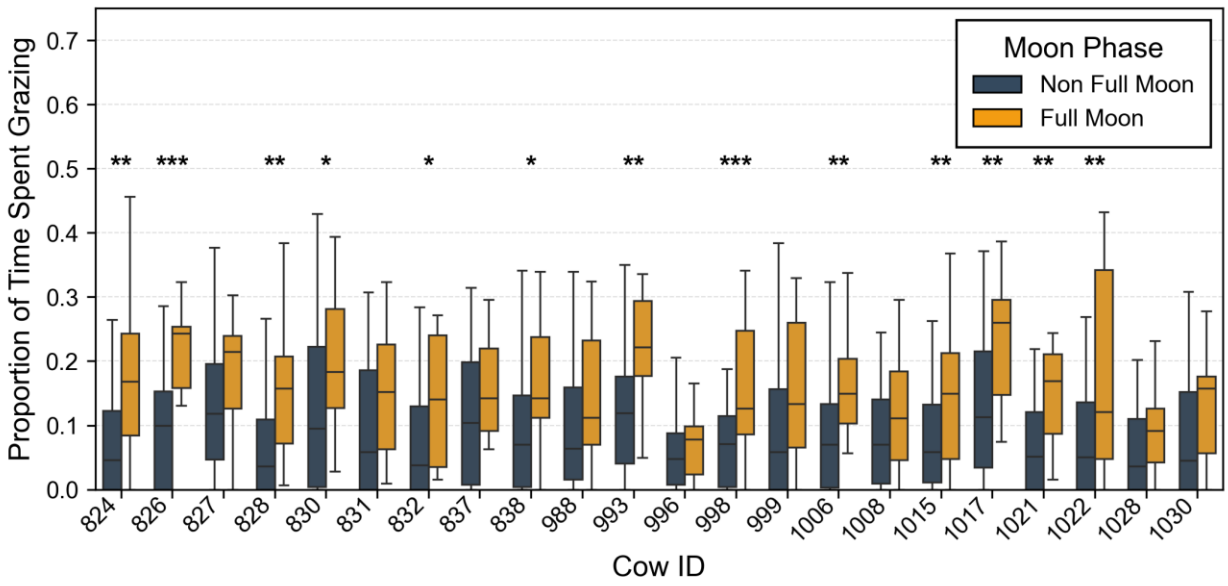

Figure 24. Nighttime grazing for cows as predicted by the OPS LSTM split by the presence of a full moon.

Effect of Time of Day and Temperature on Cattle Behavior: To quantify the effect that time and temperature have on the activity profile of the animals, GLMMs were made which estimate the probability of an animal being in a state given the time and temperature of the

observation. For more information on what GLMMs are and how they were implemented, refer to the corresponding HMM section.

The parameters and statistical significances are provided in

Table 16 for the OPO model and
Table 17 for the OPS model. The symbol `***` indicates a significance <0.001, `**` indicates a significance <0.01, `*` indicates a significance of < 0.05, and `.` indicates a significance of <0.1. Parameters which did not have an alpha <0.1 were not used for computing the curves in Figure 25 or Figure 26.

Table 16. Fitted GLMM parameter estimates and the significance of the measured effect for the OPO LSTM.

Parameter	**Grazing**		**Resting**		**Traveling**	
Intercept	-0.74	***	0.429	***	-3.248	***
Time	0.7667	***	-0.7733	***	0.2085	***
Time^2	-3.618	***	3.766	***	-2.153	***
Time^3	3.589	***	-3.33	***	-0.051	(NS)
Time^4	-3.76	***	3.643	***	-0.394	***
Temperature	-0.0242	**	0.0863	***	-0.2235	***
Temperature x Time	-0.276	***	0.337	***	-0.118	***
Temperature x Time^2	0.773	***	-0.839	***	-0.028	(NS)
Temperature x Time^4	-0.765	***	0.711	***	0.029	(NS)

Table 17. Fitted GLMM parameter estimates and the significance of the measured effect for the OPS LSTM.

Parameter	**Grazing**		**Resting**		**Traveling**	
Intercept	-0.939	***	0.481	***	-3.097	***
Time	0.8669	***	-0.7695	***	-0.0288	**
Time^2	-4.188	***	3.873	***	-2.097	***

Parameter	Grazing		Resting		Traveling	
Time^3	4.01	***	-3.368	***	-0.674	***

Table 17. Continued.

Parameter	Grazing		Resting		Traveling	
Time^4	-4.207	***	3.847	***	-0.932	***
Temperature	-0.0121	(NS)	0.0654	***	-0.0718	***
Temperature x Time	-0.316	***	0.269	***	0.027	(NS)
Temperature x Time^2	0.89	***	-0.853	***	0.227	***
Temperature x Time^4	-0.698	***	0.676	***	0.268	***

The behavior profile for the cows with respect to temperature and time is very similar between the OPO and OPS models for grazing and resting classes. The traveling class has a more bimodal pattern with cooler temperatures in the OPO model, but that trend does not present itself in the OPS model. In general, the probability of the traveling class is lower in the OPS model than the OPO model.

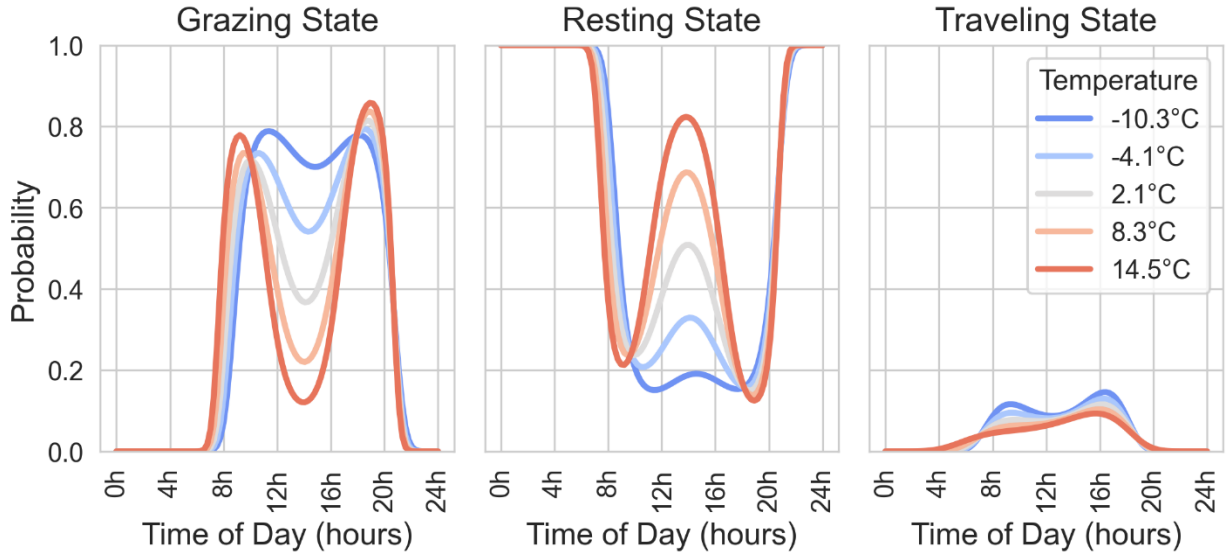

Figure 25. Probabilities of each activity as predicted by the GLMM for the OPO LSTM.

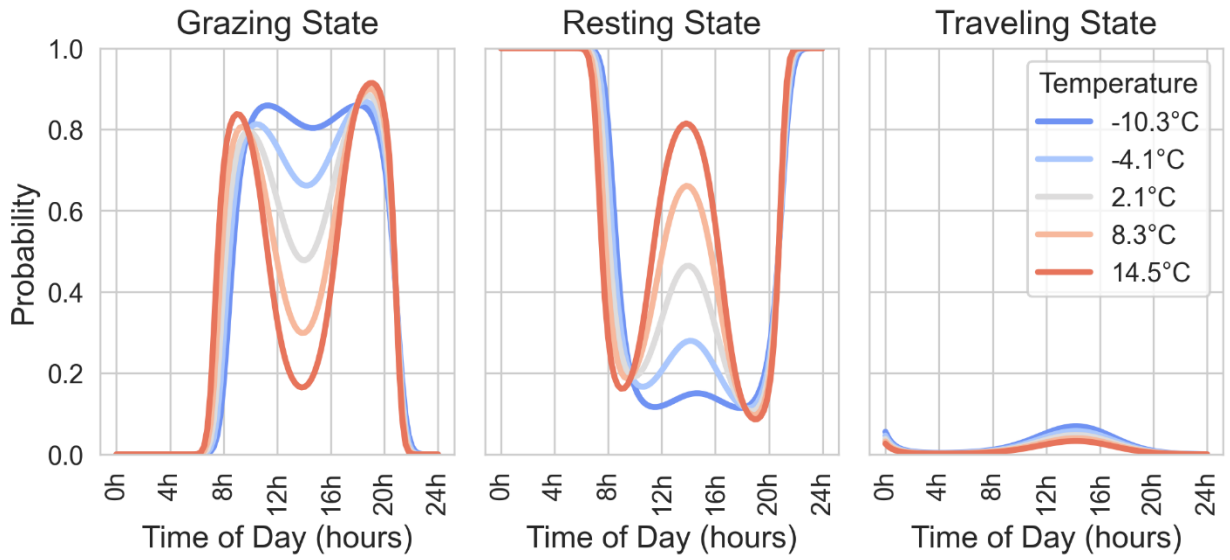

Figure 26. Probabilities of each activity as predicted by the GLMM for the OPS LSTM.

This chapter outlined the structure, validation results, and predictions derived from LSTM networks. The LSTM models showed strong performance in both cross validation and unlabeled dataset performance, but the lack of model accessibility is a limitation of LSTM models. In the next chapter, we integrate insights from both HMMs and LSTMs, providing a discussion comparing their relative strengths, limitations, and implications for precision livestock behavior classification.

DISCUSSION

Our results establish both HMMs and LSTMs as effective tools for classifying cattle behavior using GPS and accelerometer data. The classification models achieve competitive overall accuracies of 92.72% for the HMM, 94.19% for the OPO LSTM, and 94.59% for the OPS LSTM.

Outlier animals

Three cows stood out for their notably lower LOOCV performance, shown in Table 18. Cow 1006 was consistently misclassified across all models. Cow 827 was poorly predicted by both LSTMs, but the HMM performed better. Conversely, cow 1021 was poorly handled by the HMM, while both LSTMs achieved perfect accuracy.

Table 18. Three cows with notably lower performance in the LOOCV folds.

Cow ID	Number of labels	HMM	OPO	OPS
827	9	88.88%	55.56%	66.67%
1006	18	61.11%	83.33%	66.67%
1021	31	64.52%	100%	100%

Cow 827 had only 9 labeled datapoints, so its accuracy is sensitive to even one misclassification. But this alone doesn't explain why both LSTMs struggled more than the HMM. Looking at Figure 5, cow 827 shows considerable spread in the VarSVM feature for the Resting class. The LSTMs may have overfit or failed to generalize from this noisy pattern, while the HMM's parametric assumptions could have smoothed it out.

Cow 1006 is more problematic. It shows wide variation in both MeanSVM and VarSVM for the Resting class, and relatively weak separation between Resting and Grazing. This would undermine both emission modeling (for HMMs) and learned feature discriminability (for LSTMs), making it hard to assign class boundaries. It's plausible that the behavioral labeling for this cow

was noisier, or that it simply exhibited ambiguous movement patterns—though hardware variance can't be ruled out either.

Cow 1021 shows unusually low MeanSVM and VarSVM for Grazing, blurring the distinction from Resting in terms of movement intensity. This likely hurt the HMM, which relies on distinct statistical signatures for each state. However, the LSTMs, which can incorporate temporal dynamics more flexibly, managed perfect classification—suggesting that how the features evolve may have compensated for their overlap in magnitude.

We found no demographic or physiological factors (e.g., age or weight) that distinguished these cows. Instead, the deviations in their feature distributions—clearly visible in Figure 5—point to underlying differences in either behavior expression, sensor reliability, or labeling consistency. While we cannot definitively attribute the errors to any one cause, the data strongly suggest that model assumptions (statistical vs. sequential learning) and class label ambiguity both played a role.

Comparison of Training and Cross Validation

Once predictions are made for each of the test cows, they are concatenated and compared with the ground truth data to assess model performance. All three models achieve low-mid 90's in terms of overall accuracy. The HMM had the worst class-wise performance, and the OPS LSTM had the best class-wise performance. Traveling was consistently the worst performing class with F1-scores of 0.864, 0.921, and 0.944 for the HMM, OPO, and OPS models respectively. All of the models were more specific than sensitive when predicting Traveling. That is to say—false positives were more common than false negatives. For the Grazing class, the F1-Scores were 0.925, 0.943, and 0.954 for the HMM, OPO and OPS models. The models were all more sensitive than specific when predicting Grazing. The best performing class was Resting with F1-scores of 0.941, 0.954, and 0.961. Each of the models was more specific than sensitive when predicting Resting.

For the HMM, a Newton-type algorithm was used to estimate initial state probabilities, transition probabilities, and emission distribution parameters which maximized the likelihood of observed sequences. The decoding step involved applying the Viterbi algorithm to infer the most likely sequence of hidden behavioral states for the held-out cow, with the process repeated for all 22 cows.

The LSTMs used BPTT to update the weights and biases of the models with the goal of maximizing the micro F1 score the model attains on the validation data. Once the models' parameters are found, the training data is evaluated. Likelihoods are computed for each behavior at each timestep and the most likely sequence of behaviors is returned for the held-out cow.

Comparison of Unlabeled Dataset Predictions

The application of the final models to the complete dataset of 22 cows over a nine-week period further reinforced confidence in both the modelling approaches and the modelling effectiveness. All three models estimated an average daily grazing duration of 8-9 hours, which aligns closely with expected cattle grazing behavior during the dormant season (Adams et al., 1986; DelCurto et al., 1990). The heatmaps in Figure 7, Figure 17, and Figure 18 show that each of the three models predicted similar trends across the days, but dissimilar patterns across cows. The OPO model predicted the least variability between cows, then the HMM, and then the OPS model predicted greatest variability between cows.

The radar plots from the HMM, OPO and OPS models in Figure 8, Figure 19, and

Figure 20 show a consensus that grazing behavior predominantly occurs before sunset and after sunrise, while resting behavior is more frequent during midday and nighttime. These patterns align with well-established expectations of cattle behavior (Brennan et al., 2021; Iqbal et al., 2023;

Kilgour and Dalton, 2019; Linnane et al., 2001) further validating each model's ability to capture real-world behavioral dynamics.

Additionally, our analysis revealed that moon phases—and the associated levels of moonlight—have a statistically significant effect on nighttime grazing behavior. Grazing time almost doubles during nights with maximum moonlight compared to other nights. Each of the three models show an increase in grazing when there was a full moon, though the magnitudes varied. The HMM predictions had the greatest difference between full moon and non-full moon grazing by cows. This finding that full moons increase the amount of nighttime grazing aligns with (Gibb, 2007), who concluded a third grazing bout was observed in dairy cows when there was a full moon. Other large ruminants are also known to express increased foraging behaviors during full moons such as White Tailed Deer (Lashley et al., 2014). There is literature which disputes this claim. In (Dwyer, 1960), it is reported that moon phase has no significant effect on cattle grazing behavior.

Previous research shows a complex relationship between ambient temperature and cattle behavior. (Adams et al., 1986) reported a decline in mean daily grazing time with decreasing minimum daily temperature (MDT) during the winter, along with delayed start of morning grazing at higher MDTs. (Dunn et al., 1988) observed no significant association between daily temperature and total grazing time during the mid-winter period, but did find that temperature influenced within-day grazing patterns. (Prescott et al., 1994) found a modest positive relationship between ambient temperature and daily grazing time, with cows grazing more in fall trials. More recently, (Keren and Olson, 2006) concluded that cattle modulate behavior in response to complex environmental stimuli, including temperature, to reduce energy demands during winter. (Parsons et al., 2003) compared grazing behavior in early and late summer, noting changes in the diurnal

pattern of grazing activity, potentially driven by ambient temperature and photoperiod, despite consistent total grazing time. (Nakajima et al., 2019) showed that cattle exposed to cold environments engaged in more grazing and travel activity than those kept in confinement, highlighting behavioral adjustments to thermal stress. (Sprinkle et al., 2021) reported that in years characterized by more severe winter conditions, cows started grazing earlier and exhibited more intense evening grazing bouts.

To investigate the interaction between temperature and diurnal behavior, we applied GLMMs to behavioral states classified using our models. The GLMMs for each set of predictions shows the diurnal pattern with a less intense resting valley on colder days, and sharper peaks on warmer days. Our findings contribute to the growing body of literature on environmental influences on cattle behavior and provide additional evidence that ambient temperature plays a significant role in shaping behavioral patterns. The strong overlap between our results and prior studies reinforces both the validity of our models and the broader utility of both HMMs and LSTMs for behavioral inference.

CONCLUSION

Overall, our work highlights the potential of serially dependent classifiers for livestock behavior monitoring, demonstrating both accuracy and interpretability in classifying complex movement and activity patterns. Beyond achieving high overall accuracy, HMMs offer additional benefits as a statistical modeling approach. Their probabilistic structure makes them robust to noise and missing data, common in sensor-based studies. They model state transitions and emissions explicitly, enabling insights into behavioral dynamics by simply studying the parameters of the distributions. Finally, their emission probabilities which can be visualized with graphics such as Figure 6 are interoperable and build trust in the model outputs and facilitate reasoning about individual differences, such as the misclassification patterns observed in two animals.

While both LSTMs and HMMs achieved similarly high accuracy on cross-validation and full-dataset predictions, they differ fundamentally in complexity. HMMs, especially as implemented in momentuHMM, are transparent and interpretable, but that clarity hinges on well-characterized data distributions. When features are difficult to model with statistical distributions such as multimodal features or frequency-domain features, the elegance of HMMs breaks down. Fitting custom distributions, approximating marginal likelihoods, or engineering workarounds not only complicates the implementation but undermines the maintainability and transparency that make HMMs appealing in the first place. By contrast LSTMs make no promises of interpretability. LSTMs manage to perform well without interoperable emission distributions or transition matrices. Their strength lies in absorbing raw or minimally processed inputs, sidestepping the need for extensive feature engineering. This makes LSTMs easier to implement and scale as the feature space grows.

This study demonstrates the effectiveness of serial classifiers for free-range cattle behavior classification, a modeling approach that remains underutilized despite its suitability for sequential animal behavior modeling. To study the effectiveness and classification performance of HMMs and LSTMs, we used sensor data collected from 22 free-range Angus-based cows during a nine-week winter period at the Red Bluff Research Ranch in Norris, MT, USA. The model development process involved two key steps: learning and decoding. The learning phase addressed parameter estimation using Leave-One-Out Cross-Validation (LOOCV), where data from one cow was held out while the remaining 21 cows' data were used to initialize model parameters.

Following cross-validation, final HMM and LSTM models were trained using the entire labeled dataset from all 22 cows to classify behaviors across the full nine-week experimental period. We analyzed behavior patterns across the entire dataset, investigating cattle responses to diurnal cycles, moon phases, and temperature. These results were further contextualized through comparisons with existing findings in the literature.

While both HMMs and LSTMs proved highly effective for the given task, HMMs are likely the better tool in this specific setting. The simplicity, transparency, and robustness of HMMs is more beneficial than marginal improvements in accuracy while having comparable results in unlabeled dataset performance. If the context of the problem was larger, such as using more types of sensors or integrating additional data sources into the model, LSTMs would be the better tool for the task as they are able to directly handle these more complicated features. With regards to the OPO versus OPS LSTM, the OPS LSTMs are the preferable architecture for this dataset. The low sampling frequency of 1/60Hz for accelerometer data and 1/300Hz for GPS data over 67 days does not create large enough datasets to really cause strain on a modern computer. Some studies work with much higher frequency data and create orders of magnitude more data which would

cause more computational pressure when training models. Because of the lack of computational pressure for this dataset, there is not a need to compromise on the amount of historical data used to predict each label by using the OPO structure. The benefits of the OPS structure, using fixed number of historical observations for every prediction, can be fully realized on this dataset despite the increased computational requirements. OPS models can also be optimized, trained, and applied in a reasonable amount of compute time.

Limitations

While both the HMMs and LSTMs were performant, there are limitations to the study. All the labeled data was collected from a single pasture around 11:00am on two days in late March. Multiple observers labeled activity, and there is no information about the tendencies and biases of each observer. Figuring out ways to sample activity from a more diverse range of times would provide more holistic training data, potentially improving the model's generalizability and ability to represent the activity profiles of the animals over the year. Additionally, while more than three labels were recorded, there were too few labels for niche behaviors such as drinking, scratching, or fighting to be able to train ML models to predict these classes. By only predicting Grazing, Resting, and Traveling, the potential applications of this labeled dataset are limited. Also, a limitation of the momentuHMM package is that a single distribution had to be used for an entire feature. As an example, grazing and resting, which are the dominant classes in this dataset, were get fit by a lognormal distribution. The traveling class only constitutes about 10% of the labels, but it likely would have been better fit with a gamma distribution (Figure 6).

Future Work

A very intriguing application of LSTMs is featureless modeling. Unlike traditional models that rely on handcrafted features such as step size or signal variance, LSTMs can ingest raw GPS coordinates and accelerometer readings directly and still achieve high performance (Peng et al., 2019). Exploring end-to-end architectures that bypass the feature engineering step entirely could streamline the data pipeline and reduce preprocessing overhead, making large-scale deployment more feasible.

Additionally, integrating real-time data processing with embedded anomaly detection could enable early identification of abnormal behavior such as illness, injury, or estrus. Developing lightweight models capable of running on edge devices could make these insights available to ranchers on demand, improving both animal welfare and operational efficiency (Arablouei et al., 2021).

Another promising direction is unsupervised learning. Hidden Markov Models in particular are well-suited for unsupervised behavior segmentation and have been shown to detect latent behavioral patterns without labeled data. Reducing the reliance on human annotations not only lowers the cost of model deployment but also makes the approach more scalable to larger herds or longer time periods (Scriban et al., 2024).

Generalizability remains a critical concern (Kilgour et al., 2012). Current models are trained and validated on a single herd within a specific environmental and temporal context. Future studies should investigate how well these models transfer across different pastures, herds, breeds, and seasonal conditions (Arablouei et al., 2024). Without such testing, model predictions may be misleading when applied beyond the original context.

REFERENCES CITED

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems [Software]. Retrieved from https://www.tensorflow.org/.

Adams, D.C., Nelsen, T.C., Reynolds, W.L., Knapp, B.W., 1986. Winter Grazing Activity and Forage Intake of Range Cows in the Northern Great Plains. Journal of Animal Science 62, 1240-1246. https://doi.org/10.2527/jas1986.6251240x.

Aquilani, C., Confessore, A., Bozzi, R., Sirtori, F., Pugliese, C., 2022. Review: Precision Livestock Farming technologies in pasture-based livestock systems. Animal 16, 100429. https://doi.org/10.1016/j.animal.2021.100429.

Arablouei, R., Bishop-Hurley, G.J., Bagnall, N., Ingham, A., 2024. Cattle behavior recognition from accelerometer data: Leveraging in-situ cross-device model learning. Computers and Electronics in Agriculture 227, 109546. https://www.sciencedirect.com/science/article/pii/S0168169924009372.

Arablouei, R., Currie, L., Kusy, B., Ingham, A., Greenwood, P.L., Bishop-Hurley, G., 2021. In-situ classification of cattle behavior using accelerometry data. Computers and Electronics in Agriculture 183, 106045. https://www.sciencedirect.com/science/article/pii/S0168169921000636.

Awad, M., Khanna, R., 2015. Hidden Markov Model, Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Apress, Berkeley, CA, pp. 81-104 https://doi.org/10.1007/978-1-4302-5990-9_5.

Barwick, J., Lamb, D.W., Dobos, R., Welch, M., Trotter, M., 2018. Categorising sheep activity using a tri-axial accelerometer. Computers and Electronics in Agriculture 145, 289-297. https://doi.org/10.1016/j.compag.2018.01.007.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67, 1-48. https://doi.org/10.18637/jss.v067.i01.

Berckmans, D., 2017. General introduction to precision livestock farming. Animal Frontiers 7, 6-11. https://doi.org/10.2527/af.2017.0102.

Beumer, L.T., Pohle, J., Schmidt, N.M., Chimienti, M., Desforges, J.-P., Hansen, L.H., Langrock, R., Pedersen, S.H., Stelvig, M., van Beest, F.M., 2020. An application of upscaled optimal foraging theory using hidden Markov modelling: year-round

behavioural variation in a large arctic herbivore. Movement Ecology 8, 25. https://doi.org/10.1186/s40462-020-00213-x.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24, 127-135. https://doi.org/10.1016/j.tree.2008.10.008.

Brennan, J., Johnson, P., Olson, K., 2021. Classifying season long livestock grazing behavior with the use of a low-cost GPS and accelerometer. Computers and Electronics in Agriculture 181, 105957. https://doi.org/10.1016/j.compag.2020.105957.

Chang, A.Z., Fogarty, E.S., Moraes, L.E., García-Guerra, A., Swain, D.L., Trotter, M.G., 2022. Detection of rumination in cattle using an accelerometer ear-tag: A comparison of analytical methods and individual animal and generic models. Computers and Electronics in Agriculture 192, 106595. https://doi.org/10.1016/j.compag.2021.106595.

Chollet, F. (2015). Keras [Software]. Retrieved from https://keras.io.

Cusack, J.J., Dickman, A.J., Rowcliffe, J.M., Carbone, C., Macdonald, D.W., Coulson, T., 2015. Random versus Game Trail-Based Camera Trap Placement Strategy for Monitoring Terrestrial Mammal Communities. PLOS ONE 10, e0126373. https://doi.org/10.1371/journal.pone.0126373.

da Silva Santos, A., de Medeiros, V.W.C., Gonçalves, G.E., 2023. Monitoring and classification of cattle behavior: a survey. Smart Agricultural Technology 3, 100091. https://doi.org/10.1016/j.atech.2022.100091.

DelCurto, T., Cochran, R.C., Nagaraja, T.G., Corah, L.R., Beharka, A.A., Vanzant, E.S., 1990. Comparison of soybean meal/sorghum grain, alfalfa hay and dehydrated alfalfa pellets as supplemental protein sources for beef cattle consuming dormant tallgrass-prairie forage. Journal of Animal Science 68, 2901-2915. https://doi.org/10.2527/1990.6892901x.

Delignette-Muller, M.L., Dutang, C., 2015. fitdistrplus: An R package for fitting distributions. Journal of Statistical Software 64, 1–34. https://doi.org/10.18637/jss.v064.i04.

Diskin, M.G., Kenny, D.A., 2014. Optimising reproductive performance of beef cows and replacement heifers. animal 8, 27-39. https://doi.org/10.1017/S175173111400086X.

Dunn, R.W., Havstad, K.M., Ayers, E.L., 1988. Grazing behavior responses of rangeland beef cows to winter ambient temperatures and age. Applied Animal Behaviour Science 21, 201-207. https://doi.org/10.1016/0168-1591(88)90109-8.

Dwyer, D.D., 1960. Activities and grazing preferences of cows with calves in northern Osage County, Oklahoma. Texas A&M University.

El Moutaouakil, K., Noureddine, F., 2024. A comparative study on time series data-based artificial intelligence approaches for classifying cattle feeding behavior. Indonesian

Journal of Electrical Engineering and Computer Science 33, 324-332. http://dx.doi.org/10.11591/ijeecs.v33.i1.pp324-332.

Gao, G., Wang, C., Wang, J., Lv, Y., Li, Q., Ma, Y., Zhang, X., Li, Z., Chen, G., 2023. CNN-Bi-LSTM: A Complex Environment-Oriented Cattle Behavior Classification Network Based on the Fusion of CNN and Bi-LSTM. Sensors 23, 7714. https://doi.org/10.3390/s23187714.

Geroldinger, A., Lusa, L., Nold, M., Heinze, G., 2023. Leave-one-out cross-validation, penalization, and differential bias of some prediction model performance measures—a simulation study. Diagnostic and Prognostic Research 7, 9. https://doi.org/10.1186/s41512-023-00146-0.

Gibb, M.J., 2007. Grassland management with emphasis on grazing behaviour.

Giovanetti, V., Decandia, M., Molle, G., Acciaro, M., Mameli, M., Cabiddu, A., Cossu, R., Serra, M.G., Manca, C., Rassu, S.P.G., Dimauro, C., 2017. Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer. Livestock Science 196, 42-48. https://doi.org/10.1016/j.livsci.2016.12.011.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Graves, A., 2012. Long Short-Term Memory, Supervised Sequence Labelling with Recurrent Neural Networks. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 37-45 https://doi.org/10.1007/978-3-642-24797-2_4.

Grossi, G., Goglio, P., Vitali, A., Williams, A.G., 2018. Livestock and climate change: impact of livestock on climate and mitigation strategies. Animal Frontiers 9, 69-76. https://doi.org/10.1093/af/vfy034.

Guo, Y., Poulton, G., Corke, P., Bishop-Hurley, G.J., Wark, T., Swain, D.L., 2009. Using accelerometer, high sample rate GPS and magnetometer data to develop a cattle movement and behaviour model. Ecological Modelling 220, 2068-2075. https://doi.org/10.1016/j.ecolmodel.2009.04.047.

Hlimi, A., El Otmani, S., Elame, F., Chentouf, M., El Halimi, R., Chebli, Y., 2024. Application of Precision Technologies to Characterize Animal Behavior: A Review. Animals 14, 416. https://doi.org/10.3390/ani14030416.

Hosseininoorbin, S., Layeghy, S., Kusy, B., Jurdak, R., Bishop-Hurley, G.J., Greenwood, P.L., Portmann, M., 2021. Deep learning-based cattle behaviour classification using joint time-frequency data representation. Computers and Electronics in Agriculture 187, 106241. https://doi.org/10.1016/j.compag.2021.106241.

Iqbal, M.W., Draganova, I., Henry Morel, P.C., Todd Morris, S., 2023. Variations in the 24 h temporal patterns and time budgets of grazing, rumination, and idling behaviors in

grazing dairy cows in a New Zealand system. J Anim Sci 101. https://doi.org/10.1093/jas/skad038.

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.-A., 2019. Deep learning for time series classification: a review. Data Mining and Knowledge Discovery 33, 917-963. https://doi.org/10.1007/s10618-019-00619-1.

Ito, H., Takeda, K.-i., Tokgoz, K.K., Minati, L., Fukawa, M., Chao, L., Bartels, J., Rachi, I., A, S., 2022. Japanese Black Beef Cow Behavior Classification Dataset, v2.0.0 ed. Zenodo.

Jaén-Vargas, M., Reyes Leiva, K., Fernandes, F., Gonalves, S., Silva, M., Lopes, D., Serrano Olmedo, J., 2021. A Deep Learning Approach to Recognize Human Activity Using Inertial Sensors and Motion Capture Systems.

Jiang, B., Tang, W., Cui, L., Deng, X., 2023. Precision Livestock Farming Research: A Global Scientometric Review. Animals (Basel) 13. https://doi.org/10.3390/ani13132096.

Jolly, E., 2018. Pymer4: Connecting R and Python for Linear Mixed Modeling. Journal of Open Source Software 3, 862. https://doi.org/10.21105/joss.00862.

Karantonis, D.M., Narayanan, M.R., Mathie, M., Lovell, N.H., Celler, B.G., 2006. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Transactions on Information Technology in Biomedicine 10, 156-167. https://doi.org/10.1109/titb.2005.856864.

Keren, E.N., Olson, B.E., 2006. Thermal balance of cattle grazing winter range: Model development. Journal of thermal biology. 31, 371-377. https://doi.org/10.1016/j.jtherbio.2005.11.029.

Kilgour, R., 1984. Livestock Behaviour: A Practical Guide. 1st ed. CRC Press.

Kilgour, R., Dalton, C., 2019. Livestock Behaviour: a practical guide.

Kilgour, R.J., Uetake, K., Ishiwata, T., Melville, G.J., 2012. The behaviour of beef cattle at pasture. Applied Animal Behaviour Science 138, 12-17. https://www.sciencedirect.com/science/article/pii/S0168159111004217.

Koger, B., Deshpande, A., Kerby, J.T., Graving, J.M., Costelloe, B.R., Couzin, I.D., 2023. Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision. Journal of Animal Ecology 92, 1357-1371. https://doi.org/10.1111/1365-2656.13904.

Lashley, M.A., Chitwood, M.C., Biggerstaff, M.T., Morina, D.L., Moorman, C.E., DePerno, C.S., 2014. White-Tailed Deer Vigilance: The Influence of Social and Environmental Factors. PLOS ONE 9, e90652. https://doi.org/10.1371/journal.pone.0090652.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436-444. https://doi.org/10.1038/nature14539.

Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T.A., Watanabe, Y.Y., Murgatroyd, M., Papastamatiou, Y.P., 2017. Analysis of animal accelerometer data using hidden Markov models. Methods in Ecology and Evolution 8, 161-173. https://doi.org/10.1111/2041-210X.12657.

Lewis, J.S., Rachlow, J.L., Garton, E.O., Vierling, L.A., 2007. Effects of habitat on GPS collar performance: using data screening to reduce location error. Journal of Applied Ecology 44, 663-671. https://doi.org/10.1111/j.1365-2664.2007.01286.x.

Linnane, M.I., Brereton, A.J., Giller, P.S., 2001. Seasonal changes in circadian grazing patterns of Kerry cows (Bos Taurus) in semi-feral conditions in Killarney National Park, Co. Kerry, Ireland. Applied Animal Behaviour Science 71, 277-292. https://doi.org/10.1016/S0168-1591(00)00188-X.

Lipton, Z.C., Elkan, C., Naryanaswamy, B., 2014. Optimal Thresholding of Classifiers to Maximize F1 Measure. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 225-239.

Mahmud, M.S., Zahid, A., Das, A.K., Muzammil, M., Khan, M.U., 2021. A systematic literature review on deep learning applications for precision cattle farming. Computers and Electronics in Agriculture 187. https://doi.org/10.1016/j.compag.2021.106313.

Mancuso, D., Castagnolo, G., Porto, S.M.C., 2023. Cow Behavioural Activities in Extensive Farms: Challenges of Adopting Automatic Monitoring Systems. Sensors 23, 3828. https://doi.org/10.3390/s23083828.

McClintock, B.T., Michelot, T. (2018). momentuHMM: R package for generalized hidden Markov models of animal movement (6) [Software]. Retrieved from https://CRAN.R-project.org/package=momentuHMM.

Michelot, T., Langrock, R., Patterson, T.A., 2016. moveHMM: An R package for the statistical modelling of animal movement data using hidden Markov models.

Monteiro, A., Santos, S., Gonçalves, P., 2021. Precision Agriculture for Crop and Livestock Farming—Brief Review. Animals 11, 2345. https://doi.org/10.3390/ani11082345.

Mor, B., Garhwal, S., Kumar, A., 2021. A Systematic Review of Hidden Markov Models and Their Applications. Archives of Computational Methods in Engineering 28, 1429-1448. https://doi.org/10.1007/s11831-020-09422-4.

Nakajima, N., Doi, K., Tamiya, S., Yayota, M., 2019. Effects of direct exposure to cold weather under grazing in winter on the physiological, immunological, and behavioral conditions of Japanese Black beef cattle in central Japan. Animal Science Journal 90, 1033-1041. https://doi.org/10.1111/asj.13248.

Nocedal, J., Wright, S., 2006. Numerical Optimization. Springer New York.

Nyamuryekung'e, S., 2024. Transforming ranching: Precision livestock management in the Internet of Things era. Rangelands 46, 13-22. https://doi.org/10.1016/j.rala.2023.10.002.

OECD/FAO (2021), *OECD-FAO Agricultural Outlook 2021-2030*, OECD Publishing, Paris, *https://doi.org/10.1787/19428846-en*.

Parsons, C.T., Momont, P.A., Delcurto, T., McInnis, M., Porath, M.L., 2003. Cattle distribution patterns and vegetation use in mountain riparian areas. Rangeland Ecology & Management/Journal of Range Management Archives 56, 334-341. http://dx.doi.org/10.2307/4004036.

Parsons, I.L., Menendez, H.M., Brennan, J.R., Ehlert, K., 2023. PSX-15 Precision Beef Dry Matter Intake Estimation on Extensive Rangelands. J Anim Sci 101, 501-502. https://doi.org/10.1093/jas/skad281.591.

Peng, Y., Kondo, N., Fujiura, T., Suzuki, T., Wulandari, Yoshioka, H., Itoyama, E., 2019. Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units. Computers and Electronics in Agriculture 157, 247-253. https://doi.org/10.1016/j.compag.2018.12.023.

Prescott, M.L., Havstad, K.M., Olson-Rutz, K.M., Ayers, E.L., Petersen, M.K., 1994. Grazing behavior of free-ranging beef cows to initial and prolonged exposure to fluctuating thermal environments. Applied Animal Behaviour Science 39, 103-113. https://doi.org/10.1016/0168-1591(94)90130-9.

Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77, 257-286. https://doi.org/10.1109/5.18626.

Rautiainen, H., Alam, M., Blackwell, P.G., Skarin, A., 2022. Identification of reindeer fine-scale foraging behaviour using tri-axial accelerometer data. Movement Ecology 10, 40. https://doi.org/10.1186/s40462-022-00339-0.

Reimers, N., Gurevych, I., 2017. Optimal Hyperparameters for Deep LSTM-Networks for Sequence Labeling Tasks. https://doi.org/10.48550/arXiv.1707.06799.

Rhodes, B. (2011). PyEphem: Astronomical Ephemeris for Python [Software]. Retrieved from https://rhodesmill.org/pyephem/.

Riaboff, L., Couvreur, S., Madouasse, A., Roig-Pons, M., Aubin, S., Massabie, P., Chauvin, A., Bédère, N., Plantier, G., 2020. Use of Predicted Behavior from Accelerometer Data Combined with GPS Data to Explore the Relationship between Dairy Cow Behavior and Pasture Characteristics. Sensors 20, 4741. https://doi.org/10.3390/s20174741.

Riaboff, L., Shalloo, L., Smeaton, A.F., Couvreur, S., Madouasse, A., Keane, M.T., 2022. Predicting livestock behaviour using accelerometers: A systematic review of processing techniques for ruminant behaviour prediction from raw accelerometer data. Computers and Electronics in Agriculture 192, 106610. https://doi.org/10.1016/j.compag.2021.106610.

Roux, S.P.l., Wolhuter, R., Niesler, T., 2019. Energy-Aware Feature and Model Selection for Onboard Behavior Classification in Low-Power Animal Borne Sensor Applications. IEEE Sensors Journal 19, 2722-2734. http://dx.doi.org/10.1109/JSEN.2018.2886890.

Ruder, S., 2017. An overview of gradient descent optimization algorithms.

Scriban, A., Nabeneza, S., Cornelis, D., Delay, E., Vayssières, J., Cesaro, J.-D., Salgado, P., 2024. GPS-Based Hidden Markov Models to Document Pastoral Mobility in the Sahel. Sensors 24, 6964. https://doi.org/10.3390/s24216964.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms, Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. Curran Associates Inc., Lake Tahoe, Nevada, pp. 2951–2959.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Information Processing & Management 45, 427-437. https://doi.org/10.1016/j.ipm.2009.03.002.

Soruri, M., Zahiri, S., Sadri, J., 2013. A new approach of training Hidden Markov Model by PSO algorithm for gene Sequence Modeling.

Sprinkle, J.E., Sagers, J.K., Hall, J.B., Ellison, M.J., Yelich, J.V., Brennan, J.R., Taylor, J.B., Lamb, J.B., 2021. Protein Supplementation and Grazing Behavior for Cows on Differing Late-Season Rangeland Grazing Systems. Animals 11, 3219. https://doi.org/10.3390/ani11113219.

Sundman, E.R., Dewell, G.A., Dewell, R.D., Johnson, A.K., Thomson, D.U., Millman, S.T., 2024. The welfare of ill and injured feedlot cattle: a review of the literature and implications for managing feedlot hospital and chronic pens. Frontiers in Veterinary Science Volume 11 - 2024. https://doi.org/10.3389/fvets.2024.1398116.

Tatler, J., Cassey, P., Prowse, T.A.A., 2018. High accuracy at low frequency: detailed behavioural classification from accelerometer data. Journal of Experimental Biology 221. https://doi.org/10.1242/jeb.184085.

Terrell, S.P., Reinhardt, C.D., Larson, C.K., Vahl, C.I., Thomson, D.U., 2017. Incidence of lameness and association of cause and severity of lameness on the outcome for cattle on six commercial beef feedlots. J Am Vet Med Assoc 250, 437-445. https://doi.org/10.2460/javma.250.4.437.

Thompson, I.D., Bakhtiari, M., Rodgers, A.R., Baker, J.A., Fryxell, J.M., Iwachewski, E., 2012. Application of a high-resolution animal-borne remote video camera with global positioning for wildlife study: Observations on the secret lives of woodland caribou. Wildlife Society Bulletin 36, 365-370. https://doi.org/10.1002/wsb.130.

Versluijs, E., Niccolai, L.J., Spedener, M., Zimmermann, B., Hessle, A., Tofastrud, M., Devineau, O., Evans, A.L., 2023. Classification of behaviors of free-ranging cattle using

accelerometry signatures collected by virtual fence collars. Frontiers in Animal Science 4. https://doi.org/10.3389/fanim.2023.1083272.

Wang, L., Arablouei, R., Alvarenga, F.A.P., Bishop-Hurley, G.J., 2023. Classifying animal behavior from accelerometry data via recurrent neural networks. Computers and Electronics in Agriculture 206, 107647. https://doi.org/10.1016/j.compag.2023.107647.

Werbos, P.J., 1990. Backpropagation through time: what it does and how to do it. Proceedings of the IEEE 78, 1550-1560. https://doi.org/10.1109/5.58337.

Williams, M.L., James, W.P., Rose, M.T., 2017. Fixed-time data segmentation and behavior classification of pasture-based cattle: Enhancing performance using a hidden Markov model. Computers and Electronics in Agriculture 142, 585-596. https://doi.org/10.1016/j.compag.2017.11.010.

Williams, M.L., Mac Parthaláin, N., Brewer, P., James, W.P.J., Rose, M.T., 2016. A novel behavioral model of the pasture-based dairy cow from GPS data using data mining and machine learning techniques. Journal of Dairy Science 99, 2063-2075. https://doi.org/10.3168/jds.2015-10254.

Wyffels, S.A., Boss, D.L., Sowell, B.F., DelCurto, T., Bowman, J.G.P., McNew, L.B., 2020. Dormant season grazing on northern mixed grass prairie agroecosystems: Does protein supplement intake, cow age, weight and body condition impact beef cattle resource use and residual vegetation cover? PLOS ONE 15, e0240629. https://doi.org/10.1371/journal.pone.0240629.

Yamato, J., Ohya, J., Ishii, K., 1992. Recognizing human action in time-sequential images using hidden Markov model, Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 379-385.

Zhang, W., Yang, L., Geng, S., Hong, S., 2024. Self-Supervised Time Series Representation Learning via Cross Reconstruction Transformer. IEEE Transactions on Neural Networks and Learning Systems 35, 16129-16138. https://doi.org/10.1109/tnnls.2023.3292066.

Zimmerman, D.W., 1997. Teacher's Corner: A Note on Interpretation of the Paired-Samples t Test. Journal of Educational and Behavioral Statistics 22, 349-360. https://doi.org/10.3102/10769986022003349.

Zucchini, W., MacDonald, I.L., Langrock, R., 2017. Hidden Markov Models for Time Series: An Introduction Using R, Second Edition. CRC Press.

Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1, 3-14. https://doi.org/10.1111/j.2041-210X.2009.00001.x.