RESEARCH

Dairy Science and Management





Open-source carbon dioxide and volatile organic compound sensing and associations with defecation and urination events in horses

R. K. Wright¹, A. Ganino¹ and R. R. White^{1*}

Abstract

Management of non-point-source emissions from pastured livestock is complicated by spatial and temporal distribution of emissions and how they interplay with equally complex landscape typological distributions. Wearable sensing of CO₂ concentrations near the tailhead may enable real-time, spatially-explicit monitoring of manure emissions, if concentrations correlate with defecation and urination events. The objective of this research was to explore the association between measured CO₂ concentrations from wearable sensors placed on the tailhead of horses and the occurrence of defecation and urination events. CO₂ sensors consisted of a TTGO-T-Beam microprocessor equipped with GPS and LoRa radio, soldered to a CJMCU-8128 environmental sensing board capable of measuring temperature, pressure, relative humidity, CO₂ and total volatile organic compounds (TVOC). Tail wraps were placed on 4 stalled horses for a total of 9 days. Surveillance videos were collected over the same time frame and viewed to determine the time of defecation and urination occurrence. Data were analyzed visually for coherence, and quantitatively using analysis of variance, random forest regression, support vector machines, and extreme gradient boosting. Because defecation and urination events were in much lower quantity than non-events, random oversampling and undersampling were attempted on the classification approaches to improve accuracy and precision of signaling algorithms. Visual inspection revealed that although defecation and urination events corresponded to CO_2 peaks, there was considerable noise in CO₂ data suggesting that peaks in CO₂ also frequently occur in the absence of defecation and urination events. All classification algorithms showed poor accuracies (0.50 to 0.51), which were only marginally improved by over- (<0.51) and undersampling (<0.69). This preliminary assessment revealed considerable noise in sensing CO_2 emissions in production settings, which may preclude usefulness in manure sensing.

Keywords Manure, Sensing, Machine learning, Unbalanced classification, Horses

Background

Pasture based livestock are an important source of both food and income for many people [1]; however, they can also present certain environmental risks [2]. Although manure from livestock on pasture can be beneficial to

*Correspondence:

R. R. White

rrwhite@vt.edu

the land, it can also cause damage to waterways through runoff of components such as phosphorus and nitrogen, which can then affect the health of humans and animals that rely on the water source [3]. If nutrient emissions from livestock species, coming in the form of feces and urine, can be monitored, a more comprehensive understanding of when and where these emissions occur can be developed. Improved understanding of the geospatial and temporal emission of nutrients in livestock manure would allow more precise management of pasture



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

¹ School of Animal Sciences, Virginia Polytechnic Institute and State University, 101 Food Science and Technology Building, 360 Duck Pond Dr, Blacksburg, VA 24060, USA

landscapes, allowing operators to optimize the positive role of manure within the ecosystem.

Data exploring the location of defecation and urination events contributes to a greater capacity to manage grazing systems for sustainability objectives due to the geospatial heterogeneity of pasture systems. Identifying the location of manure emissions allows for understanding of the interplay between the distribution of manure deposition within the field, and the overlap with hydrologically sensitive areas. Although previous studies have attempted to explore the spatial distribution of manure through monitoring animal location, leveraging aerial imagery, and through manual observations, each of these approaches has some limitations [4–7]. Probabilistic modeling of manure location based on animal location and time budgets does not provide explicit monitoring of manure and can be prone to missing manure emission events that occur in low time-budget areas. Further, aerial imagery relies on drone technologies which do not consider the time dimension of emissions, nor are they robust during weather events which coincide with the greatest need to locate emissions due to potential for surface runoff. As such, there is a need to explore wearable behavioral sensing as a strategy for monitoring and locating manure emission events in grazing systems.

Wearable sensors for livestock behavioral monitoring are commonplace in confinement systems, where data transmissions can be completed using WiFi or Bluetooth, and data processing can be computationally intensive and housed centrally. In pasture systems, however, the need for low-energy, long-range data transmission means that alternative approaches to wearable behavior monitoring should be considered. Long range radio (LoRa) networks have been successful in pasture settings as a means of data transmission, and open source sensing systems operating on these LoRa networks show promise as a strategy to explore sensing options for animal behavioral detection based on sensed data [8]. However, LoRa networks require concise data packages for transmission, meaning that either small data must be sent for centralized signal processing to flag a behavior of interest or that edge data processing must occur so that only parsimonious signals are sent. Although working toward edge processing as a strategy to enable accurate and precise localization of manure emissions is a long-term goal, a step toward that goal requires exploration of the time-series of data obtained from sensors related to manure emissions, and subsequent analysis of that data for appropriate signals which could be used for edge processing.

An ideal sensor signal for monitoring manure emissions from livestock would have a clear signature associated only with the deposition of manure, which would allow for precise and accurate identification of emissions events. Although accelerometer-based behavioral monitoring has been used in numerous applications in the livestock industry [9, 10], there is some doubt about whether there is a clear motion signature associated with defecation and urination because tail motions associated with defecation or urination can be similar to other behaviors. Additionally, there is a need to differentiate between these types of events when considering potential environmental management. Although nutrients in defecation events are more easily managed because feces can be picked up and removed from the farm, nutrients in urine cannot be easily managed and are likely to infil-

The distinct nature of these nutrient emissions sources necessitates their individual identification. An alternative sensor, which may have greater specificity to defecation or urination behavior, would be sensing of CO_2 and total volatile organic compounds in the area around the tailhead. During the emission of feces and urine, volatile organic compounds (VOC) are released, suggesting that a strong signal should occur during these events. Further, there are few other livestock activities which result in localized spiking of VOC concentrations, suggesting this may be a precise and accurate signal to investigate when identifying defecation and urination events.

trate into the soil, affecting local soil nutrient contents.

The objective of this work was to explore, under controlled settings with minimal airflow interference, wearable CO_2 and VOC sensors as a strategy for identifying defecation and urination events in horses. Although we anticipated the sensors to have sensitive and specific signals for defecation and urination events, the possibility that signals would be complicated by environmental factors as well as emissions from other livestock justified the placement of sensors on the tailhead of individual animals.

Methods

Animals and experimental design

The animals involved in this study were under the management approved by the Virginia Polytechnic Institute and State University Institutional Animal Care and Use Committee (Protocol #19–159). Four mature (8 to 16 years old) geldings (n=2 Thoroughbred, n=2 Warmblood), were recruited for this study. Horses were housed in 12×12 box stalls in a well-ventilated barn. Stalls were used for this study, despite the interest in exploration in grazing environments, to explore a scenario where there would be minimal likelihood of interference with airflow and limited geographical area for the animals to cover. These conditions were selected to give the sensor the best chance at detecting defecation and urination events. Data collection occurred over a period of 2 weeks, during which time, horses were outfitted with an individual sensor on the tailhead and monitored by camera and visual assessment for 4 to 6 h. Monitoring sessions occurred between 08:00 AM and 17:00 PM. Gaps in monitoring occurred if horses were removed from their stalls due to use in lessons, vet/farrier visits, etc. Sensors were placed on tail wraps (Professional's Choice Inc, El Cajon, CA), which were then affixed around the tail head using a hook-and-loop closure. Tail wraps were placed with the CO_2 sensors fixed on the top side of the tail, facing away from the animal. Tail wraps equipped with sensors were removed following each collection period and placed again at the start of the next collection.

Sensor design and construction

The sensors consisted of a TTGO-T-Beam microprocessor (Shenzhen Xin Yuan Electronic Technology Co, China). This microprocessor is equipped with onboard GPS and RMF95 LoRa radio, which allowed for spaceand power-efficient design. A generic CJMCU-8128 environmental sensing board was then soldered to the microprocessor for sensing of the desired outcomes. The CJMCU-8128 integrates the CCS811 metal oxide gas sensor, the HDC1080 digital sensor, and the BMP280 absolute barometric pressure sensor to measure temperature, pressure, relative humidity, CO₂, and total volatile organic compounds (TVOC). The detection limits of temperature, pressure, relative humidity, CO2, and TVOC were -40 to 125 Celsius (C), 300 to 1,100 hectopascals (hPa), 0 to 100%, 400 to 8192 parts per million (ppm), and 0 to 1187 parts per billion (ppb), respectively. The sensor had an accuracy of $\pm 2\%$ for relative humidity, ± 0.12 hPa for pressure, and ± 0.2 C for temperature. Studies validating the use of these sensors have reported accuracies of ± 2 ppm for TVOC and up to 95% for CO₂, which is calculated rather than directly measured [11, 12]. The sensor suite was programmed using the Arduino IDE software, and formatted to read sensor data at 100 Hz, packaging those higher-density readings into an average of each 20 s interval. Although this sampling frequency rate is likely higher than necessary, it was selected considering the lack of previous work with gas sensors for detection of these types of events and the uncertainty around the results that would be obtained. The averaging was selected given the allowable frequency of LoRa data transmission. Upon completion of each 20 s interval of data averaging, the GPS was leveraged to obtain a location and time stamp, and data were packaged to send via LoRa. A Dragino LoRa gateway was used to receive data from the sensors using a standard LoRa protocol as described previously [8]. Data were then forwarded to the Virginia Tech Biological Systems Engineering server using an MQTT protocol. Sensors were powered by an 18,650 lithium ion battery and were charged via micro universal serial bus (USB) attachment prior to each deployment.

Determination of ground truth

To properly evaluate sensor efficacy at detecting defecation and urination events, sensor data was matched with ground-truth observations. In this study, video surveillance footage was recorded during the duration of data collection and used for ground-truth observation. While the sensors were active, surveillance videos were also active. Videos were manually observed and used to identify time timing of each defecation and urination event occurring during the sampling period. Time stamps from the sensors and surveillance footage were synchronized prior to data collection to allow alignment of sensor data and animal behaviors from video footage. Defecation events were marked by the passage of feces, while urination events were marked by the generation of urine. Other behaviors (i.e., standing, laying, etc.) were not differentiated, and coded only as non-events. The timestamps of defecation and urination events were used to correspond with the time stamps from the sensor to facilitate data analysis to explore associations among the CO₂ and TVOC readings and the occurrence of defecation and urination events. One to four defecation or urination events were observed per horse per period, with an average of 1.46 events per horse per period and an average of 4.5 events per horse across all sampling periods.

Data preparation

All analysis was completed in R, v 4.2.1. The server stored data files on a daily basis, and data were compiled by binding rows of data from each day file into a master dataset containing all received rows of data. The LoRa chirp sent by the sensor, forwarded by the router, and stored in the server was in the form of a string with character codes differentiating numerical data. Due to collisions among chips, occasionally these strings become corrupted, preventing their decoding. Corrupted strings were omitted from the analysis as a part of the first data cleaning step, resulting in 19,771 records available for analysis. During the second phase of data cleaning, the sensor data were visualized to determine raw distributions. Sensor data with erroneous CO₂ concentrations (>3,000 ppm) or temperature readings (>45 C) were omitted from the analysis as signal errors. Threshold values for erroneous sensor readings were determined by ensuring that all values in a normal range were accounted for and only those that could not be environmentally accurate were removed. The average global environmental CO₂ concentration is 421 ppm and the average temperature during the time of data collection was 13.3 C. After removing these erroneous data, 16,569 records were remaining.

Another anticipated type of data error was failed GPS measurements. Because this preliminary exploration of CO₂ and TVOC sensing was conducted in a barn, the GPS signal was often precluded by the barn roof. As a backup plan to allow for benchmarking the timestamp of data, the chirp included a locally logged millisecond counter and the time each sensor was turned on was manually recorded daily. Due to broad spectrum failure of GPS data, the starting time of each sensor and the locally logged timestep were used to compute the timestamps for each sensor. Behavioral ground truth data and the starting times for each sensor were recorded with minute, not second, precision, and therefore received chirp data were averaged by minute. Data were then merged with ground-truth observational data by animal and minute. The resulting dataset contained minutespecific sensor data aligned with codes reflecting a nonevent, a defecation event, or a urination event.

It was expected that the sensed signals for defecation or urination events may lag behind the observation of the actual event, due to delayed time for gasses to diffuse through the air. This delay was expected to be less than a minute, but could be up to several minutes. To explore the possibility that signals were delayed, rolling average CO2 and TVOC values were estimated using the previous 3, 5, 10, or 15 available observations. For these same time ranges, the standard deviation of observed values was also calculated, because it was postulated that the rolling average procedure might smooth out shortlived peaks in the data indicating defecation or urination events. To explore how a frameshift of data might be leveraged to relate to defecation or urination events, the 5 lagging and leading data points were also retained. Leading and lagging data points were retained by selecting the 5 data points before and 5 data points after the timestamp of a defecation or urination event. An example of the expected variation and feature engineering is shown in Fig. 1.

Data analysis

The data analysis was conducted in three steps. First, as a means of exploring whether a simple numerical cutoff on any particular measurement could be utilized, analysis of variance was used to explore the variation in measurements by observation type (defecation, urination, other). Estimated marginal means were calculated using the emmeans package and Tukey's pairwise comparisons were used to determine differences in each response variable among observation types. The signal values used included the rolling average and standard deviation of CO_2 and TVOC, as well as the raw and leading and lagging values.



Fig. 1 Example of variation in a single sensor's carbon dioxide (CO₂) concentrations (parts per million (ppm)) over a twenty-minute period and the engineered features designed to account for the variation. A defecation event occurs at 156 min, denoted by the vertical blue line and point 0. The trendline is represented in red. "RA-X" represents rolling averages, "RSD-X" represents rolling standard deviations, "Lag-X" represents lagging values, and "Lead-X" represents leading values. The data points used to calculate each of the engineered features are noted in both the table and figure with corresponding colors

It was not expected that a simple numerical cutoff would be identifiable to flag defecation and urination events, therefore, we also explored a number of machine learning approaches to tease out signals from the data. These approaches included extreme gradient boosting [13], random forest regression [14], and a support vector machine [15], and were selected based on the historical strength of these algorithms in similar classification tasks. Extreme gradient boosting is a robust machine learning algorithm that is designed to improve speed and model performance by increasing the efficiency of gradient boosting. Gradient boosting and extreme gradient boosting are both based on classifications or regressions that use weak predictions (decision trees) that make very few assumptions about data and decrease any bias [16]. In doing so, it is expected that each resulting decision tree has a lower prediction error and is closer to the target classification. This algorithm was used due to our previous success with this approach in low- and unbalanced data environments [17], and because it is an efficient stochastic boosting ensemble method with a good history of success in classification problems. Random forest regression uses ensemble learning to generate multiple random decision trees, each of which should be more reliable than any single model and has also shown promise in previous work on unbalanced datasets [18]. This algorithm works by selecting random samples from a given dataset and constructing a decision tree for each. Based on these, the final output is decided by majority voting to identify the best prediction, guaranteeing an accuracy of at least seventy percent [19]. Support vector machine is a supervised machine learning model for twogroup classification problems that works by identifying the optimal decision boundary that separates different classes. Within the sample space, a particular hyperplane can be identified that best separates the intended classes [20]. The support vector machine was selected due to its strength in highly dimensional data, as we expected the time-series element of the dataset would inflate dimensionality within the data. Each approach was trained on 60 percent of the available data and tested on 40 percent. The training dataset returned 56 defecation observations, 13 urination observations and 9,141 other observations. The testing dataset had 40 defecation, 7 urination, and 5972 other observations. To confirm the stability of the parameters used in each algorithm, model tuning was performed to explore shifts in predictive capacity under different settings. The support vector machine was checked using the tune function of the e1071 package [21], for costs ranging from 0.001 to 100. Similarly, the class weights, number of trees, and branches per tree in the random forest were systematically varied manually to explore different combinations of weighting and tree complexity as a strategy to improve algorithm performance using the randomForest package [22]. Finally, the extreme gradient boosting algorithm was trained over 100 rounds of cross validation using the xgb.cv function from the xgboost package [23] with differing weights for the class variables.

Because the defecation and urination data were very low incidence within the dataset, over- and under-sampling approaches were conducted on training data for each of the three algorithms. The upSample and down-Sample functions from the package groupdata2 were used to generate the training data for each approach [24]. With both techniques aiming to level unbalanced datasets, oversampling is used to add new samples to the minority class [25]. Opposingly, undersampling is used to reduce the number of samples in the majority class. The testing data were retained with their originally sampled distribution when computing accuracy and precision metrics, irrespective of classification algorithm used or of training data modification. The oversampling approach randomly over-sampled data from the low-incidence classes to return a dataset with equal representation among all classes. Similarly, the undersampling approach randomly omitted data from the over-represented classes until all classes had equal representation. All models were compared based on within-class accuracy, and additional metrics such as the sensitivity, specificity, positive and negative predictive values were reported for reference. Additionally, a general linear model was implemented to explore the main effects, two-, and three-way interactions of day, animal, and defecation and urination events on measured CO_2 obtained from the sensors.

Results

Trends in carbon dioxide and associations with defecation and urination

The mean and distribution of CO_2 emissions detected by the sensors in this study matched likely environmental CO_2 concentrations within the barn (Table 1). Although no ground-truth observation of barn-level CO_2 was collected during the sensing period, the profile and variation of CO_2 sensed by these tail-mounted sensors was consistent with previous barn-based observations (Fig. 2). Furthermore, the three-way interaction of animal, day, and defecation and urination events was significant (P < 0.001) suggesting that shifts in CO_2 associated with defecation and urination events were not consistent across animals and days.

Visual inspection of data revealed that although there were peaks in CO_2 during defecation and urination events, there was considerable noise within the data, possibly attributed to other behavioral events (Fig. 3). Despite this, the analysis of variance suggested consistent

Measurement	Mean	Median	SD ^c	Min ^c	Q1 ^c	Q3 ^c	Max ^c
CO ₂ , ppm ^a	566	496	248	400	448	575	2,936
TVOC, ppm ^b	24.7	14.0	37	0.0	7.0	26.0	386
Temperature, C	18.4	18.4	3.3	10.1	16.1	20.7	27.9

Table 1 Summary statistics for available sensed data collected during experimental period

 $\overline{\ }^{a}$ Carbon dioxide concentrations sensed by the open source sensor (CO₂)

^b Total volatile organic compound concentrations (TVOC) sensed by the open source sensor

^c Standard deviation (SD), minimum (Min), quartile 1 (Q1), quartile 3 (Q3), and maximum (Max) values sensed



Fig. 2 Smoothed carbon dioxide concentrations measured among sensors across experimental days. The average pattern of CO_2 sensing showed a gradual rise in CO_2 concentrations, likely driven by bringing the animal into the barn at the start of sampling. Thereafter, there is considerable natural noise within the data reflecting the various potential CO_2 sources in the environment



Fig. 3 An example day of observed CO_2 concentration data, with defecation and urination events marked with the symbol "E" and a vertical line. The lines show general coherence with spikes in CO_2 , however numerous spikes also exist in the absence of emission events

differences in sensor readings associated with defecation events, but not urination events (Table 2). The visual inspection of available data also suggests that the analysis of variance was not effective as a means of identifying robust strategies to differentiate defecation and urination events from other behaviors (Fig. 3).

Comparison of machine learning approaches to classify defecation and urination

Random forest classification, support vector machine, and extreme gradient boosting all largely failed to adequately classify defecation and urination events within the dataset (Table 3). When trained on the raw, highly imbalanced data, all three algorithms resulted in complete prediction of non-events. Although this resulted in high accuracy (0.994), the balanced, class-specific accuracies were 50%.

All models were tuned to confirm the stability of the parameters. The tuning of the support vector machine resulted in consistent results (data not shown). Much like the support vector machines, modifying the parameters of the random forest regression did not improve model performance. The extreme gradient boosting algorithm was trained over 100 rounds of cross validation, also

Table 2	Carbon dioxide,	TVOC, and temperatu	re associations with no	n-events, defecation,	, and urination
---------	-----------------	---------------------	-------------------------	-----------------------	-----------------

Measurement	Time Scale	Other	SE	Defecation	SE	Urination	SE	P-Value
CO ₂ , ppm ⁺	0	565	1.98	613	24.6	539	29.5	0.134
	3	565ª	1.59	619 ^b	19.8	517 ^a	39.6	0.013
	5	566ª	1.44	606 ^b	18.0	506 ^{ab}	35.8	0.022
	10	556 ^a	1.33	586 ^{ab}	16.8	495 ^b	34.2	0.056
	15	566	1.28	583	16.1	516	35.3	0.219
TVOC, ppm [*]	3	24.7	0.30	32.0	3.75	20.7	7.55	0.133
	5	24.7 ^a	0.24	32.8 ^b	3.01	17.3 ^a	6.04	0.013
	10	24.8 ^a	0.22	30.8 ^b	2.74	15.7 ^{ab}	5.46	0.022
	15	24.9	0.20	27.3	2.46	17.2	5.38	0.222
Temperature, C	0	18.4 ^a	0.026	19.7 ^b	0.033	17.9 ^a	0.66	< 0.001
	3	18.4 ^a	0.026	19.7 ^b	0.32	17.8 ^a	0.64	< 0.001
	5	18.4 ^a	0.025	19.7 ^b	0.31	17.7 ^a	0.62	< 0.001
	10	18.4 ^a	0.025	19.6 ^b	0.31	17.7 ^a	0.64	< 0.001
	15	18.4 ^a	0.025	19.7 ^b	0.31	17.9 ^a	0.69	< 0.001

⁺ Carbon dioxide concentrations sensed by the open source sensor (CO₂)

* Total volatile organic compound concentrations (TVOC) sensed by the open source sensor

Table 3 Predictive capa	city of algorithms fit usin	g raw distribution inherent	in sampled data
-------------------------	-----------------------------	-----------------------------	-----------------

	Random Forest			Support Vector Machine			Extreme Gradient		
	Other	Defª	Urine	Other	Def ^a	Urine	Other	Def ^a	Urine
Predicted									
Other	5971	39	7	5972	40	7	5970	39	7
Def ^a	1	1	0	0	0	0	2	1	0
Urine ^a	0	0	0	0	0	0	0	0	0
Class Statistics									
Sensitivity	1.00	0.03	0.00	1.00	0.00	0.00	1.00	0.025	0.00
Specificity	0.02	1.00	1.00	0.00	1.00	1.00	0.021	1.00	1.00
PPV ^a	0.992	0.50	0.00	0.992	0.00	0.00	0.992	0.33	0.00
NPV ^a	0.50	0.994	0.999	0.00	0.993	0.999	0.33	0.994	0.999
Accuracy	0.51	0.51	0.50	0.50	0.50	0.50	0.51	0.51	0.50
Overall Statistics									
Accuracy	0.992			0.992			0.992		
NIR ^a	0.992			0.992			0.992		

^a Defecation (Def) and urination (Urine) events, positive predictive value (PPV), negative predictive value (NPV), and null information rate (NIR)

resulting in no meaningful difference from the baseline results. As an alternative approach to handling the imbalance within the data, over- (Table 4) and under- (Table 5) sampling were used to train the models. Although oversampling resulted in marginal shifts in the class-specific balanced accuracy, resulting in improved predictions of defecation and urination events by the support vector machine (Table 4), the resulting accuracies were still well below appropriate to justify use of the sensor for prediction purposes. The accuracies achieved by the random forest and the extreme gradient boosting algorithms did not change with oversampling (Table 4). Much like oversampling, the undersampling was not effective at substantively enhancing model performance in classifying defecation and urination events. Although the within-class accuracies for predicting defecation events improved to 0.69 and 0.62 for the random forest and extreme gradient boosting, respectively (Table 5), the overall accuracies were extremely poor on account of the large false positive rates.

Table 4 Predictive of	capacity of algorithms	fit using data ov	versampled to balance	across classes in s	ampled data
-----------------------	------------------------	-------------------	-----------------------	---------------------	-------------

	Random Forest			Support Vector Machine			Extreme Gradient		
	Other	Defª	Urine	Other	Def ^a	Urine	Other	Def ^a	Urine
Predicted									
Other	5971	39	7	5972	40	7	5972	40	7
Def ^a	0	1	0	0	0	0	0	0	0
Urine ^a	0	0	0	0	0	0	0	0	0
Class Statistics									
Sensitivity	1.00	0.025	0.00	1.00	0.00	0.00	1.00	0.00	0.00
Specificity	0.02	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
PPV ^a	0.992	1.00	0.00	0.992	0.00	0.00	0.992	0.00	0.00
NPV ^a	1.00	0.994	0.999	0.00	0.993	0.999	0.00	0.993	0.999
Accuracy	0.51	0.51	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Overall Statistics									
Accuracy	0.992			0.992			0.992		
NIR ^a	0.992			0.992			0.992		

^a Defecation (Def) and urination (Urine) events, positive predictive value (PPV), negative predictive value (NPV), and null information rate (NIR)

Table 5	Predictive capacit	y of algorithms fit using	g data undersam	pled to balance across	classes in sampled data
---------	--------------------	---------------------------	-----------------	------------------------	-------------------------

	Random Forest			Support Vector Machine			Extreme Gradient		
	Other	Def ^a	Urine ^a	Other	Def ^a	Urine ^a	Other	Def ^a	Urine ^a
Predicted									
Other	1768	2	2	2449	14	2	1281	4	0
Def ^a	2465	32	3	1748	16	3	3241	31	6
Urine ^a	1739	6	2	1775	10	2	1450	5	1
Class Statistics									
Sensitivity	0.29	0.80	0.29	0.41	0.40	0.29	0.21	0.78	0.14
Specificity	0.91	0.59	0.71	0.66	0.71	0.70	0.91	0.46	0.76
PPV ^a	1.00	0.01	0.001	0.994	0.009	0.001	0.997	0.009	0.001
NPV ^a	0.01	0.998	0.999	0.009	0.994	0.999	0.009	0.997	0.999
Accuracy	0.61	0.69	0.50	0.53	0.55	0.49	0.56	0.62	0.45
Overall Statistics									
Accuracy	0.299			0.410			0218		
NIR ^a	0.992			0.992			0.992		

^a Defecation (Def) and urination (Urine) events, positive predictive value (PPV), negative predictive value (NPV), and null information rate (NIR)

Discussion

Trends in carbon dioxide and associations with defecation and urination

Microprocessor sensors have previously been used to explore CO₂ concentrations in confined animal feeding operations [26]. Although some studies measuring CO_2 concentration in barns show periods of time with fairly constant concentrations similar to ambient air [27], other studies have found considerable variability in concentrations throughout the day [28]. The inconsistent shifts in CO₂ associated with defecation and urination events across animals and days could indicate micro-environments within the barns which create disrupted capacity to sense gaseous emissions or concentrations effectively. Nevertheless, the capacity of these CO₂ sensors, which are typically designed for indoor use, to monitor CO₂ concentrations in the air around the animals' tailheads is supported by the range and profile of observations collected.

Although many metrics resulted in different mean observations for defecation events, which were often detected as different from urination or non-events, the standard errors around the defecation means were wide enough to suggest that a simple cutoff on one or more metrics would be insufficient to adequately flag defecation events. Furthermore, the lack of consistent difference during urination events indicated large variability around sensor readings during urination events. Although helpful as a screening exercise to explore possible variables for inclusion in the machine learning algorithms, the analysis of variance was not effective as a means of identifying robust strategies to differentiate defecation and urination events from other behaviors.

Comparison of machine learning approaches to classify defecation and urination

Classifying highly imbalanced data is a major challenge for most classification algorithms [29], and frequently requires use of complex ensemble approaches [30] or other specialized machine learning approaches [31]. However, each of the three approaches taken have previously shown promise as a strategy to classify imbalanced data [17, 31, 32]. As such, it was expected that the algorithms would perform better than was observed.

Similarly, while weighting and careful model tuning have been supported as strategies to handle lowincidence data in the past [33], neither was effective at improving algorithm performance on these data. As another strategy for managing unbalanced data, the over- and under-sampling approaches were conducted. Although oversampling has been widely reviewed as a technique to address the limitations of imbalanced data [34], many example datasets have imbalance which is less severe than observed in this case. For example, [35] reported oversampling was an effective strategy to deal with classes at a 0.56:1 ratio. In our data, the defecation class is at a 0.007:1 ratio. The failure of oversampling to improve classification accuracy for these algorithms likely reflects the challenge of simply attempting to pick a needle from a haystack. Additionally, although undersampling is frequently leveraged as a strategy to address imbalance in classification datasets, random undersampling, as used here, can lead to poor classification results because it fails to take into consideration extremely informative samples in the majority class [36]. Overall, the incidence of defecation and urination events within the dataset was so low that even these traditional strategies for handling imbalanced data were unable to significantly improve accuracy.

Conclusions

Based on the time-series data, there seems to be merit in concluding that CO_2 concentrations do spike during defecation and urination occurrence; however, these concentrations also spike for a variety of other reasons, unrelated to measured defecation and urination events. As such, the practical application of CO_2 sensing for environmental monitoring of livestock either in confined environments, as studied here, or in pastured environments, as was the longer-term goal of this initial analysis, may be impractical. Further work in pasture environments is necessary to determine whether the considerable variability in CO_2 observations reported in this study is due to the barn environment, where numerous other CO_2 sources are present, or whether pastured environments also result in similar variability.

Acknowledgements

Not applicable.

Authors' contributions

RKW contributed to data analysis and interpretation and is responsible for the drafting and revision of the presented work. AG is responsible for the data acquisition. RRW designed the study and contributed to data analysis and interpretation. RRW is also responsible for revisions and is the corresponding author. All authors read and approved the final manuscript.

Funding

This work was supported by funds appropriated to the Virginia Tech College of Agriculture and Life Sciences and by USDA-NIFA-AFRI project 2021–67021-34769.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 26 September 2024 Accepted: 14 February 2025 Published online: 03 March 2025

References

- Sloat LL, Gerber JS, Samberg LH, Smith WK, Herrero M, Ferreira LG, et al. Increasing importance of precipitation variability on global livestock grazing lands. Nat Clim Chang. 2018;8(3):214–8.
- Hubbard R, Newton G, Hill G. Water quality and the grazing animal. J Animal Sci. 2004;82(13):E255–63.
- Roche LM, Kromschroeder L, Atwill ER, Dahlgren RA, Tate KW. Water quality conditions associated with cattle grazing and recreation on national forest lands. PLoS ONE. 2013;8(6):e68127.
- Oudshoorn FW, Kristensen T, Nadimi ES. Dairy cow defecation and urination frequency and spatial distribution in relation to time-limited grazing. Livest Sci. 2008;113(1):62–73.
- Petersen R, Lucas H, Woodhouse W Jr. The Distribution of Excreta by Freely Grazing Cattle and Its Effect on Pasture Fertility: I Excretal Distribution 1. Agronomy J. 1956;48(10):440–4.
- Maire J, Gibson-Poole S, Cowan N, Krol D, Somers C, Reay DS, et al. Can nitrogen input mapping from aerial imagery improve nitrous oxide emissions estimates from grazed grassland? Precision Agric. 2022;23(5):1743–74.
- Fraser RH, Barten PK, Pinney DA. Predicting stream pathogen loading from livestock using a geographical information system-based delivery model. Wiley Online Library; 1998. Report No.: 0047–2425.
- Dos Reis BR, Easton Z, White RR, Fuka D. A LoRa sensor network for monitoring pastured livestock location and activity. Translational Animal Science. 2021;5(2):txab010.
- Bailey DW, Trotter MG, Knight CW, Thomas MG. Use of GPS tracking collars and accelerometers for rangeland livestock production research. Translational Animal Science. 2018;2(1):81–8.
- Brennan J, Johnson P, Olson K. Classifying season long livestock grazing behavior with the use of a low-cost GPS and accelerometer. Comput Electron Agric. 2021;181:105957.
- 11. Widhowati AA, Wardoyo AYP, Dharmawan HA, Nurhuda M, Budianto A, editors. Development of a Portable Volatile Organic Compounds Concentration Measurement System Using a CCS811 Air Quality Sensor. 2021 International Symposium on Electronics and Smart Devices (ISESD); 2021: IEEE.
- Komarudin M, Sulistyanti SR, Irsyad M, Septama HD, Yulianti T, editors. Improving Low-Cost Carbon Dioxide Sensor Accuracy for Environmental Air Quality Monitoring Systems. 2023 International Conference on Converging Technology in Electrical and Information Engineering (ICCTEIE); 2023: IEEE.
- Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost: extreme gradient boosting. R package version 04-2. 2015;1(4):1–4.
- Liu Y, Wang Y, Zhang J, editors. New machine learning algorithm: Random forest. Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14–16, 2012 Proceedings 3; 2012: Springer.
- Suthaharan S, Suthaharan S. Support vector machine. Machine learning models and algorithms for big data classification: thinking with examples for effective learning. 2016:207–35.
- Ramraj S, Uzir N, Sunil R, Banerjee S. Experimenting XGBoost algorithm for prediction and classification of different datasets. Int J Control Theory Appl. 2016;9(40):651–62.
- Liebe D, Hall M, White RR. Contributions of dairy products to environmental impacts and nutritional supplies from United States agriculture. J Dairy Sci. 2020;103(11):10867–81.
- Boateng EY, Otoo J, Abaye DA. Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review. J Data Anal Inform Processing. 2020;8(4):341–57.

- Schonlau M, Zou RY. The random forest algorithm for statistical learning. Stand Genomic Sci. 2020;20(1):3–29.
- Zhou Z-H. Support Vector Machine. In: Machine learning. Singapore: Springer; 2021. p. 129–53. https://doi.org/10.1007/978-981-15-1967-3_6.
- 21. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: Misc Functinos of the Department of Statistics, Probability Theory Grooup (Formerly: E1071), TU Wien. 2023.
- 22. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2(3):18–22.
- 23. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. xgboost: Extreme Gradient Boosting. 2023.
- 24. Olsen LR. groupdata2: Creating Groups from Data. 2024.
- Shelke MS, Deshmukh PR, Shandilya VK. A review on imbalanced data handling using undersampling and oversampling technique. Int J Recent Trends Eng Res. 2017;3(4):444–9.
- Pate ML, Hofstetter D, editors. A one-health approach to safety instruction: A framework for livestock housing environment monitoring using arduino-based sensors. 2020 ASABE Annual International Virtual Meeting; 2020: American Society of Agricultural and Biological Engineers.
- Mendes LB, Ogink NW, Edouard N, Van Dooren HJC, Tinoco IDFF, Mosquera J. NDIR gas sensor for spatial monitoring of carbon dioxide concentrations in naturally ventilated livestock buildings. Sensors. 2015;15(5):11239–57.
- Vtoryi V, Vtoryi S, Lantsova E, Gordeev V, editors. Effect of weather conditions on content of carbon dioxide in barns. Proceedings 15th International Scientific Conference «Engineering for rural development; 2016.
- Kumar P, Bhatnagar R, Gaur K, Bhatnagar A, editors. Classification of imbalanced data: review of methods and applications. IOP conference series: materials science and engineering; 2021: IOP Publishing.
- Sun Z, Song Q, Zhu X, Sun H, Xu B, Zhou Y. A novel ensemble method for classifying imbalanced data. Pattern Recogn. 2015;48(5):1623–37.
- Ganganwar V. An overview of classification algorithms for imbalanced datasets. International Journal of Emerging Technology and Advanced Engineering. 2012;2(4):42–7.
- More A, Rana DP, editors. Review of random forest classification techniques to resolve data imbalance. 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM); 2017: IEEE.
- Effendy V, Baizal ZA, editors. Handling imbalanced data in customer churn prediction using combined sampling and weighted random forest. 2014 2nd International Conference on Information and Communication Technology (ICoICT); 2014: IEEE.
- 34. Gosain A, Sardana S, editors. Handling class imbalance problem using oversampling techniques: A review. 2017 international conference on advances in computing, communications and informatics (ICACCI); 2017: IEEE.
- Adhikari S, Thapa S, Shah BK, editors. Oversampling based classifiers for categorization of radar returns from the ionosphere. 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC); 2020: IEEE.
- Ng WW, Hu J, Yeung DS, Yin S, Roli F. Diversified sensitivity-based undersampling for imbalance classification problems. IEEE transactions on cybernetics. 2014;45(11):2402–12.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.