

1 Improving bovine health outcomes and management strategies using machine learning in cattle
2 production systems
3

4
5 by
6

7
8 Lilli Heinen
9

10
11
12 B.S., Kansas State University, 2022
13 DVM, Kansas State University, 2024
14

15
16
17 AN ABSTRACT OF A DISSERTATION
18

19
20
21 submitted in partial fulfillment of the requirements for the degree
22

23
24
25 DOCTOR OF PHILOSOPHY
26

27
28 Department of Diagnostic Medicine and Pathobiology
29 College of Veterinary Medicine
30

31
32
33 KANSAS STATE UNIVERSITY
34 Manhattan, Kansas
35

36
37 2025
38
39

Abstract

The objective of this dissertation is to explore the use of machine learning to enhance health and management of cattle in beef and dairy systems. One chapter is dedicated to a review of the literature of machine learning application to the diagnosis and prognosis of cattle diseases in the beef and dairy production sectors. A total of four studies were designed and conducted to determine the predictive ability of machine learning algorithms to determine health outcomes or ideal management strategies in beef cattle in confined feeding operations at both group- and individual animal-levels. These four studies focused on bovine respiratory disease as the major health concern. The remaining study evaluated post-treatment intervals in stocker cattle treated with a novel antimicrobial for bovine respiratory disease.

Two studies focused on the prediction of a group-level outcome using various machine learning algorithms. One study aimed to predict morbidity class (high = $\geq 15\%$ or low = $< 15\%$) of pens of feedlot cattle when given demographic data and feed delivery data from the first 15 days on feed. The algorithms tested included advanced perceptron, logistic regression, neural network, boosted decision tree, and random forest models. Model performance was poor as determined by area under the receiver operating characteristics curve (AUC) but showed promise for determining which pens of cattle would have low morbidity using specificity as a performance metric. Non-linear algorithms tended to perform better than logistic regression. The second study aimed to predict optimal metaphylaxis application strategy (yes or no) to groups of cattle arriving to the feedlot given demographic, origin, and external economic data. Algorithms tested were similar to the first study. High AUC values were achieved when models were given demographic data alone. Model performance based on AUC increased given external economic data but decreased given origin data. Again, non-linear models had higher predictive power than logistic regression.

Two studies focused on the prediction of individual animal outcome at the time of first or second treatment for bovine respiratory disease in feedyard cattle. One study aimed to predict mortality or early culling (did not finish = yes or no) when given individual animal-, group-, and feedyard-level data on demographics. Additionally, the addition of weather data, the impact of a balancing strategy applied to the training data, and the development of individualized models on predictive model performance was tested. Similar algorithms to the previously mentioned studies

were utilized. Models performed well with moderate AUC values. No improvement in AUC values were noted with the addition of weather data or the application of a balancing strategy. Creation of individualized models for each feedyard resulted in increased predictive performance for some feedyards but not others. The second study determined the economic effect of using a predictive model to cull sick animals at the time of diagnosis for first or second treatment of bovine respiratory disease. In this study, only a boosted decision tree model was investigated. It was found that minimizing the number of false positives (i.e. animals that the model has determined will die but would have actually lived) is the biggest driver of positive net return. Therefore, the specificity of the predictive model is the most important metric when the goal of using a predictive model to make culling decisions is profit.

A final study was a prospective trial to investigate the effect of post-treatment interval on health outcomes following administration of pradofloxacin for the treatment of bovine respiratory disease in stocker cattle. Three-day, six-day, and nine-day post-treatment intervals were investigated in a multi-site trial. Post-treatment interval had no effect on first treatment success or case fatality. Additionally, there was no difference in number of days to death from initial treatment amongst the mortalities by post-treatment interval.

In conclusion, the current dissertation has contributed to the body of literature surrounding the use of machine learning algorithms for health outcomes in cattle production systems. By investigating various algorithms, data types, and strategies for machine learning application to cattle health, this work will provide a foundation for future research. Additionally, this work provides a background on predictive model development and proper model training strategies that could be applied to other livestock production species.

Improving bovine health outcomes and management strategies using machine learning in cattle
production systems

by

Lilli Heinen

B.S., Kansas State University, 2022
DVM, Kansas State University, 2024

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Diagnostic Medicine and Pathobiology
College of Veterinary Medicine

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Major Professor
Dr. Brad J. White

132

Copyright

133

© Lilli Heinen 2025.

134

135

136

Abstract

The objective of this dissertation is to explore the use of machine learning to enhance health and management of cattle in beef and dairy systems. One chapter is dedicated to a review of the literature of machine learning application to the diagnosis and prognosis of cattle diseases in the beef and dairy production sectors. A total of four studies were designed and conducted to determine the predictive ability of machine learning algorithms to determine health outcomes or ideal management strategies in beef cattle in confined feeding operations at both group- and individual animal-levels. These four studies focused on bovine respiratory disease as the major health concern. The remaining study evaluated post-treatment intervals in stocker cattle treated with a novel antimicrobial for bovine respiratory disease.

Two studies focused on the prediction of a group-level outcome using various machine learning algorithms. One study aimed to predict morbidity class (high = $\geq 15\%$ or low = $< 15\%$) of pens of feedlot cattle when given demographic data and feed delivery data from the first 15 days on feed. The algorithms tested included advanced perceptron, logistic regression, neural network, boosted decision tree, and random forest models. Model performance was poor as determined by area under the receiver operating characteristics curve (AUC) but showed promise for determining which pens of cattle would have low morbidity using specificity as a performance metric. Non-linear algorithms tended to perform better than logistic regression. The second study aimed to predict optimal metaphylaxis application strategy (yes or no) to groups of cattle arriving to the feedlot given demographic, origin, and external economic data. Algorithms tested were similar to the first study. High AUC values were achieved when models were given demographic data alone. Model performance based on AUC increased given external economic data but decreased given origin data. Again, non-linear models had higher predictive power than logistic regression.

Two studies focused on the prediction of individual animal outcome at the time of first or second treatment for bovine respiratory disease in feedyard cattle. One study aimed to predict mortality or early culling (did not finish = yes or no) when given individual animal-, group-, and feedyard-level data on demographics. Additionally, the addition of weather data, the impact of a balancing strategy applied to the training data, and the development of individualized models on predictive model performance was tested. Similar algorithms to the previously mentioned studies

167 were utilized. Models performed well with moderate AUC values. No improvement in AUC
168 values were noted with the addition of weather data or the application of a balancing strategy.
169 Creation of individualized models for each feedyard resulted in increased predictive performance
170 for some feedyards but not others. The second study determined the economic effect of using a
171 predictive model to cull sick animals at the time of diagnosis for first or second treatment of
172 bovine respiratory disease. In this study, only a boosted decision tree model was investigated. It
173 was found that minimizing the number of false positives (i.e. animals that the model has
174 determined will die but would have actually lived) is the biggest driver of positive net return.
175 Therefore, the specificity of the predictive model is the most important metric when the goal of
176 using a predictive model to make culling decisions is profit.

177 A final study was a prospective trial to investigate the effect of post-treatment interval on
178 health outcomes following administration of pradofloxacin for the treatment of bovine
179 respiratory disease in stocker cattle. Three-day, six-day, and nine-day post-treatment intervals
180 were investigated in a multi-site trial. Post-treatment interval had no effect on first treatment
181 success or case fatality. Additionally, there was no difference in number of days to death from
182 initial treatment amongst the mortalities by post-treatment interval.

183 In conclusion, the current dissertation has contributed to the body of literature
184 surrounding the use of machine learning algorithms for health outcomes in cattle production
185 systems. By investigating various algorithms, data types, and strategies for machine learning
186 application to cattle health, this work will provide a foundation for future research. Additionally,
187 this work provides a background on predictive model development and proper model training
188 strategies that could be applied to other livestock production species.

191	Table of Contents	
192	List of Figures	xii
193	List of Tables	xiv
194	Acknowledgements	xvi
195	Dedication	xvii
196	Chapter 1 - Disease prediction in beef cattle: a narrative review of the predictive model literature	
197		1
198	Abstract	1
199	Introduction.....	2
200	Materials and Methods.....	3
201	Predictive model performance metrics	4
202	Results.....	5
203	Discussion	7
204	References	11
205	Chapter 2 - Evaluation of predictive models to determine total morbidity outcome of feedlot	
206	cattle based on pen-level feed delivery data during the first 15 days on feed	26
207	Abstract	26
208	Introduction.....	27
209	Materials and Methods.....	28
210	Data Collection	28
211	Data Transformation	28
212	Model Evaluation.....	29
213	Results and Discussion	30
214	Conclusion	32
215	References	32
216	Chapter 3 - Evaluation of predictive models to determine final outcome for feedlot cattle based	
217	on information available at first treatment for bovine respiratory disease	1
218	Abstract	1
219	Introduction.....	2
220	Methods	3

221	Data collection	3
222	Data transformation	3
223	Model-building process	4
224	Model evaluation	5
225	Weather versus Base models.....	6
226	Balanced versus Base models	6
227	Individual versus Base models.....	7
228	Results.....	7
229	Evaluation of Base models.....	7
230	Comparison of weather and Base models	7
231	Comparison of balanced versus Base models	8
232	Comparison of individual versus Base models	8
233	Discussion.....	8
234	References.....	13
235	Chapter 4 - Economic impact of mortality prediction by predictive model at first and second	
236	treatment for bovine respiratory disease.....	21
237	Abstract.....	21
238	Introduction.....	22
239	Methods	23
240	Data collection	23
241	Feedyard data transformation	24
242	Model building process.....	25
243	Economic analysis	26
244	Results.....	29
245	Evaluation of models	29
246	Economic analysis results	29
247	Discussion.....	30
248	References.....	33
249	Chapter 5 - Predictive models to determine best strategy for metaphylaxis application in cattle at	
250	arrival to a feedyard	42
251	Abstract.....	42

252	Introduction.....	43
253	Materials and methods	45
254	Ethical statement.....	45
255	Data sources	45
256	Feedyard data	45
257	Cattle price data	47
258	Feed data and cost calculation	48
259	Treatment cost.....	49
260	Economic analysis to calculate returns	49
261	Economic analysis for existing metaphylaxis status.....	49
262	Economic simulation to determine returns based on opposite metaphylaxis status	50
263	Predictive modeling	51
264	Creation of datasets for predictive models.....	51
265	Predictive modeling	52
266	Boosted decision tree (BDT)	54
267	Logistic regression (LR)	55
268	Neural network (NN)	55
269	Random forest (RF)	56
270	Parameters.....	56
271	Baseline.....	56
272	Baseline + Econ	57
273	Baseline + Origin	57
274	Baseline + Origin + Econ.....	57
275	Results.....	58
276	Descriptive statistics	58
277	Baseline.....	58
278	Baseline + Econ	59
279	Baseline Subset	59
280	Baseline + Origin	59
281	Baseline + Origin + Econ.....	60
282	Feature importance.....	60

283	Discussion.....	60
284	References.....	67
285	Chapter 6 - Evaluation of 3-, 6-, and 9-day post-treatment intervals following pradofloxacin for	
286	treatment of bovine respiratory disease in stocker cattle.....	79
287	Abstract.....	79
288	Introduction.....	80
289	Materials and Methods.....	81
290	Sample size calculation.....	81
291	Study population	82
292	Inclusion criteria	82
293	Feed and housing	83
294	Enrollment.....	84
295	Observation period.....	85
296	Statistical analysis.....	85
297	Results.....	86
298	First treatment success	86
299	Case fatality and days to death.....	87
300	Discussion.....	87
301	End Notes.....	90
302	References.....	91
303	Appendix A - Supplement to Chapter 3.....	94
304	Appendix B - Supplement to Chapter 4.....	98
305	Appendix C - Supplement to Chapter 5.....	108
306		

List of Figures

307		
308	Figure 1.1. Box-and-whisker plots of predictive model accuracy by disease outcome and	
309	algorithm category extracted from reviewed articles. Points represent individual study arms.	
310		18
311	Figure 1.2. Box-and-whisker plots of predictive model sensitivity and specificity by disease	
312	outcome and algorithm category extracted from reviewed articles. Points represent	
313	independent study arms.....	19
314	Figure 1.3. Box-and-whisker plots of predictive model negative and positive predictive values	
315	by disease outcome and algorithm category extracted from reviewed articles. Points	
316	represent independent study arms.	20
317	Figure 2.1. Timeline schematic demonstrating the feed delivery data corresponding to various	
318	feeding predictor variables. Triangles indicate data on dry matter delivered as percentages	
319	of arrival body weight (DMI-BW). Arrows indicate data on percent change in DMI-BW	
320	from day to day.	35
321	Figure 2.2. Illustration of data management, model training and evaluation process.	36
322	Figure 2.3. Receiver operating characteristic (ROC) curves for five predictive models trained to	
323	predict high ($\geq 15\%$) morbidity cohorts of feedlot cattle. The five predictive models are	
324	Advanced Perceptron, Logistic Regression, Neural Network, Decision Tree, and Decision	
325	Forest. Perf. Pred. represents the perfect predictive model.	37
326	Figure 3.1. Flowchart describing the steps followed to create the base predictive models	
327	investigated in this trial. Inclusion criteria include a first-time treatment for bovine	
328	respiratory disease with an antimicrobial.....	16
329	Figure 3.2. Flowchart describing the extra steps followed to balance the training dataset with	
330	respect to minority outcome class before model training. This process allowed for the	
331	creation of 3 model sets built with dataset of varying prevalence of did not finish: 14.2%	
332	(Base) and 50% (Oversample and Undersampled).	17
333	Figure 3.3. AUC in Base predictive models including all feedyards compared to individual	
334	feedyard models. Each point demonstrates the AUC value for 5 predictive models (specific	
335	model designated by color).	18

336	Figure 4.1. Flowchart summarizing exclusion criteria and dataflow. BRD1 refers to first	
337	treatment for BRD. BRD2 refers to second treatment for BRD.	36
338	Figure 4.2. Illustration describing the steps followed to train, test, and validate the boosted	
339	decision tree models. These steps were performed for each dataset (BRD1 and BRD2).	37
340	Figure 4.3. Confidence matrix describing the costs or profits associated with each possible	
341	model outcome for the BRD1 model. RP = railer price, FC _{TN} = adjusted feed cost for true	
342	negative, SP = sale price, TC = treatment cost, Risk _{DNF} = risk of DNF adjustment	38
343	Figure 4.4. Difference in net returns using the model compared to status quo for BRD1 and	
344	BRD2	39
345	Figure 5.1. Flowchart demonstrating the different variables, datasets, and number of cohorts	
346	included in each dataset.	74
347	Figure 5.2. Bar chart demonstrating the average AUC-ROC values for the Baseline and the	
348	Baseline + Econ models. Error bars represent the standard deviation. BDT = boosted	
349	decision tree, LR = logistic regression, NN = neural network, RF = random forest.	75
350	Figure 5.3. Bar chart demonstrating the average AUC-ROC values for the Baseline Subset,	
351	Baseline + Origin, and Baseline + Origin + Econ models. Error bars represent the standard	
352	deviation. BDT = boosted decision tree, LR = logistic regression, NN = neural network, RF	
353	= random forest	76
354		
355		

List of Tables

357	Table 1.1. Inclusion and exclusion criteria applied to articles for literature review selection	21
358	Table 1.2. Description of data collected from articles included in narrative literature review....	22
359	Table 1.3. Distribution of articles and study arms from literature review by input data type and	
360	disease outcome category.....	24
361	Table 1.4. Description of health outcomes and prevalence of health outcomes investigated by	
362	articles included in literature review	25
363	Table 2.1. Complete overview of variables in the final dataset to train and test the five predictive	
364	models	38
365	Table 2.2. Model evaluation of five predictive models trained to predict high ($\geq 15\%$) morbidity	
366	cohorts of feedlot cattle.....	40
367	Table 3.1. Summary of exclusion criteria and the rows of data removed from the study. Rows	
368	were removed in order of the criteria listed. 300,104 animals were removed from the study.	
369		19
370	Table 3.2. Summary of Base model metrics and results.....	20
371	Table 4.1. Summary of calculations for estimation of cost and prices associated with model	
372	results.	40
373	Table 4.2. Summary of model performance predicting DNF_15 at first and second treatment for	
374	BRD.	41
375	Table 5.1. Table of descriptive statistics of selected variables split by metaphylaxis choice made	
376	by feedyard and economic model.	77
377	Table 5.2. Table showing all performance metrics for each model series and model run. Models	
378	included boosted decision tree, logistic regression, neural network, and random forest.....	77
379	Table 6.1. Characteristics of cattle at study enrollment by study location	93
380	Table 6.2. Characteristics of cattle at study enrollment based on treatment group and across the	
381	study population.....	93
382	Table 6.3. Outcome means ($\pm$ SEM) by treatment group from generalized linear mixed models	
383	with random intercept to account for lack of independence in cohorts within sites.	93
384		
385	Appendix Table A.1. Complete list of variables created or obtained to describe the data.....	94

386	Appendix Table B.1. Complete list of variables included in both the BRD1 and BRD2 datasets.	
387		98
388	Appendix Table B.2. Complete list of additional variables included in only the BRD2 dataset.	
389		101
390	Appendix Table B.3. Summary of the features selected for inclusion in the BRD1 model with	
391	associated chi square correlation coefficients.	104
392	Appendix Table B.4. Summary of the features selected for inclusion in the BRD2 model with	
393	associated chi square correlation coefficients.	105
394	Appendix Table C.1. Table describing the ranking of each variable for each predictive algorithm	
395	within each dataset	108
396		
397		

Acknowledgements

399 I want to first acknowledge the men of my doctoral committee. Without their guidance, I
400 would never have made it far into my graduate program. Brad, Bob, Phillip, Dustin, and Brian
401 have advised and encouraged me throughout vet school and my graduate program. They never
402 once doubted my abilities and have made me a strong and passionate researcher. I hope to one
403 day pay it forward and help young aspiring researchers as they have.

404 No less important are the graduate students that I have had the honor of calling my peers.
405 Kristen, Blaine, and Hector greeted and welcomed me to the BCI office on my first day as a
406 student worker and quickly became role models for my graduate program. Conrad, Kyndall, and
407 Paige helped me feel seen and understood as we slugged through our programs as fellow dual
408 degree students. Maddie, Rebecca, Jordana, Katie, Hayden, Liliana, Cole and Jake welcomed me
409 back to the BCI office after a full year away for clinics and eased me back into the graduate
410 student life. BCI grad students are truly a different breed, and I would not be the researcher I am
411 without my group of peers who I know all have incredibly bright futures ahead of them.

412 Eduarda and Luis have become fast friends and important role models for my career. I
413 hope to settle into my new role as a working professional as seamlessly as they have. The rest of
414 the BCI faculty deserve special acknowledgement for their contributions to my education and
415 this wonderful program as do the faculty at K-State CVM that have supported my learning
416 throughout my time at the college.

417 I also want to acknowledge the huge support system of friends and family working
418 behind the scenes. My parents and siblings who have supported my dreams from the beginning
419 and are excited to (finally) see me finish school. My extended family in Germany who have
420 supported me from across an ocean. My very best friends, McNeilly, Grace, and Sam, who have
421 listened to me blab about my research for years and have contributed the most to my sanity for
422 these last 7 years in college. The love of my life, Marc, who continues to be my light at the end
423 of every day, good, bad or otherwise. And lastly, my grandparents, who watched from above as I
424 completed my secondary education; ich hoffe, ich habe euch stolz gemacht!

426

Dedication

427 I would like to dedicate this work to my mama and papa who not only supported my
428 dreams in every way possible but have dedicated their lives to good science and have inspired
429 that love in all their children, most of all me. And to my siblings, who I like to think I am a good
430 role model for, I hope this rather large document makes you proud!

Chapter 1 - Disease prediction in beef cattle: a narrative review of the predictive model literature

L. Heinen, B.J. White, R.L. Larson

¹Beef Cattle Institute, Kansas State University, Manhattan, KS 66506, USA

Abstract

Predictive models use historical data to predict a future event and can be applied to a wide variety of tasks. Recently, the popularity of machine learning in livestock production has increased dramatically and resulted in an expansion of literature. Several reviews of predictive model literature have already addressed specific areas within cattle production including application of machine learning to dairy cattle health and performance as well as bovine respiratory disease research. A broader evaluation of the existing literature, specifically in the beef cattle sector, is required to better understand predictive model performance across various health challenges. Additionally, review of predictor variables and types of data utilized to train predictive models is lacking in the beef cattle literature. This narrative review aims to describe predictive model performance in greater detail across various disease outcomes, input data types, and algorithms with specific attention to accuracy, sensitivity, specificity, and positive and negative predictive values. A secondary goal is to address important areas for consideration for future work investigating the application of predictive models to beef cattle health challenges. Accuracy values ranged widely across disease outcome and algorithm categories. Predictive models demonstrated variability in sensitivity and specificity values. Negative predictive values were greater than positive predictive values for most disease outcomes. This review highlights the importance of utilizing several performance metrics and concludes that future work should address prevalence of outcomes and class-imbalanced data.

Keywords: machine learning, artificial intelligence, health, bovine

Introduction

As in any sector of the livestock industry, maintenance of animal health and well-being is a priority in the beef industry. The costliest disease affecting the beef cattle industry continues to be bovine respiratory disease (USDA APHIS National Animal Health Monitoring System, 2013). In addition, other conditions such as liver abscesses affect production and profit despite no outward clinical signs (Taylor et al., 2025). Several foreign animal diseases, such as foot-and-mouth disease, threaten the North American livestock industry as well (Knight-Jones and Rushton, 2013). National databases, on-farm record-keeping servers, and laboratory or clinical testing provide vast amounts of data that could be used for earlier detection of disease or outbreaks. Modern technology makes it easy to access, store, and utilize this data to its full capabilities.

Machine learning and predictive modeling offer powerful tools for the beef industry to aid in the detection of disease and determination of prognosis. Machine learning can incorporate large, complex datasets into algorithms used to predict outcomes and make estimations. By handling large amounts of data with ease, machine learning algorithms can utilize technology to complete tasks that would otherwise require human labor that could be best used elsewhere. Potential increased efficiency makes machine learning and predictive modeling attractive for the beef production sector as labor is a precious resource.

Several reviews have described the literature of predictive modeling in various cattle production systems (Graesboll et al., 2017; Ozella et al., 2023; Prosser et al., 2025; Slob et al., 2021; Wisnieski et al., 2019; Wisnieski et al., 2021). Reviews from the dairy sector summarize the capabilities of machine learning to assist in measuring performance outcomes as well as disease detection and health monitoring (Graesboll et al., 2017; Ozella et al., 2023; Slob et al., 2021; Wisnieski et al., 2019). Predictive modeling in the beef sector focuses largely on bovine respiratory disease (Prosser et al., 2025; Wisnieski et al., 2021). Other diseases or conditions affecting performance are under-represented in these literature reviews. The objective of this review is to report the performance of machine learning algorithms to predict health outcomes in beef cattle production systems while describing input data, algorithms, and hierarchy of the outcome variable to determine areas of consideration when utilizing predictive modeling in the beef sector.

Materials and Methods

For this review, the PubMed (National Library of Medicine, 2025), CAB Direct (CABI, 2025), and Agricola (National Agricultural Library, 2025) databases were used. These databases were selected to ensure adequate coverage of primary research in the beef sector of the cattle industry.

Initial search terms were selected based on the objective statement: cattle, dairy, beef, machine learning, and disease. Search terms were entered into the MeSH database from PubMed (National Library of Medicine, 2025) to add synonyms and preferred search words. This resulted in the following search string: (((((((cattle) OR (dairy)) OR (beef)) OR (feedlot)) OR (bovine)) AND (((predictive learning model) OR (machine learning)) OR (artificial intelligence)) OR (deep learning))) AND (((disease) OR (morbidity)) OR (mortality)). This search string was used to search the databases in the article title, abstract, and keywords fields in May of 2025.

Inclusion and exclusion criteria were applied to select articles relevant to the systematic literature review objective statement. The criteria are displayed in Table 1. Two authors (LH and BJW) evaluated article titles and abstracts based on these criteria. If a consensus could not be reached, the article was evaluated by a third author (RLL) to break the tie.

Following the review process, each selected article was read in full to extract relevant data. The data were recorded in an Excel spreadsheet and structured so each row represented a “study arm” from a selected article. A study arm serves as the smallest unit within an article for which performance metrics were reported. Study arms could represent different algorithms, different outcomes, or different data inputs. The columns represented data points extracted to evaluate the systematic review objective. Table 2 describes the columns in greater detail. Additional articles were removed at this step, all articles referencing only dairy cattle or calves were removed from consideration. If an article reported using data from both dairy and beef cattle, it was retained. Articles where a clear production system could not be identified were also removed. Furthermore, only articles utilizing classification algorithms were retained in order to compare predictive model performance across the same metrics; articles describing the use of regression algorithms were removed. Finally, model performance metrics of interest were accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. Articles were removed if these values were not reported or could not be calculated based on data provided in the article.

Predictive model performance metrics

All of the following measurements were calculated using counts of possible model outcomes: true positive, true negative, false positive, and false negative. A true positive is a unit (animal, group, or operation) that is identified as belonging to the positive class by the model and is truly positive. A true negative is a unit that is identified as belonging to the negative class by the model and is truly negative. A model may also make mistakes and misclassify a unit. A false positive is a unit that the model determines to belong to the positive class but is truly negative. A false negative is a unit that the model determines to belong to the negative class but is truly positive. Using the counts of these model outcomes, one can calculate various metrics for the evaluation of predictive model performance.

Accuracy measures how well the model determines truly positive and truly negative units in the total tested population. Accuracy is calculated by dividing the number of true positives and true negatives by the total population tested using the model. Sensitivity measures the ability of the model to find the truly positive units and is calculated by dividing the number of true positives by the total positive population (true positives and false negatives). Specificity measures the ability of the model to find the truly negative units. Specificity is calculated by dividing the number of true negatives by the total negative population (true negatives and false positives).

Positive predictive value (PPV) and negative predictive value (NPV) are additional measures of model performance and are affected by prevalence of the outcome of interest. PPV refers to the probability that a positive classification by the model correctly identifies a unit as positive and it is calculated by dividing the number of true positives by the total model positive population (true positives and false positives). NPV refers to the probability that a negative classification by the model correctly identifies a unit as negative. NPV is calculated by dividing the number of true negatives by the total model negative population (true negatives and false negatives). Using all these metrics to evaluate predictive model performance is important as the prevalence may affect some metrics more than others. Additionally, an imbalanced dataset can result in the model favoring the prediction of one outcome over the other resulting in models that can appear to have great accuracy but at the expense of other metrics such as sensitivity or specificity (Haixiang et al., 2017).

Results

The initial search led to a total of 1,061 articles across the three search databases (PubMed, CAB Direct, and Agricola). Following the removal of duplicates, 776 unique article titles and abstracts were available for review. Several articles ($n = 122$) were removed from review due to inability to access the full article text or due to full article text not being written in English. This resulted in 654 unique articles to be evaluated for final inclusion using the inclusion and exclusion criteria displayed in Table 1.

Following review of 654 abstracts, two authors (LH and BJW) agreed on inclusion of 118 articles in the final evaluation. The disagreements ($n = 53$) were reviewed by a third author (RLL) and resulted in the inclusion of 27 additional articles. This left a total of 145 articles for full review and data extraction. Removal of all articles referencing dairy cattle or those referencing an unknown production system resulted in removal of 117 and 4 articles, respectively. One article was removed due to describing the use of regression algorithms for prediction. Finally, four articles were removed due to inability to calculate all metrics of interest. The remaining 19 articles were left for data collection.

Of these 19 articles, 11 studies described data collection and model building as occurring only in beef animals. The remaining studies ($n = 8$) were conducted in both dairy and beef systems. Articles were published between 2014 and 2025 and published in 12 different journals. Health outcomes of interest covered 13 unique diseases that were broadly categorized in four different disease processes (Respiratory Disease, Foreign Animal Disease, Bovine Tuberculosis, and Other). The hierarchy of the outcome also varied between articles. An operational level outcome was defined as an outcome which would be determined at the level of a farm or feedlot. A group-level outcome was considered as an outcome determined at the level of a cohort of animals that had all been managed in a similar manner but were part of a larger operation. For example, a pen of cattle at a feedlot. An individual-level outcome was one in which the outcome was determined at the level of a single animal, regardless of that animal's participation in a group or farm. The majority of articles ($n = 11$) investigated individual-level outcomes. Six studies investigated operation-level outcomes. Only two studies investigated group-level outcomes.

The input data used to train the predictive models were diverse between the investigated articles. To simplify analysis, input data were broadly categorized: images and videos, climate

data, laboratory test results, and demographic data. Several studies included input data from multiple categories. Demographic data could further be categorized into individual-level, group-level, and operation-level data. Many studies utilized individual-level input data to predict an individual-level outcome ($n = 11$). Four of those studies also utilized operation-level and group-level data. Operation-level input data alone was used to predict operational-level outcomes in six articles. Only two studies used group-level data alone to predict group-level outcomes. Table 3 describes the input data used by the 19 articles by disease outcome category in greater detail.

Prevalence of health outcome varied between the studies. Table 4 reports the average prevalence of each health outcome as well as the number of articles and study arms contributing to the prevalence. The range of prevalence for each health outcome is also displayed for the health outcomes for which more than one prevalence was reported. Several performance metrics used to evaluate predictive models should be evaluated in the context of prevalence. Specifically, positive and negative predictive values (PPV and NPV) will vary depending on a low or high prevalence outcome.

Predictive model performance metrics were evaluated by disease outcome category and by algorithm category. The “Other” and “Respiratory Disease” outcome categories included all algorithm categories. The “Bovine Tuberculosis” disease outcome category only included linear and tree-based models. The “Foreign Animal Disease” category included few articles and only covered neural network and tree-based models. Figure 1 displays box-and-whisker plots of accuracy for the different disease outcome categories and algorithm categories. Points are used to indicate individual study arms. Accuracies were clustered together among the different algorithm categories in the “Respiratory Disease” outcome but varied in the “Other” category. Non-linear algorithms (neural networks, tree-based, and bayes-based) appeared to perform with higher accuracy than linear algorithms.

Figure 2 displays box-and-whisker plots of sensitivity and specificity for the different disease outcome categories and algorithm categories. Sensitivity varied greatly for all disease outcomes, particularly within the “Respiratory Disease” category. The linear algorithm category displayed the smallest variability in sensitivity among the algorithm categories. Specificity varied less and tended to be in the upper range (> 0.50) for all disease categories; however, linear models in the “Other” disease category displayed a wide range in specificity.

Figure 3 displays box-and-whisker plots of NPV and PPV for the different disease outcome categories and algorithm categories. NPVs tended to be high across different disease outcome categories and displayed little variation. Linear models built for the “Other” disease outcome category demonstrated a wide range of NPVs. Consequently, PPVs demonstrated lower values overall and displayed greater variability. Neural network models in the “Foreign Animal Disease” and “Other” categories displayed high PPVs.

Discussion

The search terms used in this literature review revealed hundreds of relevant articles. The application of machine learning to production agriculture has become popular in recent years, resulting in the publication of many articles relevant to livestock health conditions. The majority of the 145 articles originally selected for inclusion by the three authors described the application of predictive models to dairy cattle diseases. Data analysis within the dairy industry, especially in North America, benefits from vertical integration, meaning detailed data is often available for individual animals. Additionally, a variety of automatic sensors and computer systems have been adopted by the dairy industry in recent years. The vast amount of data generated by this technology lends itself well to machine learning techniques to detect health incidents or decreased performance of individual cows and contributes to the large amount of research on the application of machine learning the dairy cattle sector.

Comparatively few studies describe the application of machine learning to beef cattle health issues. Many articles included in this review focus efforts on bovine respiratory disease. As this is the costliest and most prevalent condition affecting the beef industry, specifically in the feedlot sector, this makes sense (USDA APHIS National Animal Health Monitoring System, 2013). Bovine tuberculosis is also well-represented in this review. This condition is also widespread and costly particularly in countries where eradication efforts are underway (Department for Environment, Food & Rural Affairs, 2014; Department for Environment, Food & Rural Affairs and Animal and Plant Health Agency, 2025). An additional reason why these two conditions are well-represented in this review could be due to the availability of the data used as inputs for the models predicting BRD and bovine tuberculosis outcomes. For example, the bovine tuberculosis articles utilized national databases on previous testing information collected by government agencies tasked with the tracking of tuberculosis outbreaks in cattle.

Many articles investigating the use of predictive models in BRD outcomes utilized treatment data from feedlots.

A variety of input data were utilized across the 19 articles included in this review. The most popular category was demographic data (Amrine et al., 2014; Feuz et al., 2021; Heinen et al., 2022; Heinen et al., 2023; Heinen et al., 2024; Punyapornwithaya et al., 2022; Rojas et al., 2022; Romero et al., 2020; Romero et al., 2021; Romero et al., 2022; Romero et al., 2023; Stanski et al., 2021; Unruh et al., 2017). This broad category featured three distinct hierarchies of data: individual animal-level, group-level, and operation-level. In the beef sector, animals are almost never marketed or managed as individuals except for the treatment of sick individuals. Group-level demographic data describes the characteristics of the group such as sex, average weight, and origin, among others. These characteristics not only describe the group but also inform various operation-specific management and marketing practices. For instance, weight and sex often determine a group of cattle's risk level for high morbidity at a feedyard. Operation-level demographic data can describe unique aspects of a single operation but may not be relevant if the predictive model is built for one operation only. The complicated structure of livestock production data lends itself to machine learning, especially non-linear algorithms which may be better able to model complex relationships between variables. However, special attention must be paid to ensure that the levels of data are used in a manner that does not result in overgeneralization of a model. This becomes an issue when models are trained on the data of one or multiple operations but applied to a different group of operations to which that data is not relevant.

Climate data was utilized by several studies (Heinen et al., 2023; Romero et al., 2020; Romero et al., 2021; Romero et al., 2023; Stanski et al., 2021; Unruh et al., 2017). Environmental factors likely contribute to the spread of disease and are significant in highly transmissible diseases such as foot and mouth disease (Gale et al., 2009). Additionally, the association of weather data and BRD morbidity and mortality has been investigated using traditional inferential statistics (Cernicchiaro et al., 2012; Cusack et al., 2007). One article in the current review directly evaluates the effect of inclusion of weather data on the performance of predictive models (Heinen et al., 2023). This article concluded that there was no benefit from the addition of weather data. Several studies included climate data but did not compare the performance of models built with and without it. These studies investigated bovine tuberculosis

outbreaks in England. In the context of this outcome, climate data was reported to have high feature importance indicating that the inclusion of climate data was important for predictive model performance (Romero et al., 2020; Romero et al., 2021; Romero et al., 2023). This finding could indicate that the inclusion of climate data should be considered for disease outcomes related to outbreak status or spread of disease. Additionally, climate data may only be important for diseases in which environmental spread is of particular importance.

Laboratory testing data were included in several articles as inputs for predictive models (Centeno-Delphia et al., 2025; Romero et al., 2020; Romero et al., 2021; Romero et al., 2022; Romero et al., 2023; Stanski et al., 2021; Wilson et al., 2024). These type of data are not as easy to obtain as the demographic and climate data. Climate data are easily obtained from many sources automatically and at a low cost through automatic programming interface (API) calls. Demographic data are often recorded on many levels at an operation for book-keeping purposes or to evaluate morbidity and mortality over time. Laboratory testing data are often not contemporaneous with other data. Time must be allowed for running a laboratory test and reporting of outcomes. This can introduce errors. Additionally, the application of a predictive model is often desired at the time an intervention can be implemented; thus, waiting for a test result is not always practical. Despite these limitations, laboratory testing may provide useful information and increase the accuracy of a predictive model depending on the test and predictive model application.

Several articles reported using images as input data for disease outcome predictions (Bortoluzzi et al., 2023; Gupta et al., 2023; Senthilkumar et al., 2024; Trindade et al., 2025; Unruh et al., 2017). Image classification with machine learning has become a powerful tool in both veterinary and human medicine (Marschner et al., 2017; Pesapane et al., 2018). Additionally, studies report using images for identification of individual animals (Li et al., 2022; Qiao et al., 2021). Images or video data can capture data in a form that cannot otherwise be conveyed in a table. This unique data source is also easy to collect with the right tools and, as technology improves, obtaining high quality images becomes less cost prohibitive. A limitation of this data source is the required storage space needed to utilize it in machine learning tasks. Additionally, models trained on image data may require longer processing times to provide predicted outcomes; thereby limiting their utilization to assist with management decisions in real-time.

Prevalence of an outcome, especially a disease-related outcome, should always be considered in the context of predictive model evaluation. Oftentimes, the outcome of interest may be a rare one, with low prevalence; however, the rare outcome is of interest to model due to significant economic or population health implications. The result is predictive models trained on imbalanced datasets. Evaluation of model performance on a single metric alone (such as accuracy) is not appropriate in such cases as this does not provide the full breadth of the model's performance (Haixiang et al., 2017). For example, a model trained on a dataset in which the prevalence of an outcome of interest is 5% may favor prediction of the negative because it results in a high accuracy (95%). If that model is evaluated on accuracy alone, it appears to be a great model. However, upon evaluation of sensitivity (0%), it becomes clear that the model is predicting every row to be negative, which is not useful if the outcome of interest is presence or absence of disease.

This imbalance of the outcome of interest is apparent in the current review. Prevalences for disease outcomes in this study ranged from 6.1 to 86.0%. The application of predictive models to disease outcomes for which the positive class is rare, may result in a tendency of the predictive model to favor selection of the majority class to improve accuracy (Loyola-Gonzalez et al., 2016). In these cases, the predictive model may boast high accuracy but fails to detect any of the truly positive cases resulting in a very low sensitivity. In applications in which disease detection is important, such a predictive model becomes useless. Several articles report high prevalences which may have resulted in the opposite phenomenon, high sensitivity values while sacrificing specificity (Romero et al., 2020; Romero et al., 2021; Romero et al., 2022; Romero et al., 2023; Trindade et al., 2025).

An additional consequence of utilizing imbalanced data is captured by the trends seen in figures 2 and 3. Both NPV and PPV are metrics influenced by the sensitivity and specificity of the model as well as the prevalence of the outcome of interest. Oftentimes, the prevalence of disease-positive individuals tends to be lower than those that are disease-negative (Wisnieski et al., 2021). Very high NPVs and low PPVs demonstrate that models predicting a negative outcome had a higher probability of being correct and models predicting a positive outcome had a lower probability of being correct. Instances in which prevalence of disease positive individuals is higher would result in decreasing NPVs and increasing PPVs regardless of

sensitivity and specificity – which are metrics that remain the same regardless of prevalence of outcome of interest.

Class-imbalanced data are perhaps the biggest challenge in the application of predictive models to determine health outcomes in beef cattle. There exist several strategies to cope with this challenge. Several machine learning approaches have been developed specifically to address class-imbalance such as development of ensemble methods such as random forest or boosting (Galar et al., 2013; Krawczyk et al., 2016). Additionally, balancing of the outcome classes prior to testing using over- or undersampling techniques as also been described. Two articles included in the current review reported on the differences in model performance when utilizing a balancing technique in comparison to the baseline prevalence. The results demonstrated no improvement of predictive model performance metrics (Amrine et al., 2014; Heinen et al., 2023). Other methods such as hybrid over- and under-sampling, cluster-based, and distance-based methods also exist and have demonstrated promise in other fields (Haixiang et al., 2017). Efforts should be made to address class imbalance either through application of an ensemble method or a balancing technique in the application of predictive modeling to health outcome determination.

There has been increased interest in the application of machine learning to improve health and productivity in cattle. The dairy sector of the industry remains ahead of the beef sector in applying predictive model technology. As the beef sector, in addition to the remainder of the livestock production industry, moves forward with machine learning integration, there are several important considerations revealed by the current review. First, selection of input data should balance utility in the model, ease of collection, and quality. Contribution to predictive model performance should be measured if possible and evaluated in the context of cost of collection. Efforts should be made to remove data which could provide misleading information. Second, evaluation of predictive model performance should include several variables and consider the prevalence of the outcome of interest in order to avoid development of predictive models that favor a single outcome class. Finally, the challenges of class-imbalanced data should be addressed with algorithm selection and/or use of a balancing technique at training.

References

Amrine, D.E., White, B.J., Larson, R.L., 2014. Comparison of classification algorithms to predict outcome of feedlot cattle identified and treated for bovine respiratory disease.

Computers and Electronics in Agriculture 105, 9–19.
<https://doi.org/http://dx.doi.org/10.1016/j.compag.2014.04.009>

Bortoluzzi, E.M., Schmidt, P.H., Brown, R.E., Jensen, M., Mancke, M.R., Larson, R.L.,
 Lancaster, P.A., White, B.J., 2023. Image classification and automated machine learning
 to classify lung pathologies in deceased feedlot cattle. *Veterinary Sciences* 10, 113–124.
<https://doi.org/https://doi.org/10.3390/vetsci10020113>

CABI, 2025. CABI Digital Library [WWW Document]. URL <https://www.cabidigitallibrary.org/>

Centeno-Delphia, R.E., Glidden, N., Long, E., Ellis, A., Hoffman, S., Mosier, K., Ulloa, N.,
 Cheng, J.J., Davidson, J.L., Mohan, S., Kamel, M., Szasz, J.I., Schoonmaker, J., Koziol,
 J., Boerman, J.P., Ault, A., Verma, M.S., Johnson, T.A., 2025. Nasal pathobiont
 abundance is a moderate feedlot-dependent indicator of bovine respiratory disease in beef
 cattle. *Animal Microbiome* 7, 27–43. <https://doi.org/https://doi.org/10.1186/s42523-025-00387-y>.

Cernicchiaro, N., Renter, D., White, B., Babcock, A., Fox, J., 2012. Associations between
 weather conditions during the first 45 days after feedlot arrival and daily respiratory
 disease risks in autumn-placed feeder cattle in the United States. *Journal of Animal
 Science* 90, 1328–1337. <https://doi.org/10.2527/jas.2011-4657>.

Cusack, P., McMeniman, N., Lean, I., 2007. Feedlot entry characteristics and climate: their
 relationship with cattle growth rate, bovine respiratory disease and mortality. *Australian
 Veterinary Journal* 85, 311–316. <https://doi.org/10.1111/j.1751-0813.2007.00184.x>.

Department for Environment, Food & Rural Affairs, 2014. A strategy for achieving Officially
 Bovine Tuberculosis Free status in England [WWW Document]. GOV.UK. URL
<https://www.gov.uk/government/publications/a-strategy-for-achieving-officially-bovine-tuberculosis-free-status-for-england>

Department for Environment, Food & Rural Affairs, Animal and Plant Health Agency, 2025.
 Latest accredited official statistics on tuberculosis (TB) in cattle in Great Britain -
 quarterly [WWW Document]. GOV.UK. URL
<https://www.gov.uk/government/statistics/incidence-of-tuberculosis-tb-in-cattle-in-great-britain>

791 Feuz, R., Feuz, K., Johnson, M., 2021. Improving feedlot profitability using operational data in
792 mortality prediction modeling. *Journal of Agriculture and Resource Economics* 46, 242–
793 255. <https://doi.org/10.22004/ag.econ.304772>

794 Galar, M., Fernandez, A., Barrenechea, E., Herrera, F., 2013. EUSBoost: enhancing ensembles
795 for highly imbalanced data-sets by evolutionary undersampling. *Pattern Recognition* 46,
796 3460–3471. <https://doi.org/https://doi.org/10.1016/j.patcog.2013.05.006>

797 Gale, P., Drew, T., Phipps, L.P., David, G., Wooldridge, M., 2009. The effect of climate change
798 on the occurrence and prevalence of livestock diseases in Great Britain: a review. *Journal*
799 *of Applied Microbiology* 106, 1409–1423. [https://doi.org/https://doi.org/10.1111/j.1365-](https://doi.org/https://doi.org/10.1111/j.1365-2672.2008.04036.x)
800 [2672.2008.04036.x](https://doi.org/https://doi.org/10.1111/j.1365-2672.2008.04036.x)

801 Graesboll, K., Kirkeby, C., Nielson, S.S., Halasa, T., Toft, N., Christiansen, L.E., 2017. A robust
802 statistical model to predict the future value of the milk production of dairy cows using
803 herd recording data. *Frontiers in Veterinary Science* 4.
804 <https://doi.org/https://doi.org/10.3389/fvets.2017.00013>

805 Gupta, S., Keuhn, L.A., Clawson, M.L., 2023. Early detection of infectious bovine
806 keratoconjunctivitis with artificial intelligence. *Veterinary Research* 54, 122–126.
807 <https://doi.org/https://doi.org/10.1186/s13567-023-01255-w>

808 Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G., 2017. Learning from
809 class-imbalanced data: review of methods and applications. *Expert Syst. Appl.* 73, 220–
810 239.

811 Heinen, L., Lancaster, P.A., White, B.J., Zwiefel, E., 2022. Evaluation of predictive models to
812 determine total morbidity outcome of feedlot cattle based on cohort-level feed delivery
813 data during the first 15 days on feed. *Translational Animal Science* 6, 1–5.
814 <https://doi.org/https://doi.org/10.1093/tas/txac121>

815 Heinen, L., White, B.J., Amrine, D.E., Larson, R.L., 2023. Evaluation of predictive models to
816 determine final outcome for feedlot cattle based on information available at first
817 treatment for bovine respiratory disease. *American Journal of Veterinary Research*.
818 <https://doi.org/doi.org/10.2460/ajvr.23.05.0094>

819 Heinen, L., White, B.J., Larson, R.L., Kopp, D., Pendell, D.L., 2024. Economic impact of
820 mortality prediction by predictive model at first and second treatment for bovine

respiratory disease. American Journal of Veterinary Research.
<https://doi.org/doi.org/10.2460/ajvr.24.06.0169>

Knight-Jones, T.J.D., Rushton, J., 2013. The economic impacts of foot and mouth disease - what are they, how big are they, and where do they occur? Preventive Veterinary Medicine 112, 161–173. <https://doi.org/https://doi.org/10.1016/j.prevetmed.2013.07.013>

Krawczyk, B., Wolzniak, M., Schaefer, G., 2016. Cost-sensitive decision tree ensembles for effective imbalanced classification. Applied Soft Computing 14, 554–562.
<https://doi.org/https://doi.org/10.1016/j.asoc.2013.08.014>

Li, G., Erickson, G.E., Xiong, Y., 2022. Individual beef cattle identification using muzzle images and deep learning techniques. Animals 12.
<https://doi.org/https://doi.org/10.3390/ani12111453>

Loyola-Gonzalez, O., Martinez-Trinidad, J.F., Carrasco-Ochoa, J.A., Garcia-Borroto, M., 2016. Study of the impact of resampling methods for contrast pattern based classifiers in imbalanced databases. Neurocomputing 175, 935–947.
<https://doi.org/https://doi.org/10.1016/j.neucom.2015.04.120>

Marschner, C.B., Kokla, M., Amigo, J.M., Rozanski, E.A., Wiinberg, B., McEvoy, F.J., 2017. Texture analysis of pulmonary parenchymateous changes related to pulmonary thromboembolism in dogs - a novel approach using quatitative methods. BMC Veterinary Research 13. <https://doi.org/https://doi.org/10.1186/s12917-017-1117-1>

National Agricultural Library, 2025. AGRICOLA [WWW Document]. URL
https://search.nal.usda.gov/discovery/search?vid=01NAL_INST:MAIN

National Library of Medicine, 2025. PubMed [WWW Document]. URL
<https://pubmed.ncbi.nlm.nih.gov/>

Ozella, L., Rebulli, K.B., Forte, C., Giacobini, M., 2023. A literature review of modeling approaches applied to data collected in automatic milking systems. Animals 13.
<https://doi.org/https://doi.org/10.3390/ani13121916>

Pesapane, F., Codari, M., Sardanelli, F., 2018. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. European Radiology Experimental 2. <https://doi.org/https://doi.org/10.1186/s41747-018-0061-6>

Prosser, H.M., Bortoluzzi, E.M., Valeris-Chacin, R.J., Baker, E.C., Scott, M.A., 2025. Application of artificial intelligence and machine learning in bovine respiratory disease

prevention, diagnosis, and classification. American Journal of Veterinary Research 86, S22–S26. <https://doi.org/https://doi.org/10.2460/ajvr.24.10.0327>

Punyapornwithaya, V., Klaharn, K., Arjkumpa, O., Sansamur, C., 2022. Exploring the predictive capability of machine learning models in identifying foot and mouth disease outbreak occurrences in cattle farms in an endemic setting of Thailand. Preventive Veterinary Medicine 207. <https://doi.org/https://doi.org/10.1016/j.prevetmed.2022.105706>

Qiao, Y., Clark, C., Lomax, S., Kong, H., Su, D., Sukkarieh, S., 2021. Automated individual cattle identification using video data: a unified deep learning architecture approach. Frontiers in Animal Science 22. <https://doi.org/https://doi.org/10.3389/fanim.2021.759147>

Rojas, H.A., White, B.J., Amrine, D.E., Larson, R.L., 2022. Predicting bovine respiratory disease risk in feedlot cattle in the first 45 days post arrival. Pathogens 11, 442–456. <https://doi.org/https://doi.org/10.3390/pathogens11040442>

Romero, M.P., Chang, Y., Brunton, L.A., Parry, J., Prosser, A., Upton, P., Rees, E., Tearne, O., Arnold, M., Stevens, K., Drewe, J.A., 2020. Decision tree machine learning applied to bovine tuberculosis risk factors to aid disease control decision making. Preventive Veterinary Medicine 175. <https://doi.org/https://doi.org/10.1016/j.prevetmed.2019.104860>

Romero, M.P., Chang, Y., Brunton, L.A., Prosser, A., Upton, P., Rees, E., Tearne, O., Arnold, M., Stevens, K., Drewe, J.A., 2021. A comparison of the value of two machine learning predictive models to support bovine tuberculosis disease control in England. Preventive Veterinary Medicine 188. <https://doi.org/https://doi.org/10.1016/j.prevetmed.2021.105264>

Romero, M.P., Chang, Y., Brunton, L.A., Parry, J., Prosser, A., Upton, P., Drewe, J.A., 2022. Machine learning classification methods informing the management of inconclusive reactors at bovine tuberculosis surveillance tests in England. Preventive Veterinary Medicine 199. <https://doi.org/https://doi.org/10.1016/j.prevetmed.2021.105565>

Romero, M.P., Chang, Y., Brunton, L.A., Parry, J., Prosser, A., Upton, P., Drewe, J.A., 2023. Assessing the potential impact of applying a higher sensitivity test to selected cattle populations for the control of bovine tuberculosis in England. Preventive Veterinary Medicine 219. <https://doi.org/https://doi.org/10.1016/j.prevetmed.2023.106004>

Senthilkumar, C., C, S., Vadivu, G., Neethirajan, S., 2024. Early detection of lumpy skin disease in cattle using deep learning - a comparative analysis of pretrained models. *Veterinary Sciences* 11, 510–531. <https://doi.org/https://doi.org/10.3390/vetsci11100510>

Slob, N., Catal, C., Kassahun, A., 2021. Application of machine learning to improve dairy farm management: a systematic literature review. *Preventive Veterinary Medicine* 187. <https://doi.org/https://doi.org/10.1016/j.prevetmed.2020.105237>

Stanski, K., Lycett, S., Porphyre, T., de C. Bronsvort, B.M., 2021. Using machine learning improves predictions of herd-level bovine tuberculosis breakdowns in Great Britain. *Scientific Reports* 11.

Taylor, D.D., Groenendaal, H., Morley, P.S., Zagmutt, F., 2025. Economic costs of liver abscesses in U.S. beef feedlot cattle: a comprehensive economic analysis. *Journal of Animal Science*. <https://doi.org/https://doi.org/10.1093/jas/skaf127>

Trindade, A.S.N., de Moraes, I.M.L., Leal, L.L.L.L., Salvador, V.F., Heller, L.M., Zapa, D.M.B., de Aquino, L.M., Ferreira, L.L., Lopes, W.D.Z., 2025. TFinder App: artificial intelligence to diagnose tick fever agents and assess parasitemia/bacteremia in bovine blood smears. *Veterinary Parasitology* 334. <https://doi.org/https://doi.org/10.1016/j.vetpar.2025.110415>

Unruh, E.M., Theurer, M.E., White, B.J., Larson, R.L., Drouillard, J.S., Schrag, N., 2017. Evaluation of infrared thermography as a diagnostic tool to predict heat stress events in feedlot cattle. *American Journal of Veterinary Research* 78, 771–777.

USDA APHIS National Animal Health Monitoring System, 2013. Feedlot 2011 Part IV: Health and Health Management on U.S. Feedlots with a Capacity of 1,000 or More Head.

Wilson, R.A., Johnson, B.J., Sarturi, J.O., Crossland, W.L., Hales, K.E., Rathmann, R.J., Bratcher, C.L., Theurer, M.E., Amachawadi, R.G., Nagaraja, T.G., Speidel, S.E., Enns, R.M., Thomas, M.G., Foraker, B.A., Cleveland, M.A., Woerner, D.R., 2024. Identification of blood-based biomarkers for detection of liver abscesses in beef x dairy heifers. *Applied Animal Science* 40, 386–397. <https://doi.org/https://doi.org/10.15232/aas.2023-02504>

Wisnieski, L., Norby, B., Pierce, S.J., Becker, T., Sordillo, L.M., 2019. Prospects for predictive modeling of transition cow diseases. *Animal Health Research Reviews* 20, 19–30. <https://doi.org/10.1017/S1466252319000112>

914 Wisnieski, L., Amrine, D.E., Renter, D.G., 2021. Predictive modeling of bovine respiratory
915 disease outcomes in feedlot cattle: a narrative review. *Livestock Science* 251.
916 <https://doi.org/https://doi.org/10.1016/j.livsci.2021.104666>
917

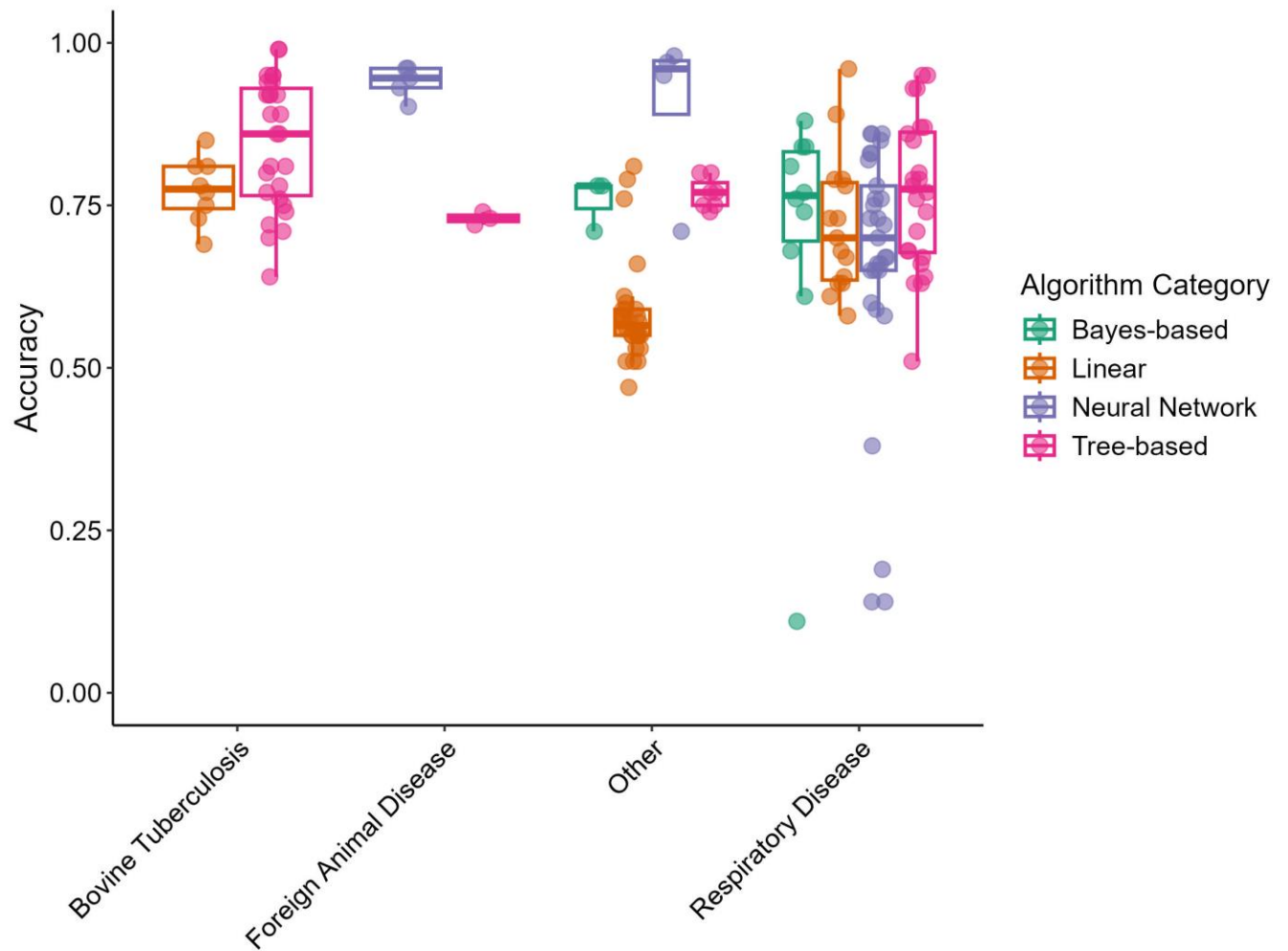


Figure 1.1. Box-and-whisker plots of predictive model accuracy by disease outcome and algorithm category extracted from reviewed articles. Points represent individual study arms.

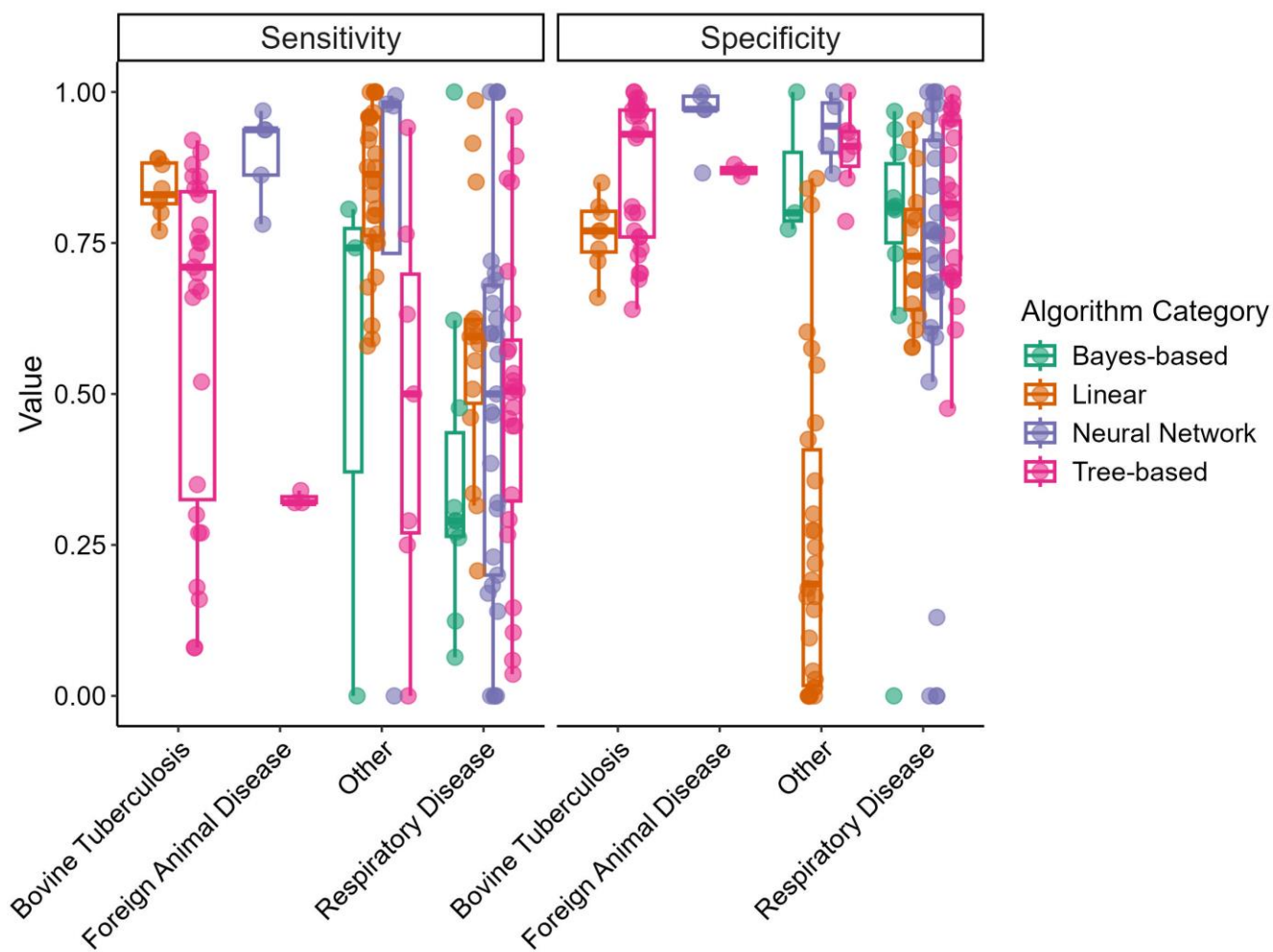


Figure 1.2. Box-and-whisker plots of predictive model sensitivity and specificity by disease outcome and algorithm category extracted from reviewed articles. Points represent independent study arms.

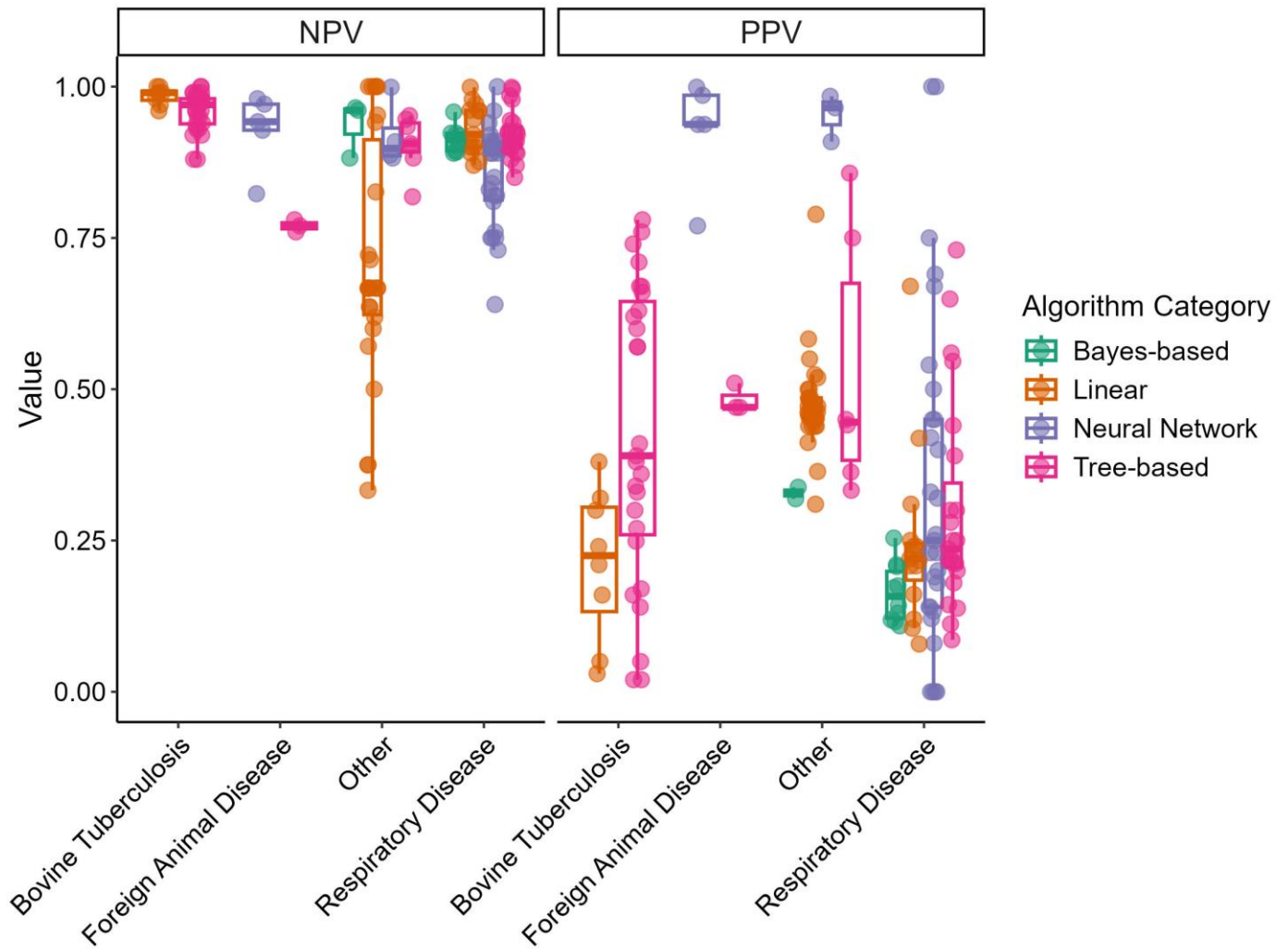


Figure 1.3. Box-and-whisker plots of predictive model negative and positive predictive values by disease outcome and algorithm category extracted from reviewed articles. Points represent independent study arms.

933 **Table 1.1. Inclusion and exclusion criteria applied to articles for literature review selection**

Inclusion	Exclusion
Primary research article	Narrative or systematic literature review, meta-analysis
Data used from dairy or beef cattle	Data used from small ruminants, poultry, companion animals, rodents, reptiles, equine, or other atypical bovids (bison, water buffalo, yaks, etc.)
Data must be at the individual animal-, group/herd, or operational level	Simulated data
Outcomes of interest are related to a specific disease or condition or overall morbidity or mortality Pregnancy loss considered a disease	Outcomes of interest are related to production performance such as carcass characteristics, milk yield, pregnancy diagnosis, weight gain, etc.
Machine learning/predictive modeling is utilized to anticipate the outcome of interest	Inferential statistics are used to determine an association between the outcome and predictor variables

934

935

936

937

938 **Table 1.2. Description of data collected from articles included in narrative literature**
939 **review**

Variable Name	Description
Title	Title of the research article as published
Authors	Last name and initials of all authors
Year	Year of publication
Journal	Title of journal in which article was published
Production_System	Broad category of animal type/use referred to in the article; beef, dairy, or both
Study_Arm	Letter to indicate study arm of the article
Outcome_Category	Word or phrase to describe the broad disease category being investigated in the article; example: lameness, respiratory disease
Image_yn	Describes if images or video data were used to train the models
Climate_Data_yn	Described if weather data were used to train the models. Included data such as temperature, humidity, wind speed, etc.
TestResults_yn	Described if laboratory test results were used to train the models. Included data such as complete blood count results, intradermal tuberculosis testing, etc.
Demographic_Data_yn	Described if demographic characteristics were used to train the models. Further split into three different hierarchies: operation, group, and individual. Included data such as number of animals on an operation, average weight of a group, rectal temperature of an individual, etc.
Number_EU	Number of experimental units in the training dataset given to the model
EU_Type	Type of experimental unit to describe data hierarchy; example: farm, individual animal
Algorithm	Machine learning or predictive model used in the study arm, as a broad category; example: tree-based, linear, etc.
Accuracy, %	Measure of number of correctly classified units over total classified units

Sensitivity, %	Proportion of the truly positive individuals (diseased) that were identified as positive by the model
Specificity, %	Proportion of the truly negative individuals (disease free) that were identified as negative by the model
Positive predictive value (PPV)	Proportion of the model-identified positive individuals that were truly positive
Negative predictive value (NPV)	Proportion of the model-identified negative individuals that were truly negative

940
941

942 **Table 1.3. Distribution of articles and study arms from literature review by input data**
943 **type and disease outcome category**

Input Data Type	Disease Outcome Category	Count of Articles	Count of Study Arms
<i>Image or Video</i>	Other	3	14
	Respiratory Disease	1	16
	Foreign Animal Disease	1	5
<i>Climate Data</i>	Other	1	11
	Respiratory Disease	1	5
	Bovine Tuberculosis	4	32
<i>Laboratory Testing Data</i>	Other	1	30
	Respiratory Disease	1	1
	Bovine Tuberculosis	5	35
<i>Operational Demographic Data</i>	Respiratory Disease	4	51
	Foreign Animal Disease	1	3
	Bovine Tuberculosis	5	35
<i>Group-Level Demographic Data</i>	Other	1	5
	Respiratory Disease	5	55
<i>Individual Animal-Level Demographic Data</i>	Other	1	11
	Respiratory Disease	4	50

Table 1.4. Description of health outcomes and prevalence of health outcomes investigated by articles included in literature review

Disease Category	Health Outcome	Count of Articles	Count of Study Arms	Average Prevalence, % (range)*
<i>Respiratory Disease</i>	Acute interstitial pneumonia	1	4	11.8 (10.7-12.9)
	Broncho-interstitial pneumonia	1	4	38.5 (35.4-40.9)
	Bovine respiratory disease	7	62	20.1 (6.1-51.9)
	Other pneumonia	1	4	13.7 (11.4-16.4)
<i>Foreign Animal Disease</i>	Foot-and-mouth disease	1	3	27.0
	Lumpy skin disease	1	5	38.6 (31.6-49.1)
<i>Bovine Tuberculosis</i>	Bovine tuberculosis	5	35	13.6 (3.58-86.0)
<i>Other</i>	Heat stress	1	11	11.8
	Infectious bovine keratoconjunctivitis	1	1	19.5
	Liver abscess	1	30	44.8
	Total morbidity	1	5	18.5
	Anaplasmosis	1	1	79.1
	Babesia	1	1	85

*Ranges only available for health outcomes with more than one entry for prevalence

Chapter 2 - Evaluation of predictive models to determine total morbidity outcome of feedlot cattle based on pen-level feed delivery data during the first 15 days on feed

L. Heinen¹, P.A. Lancaster¹, B.J. White^{1*}, and E. Zwiefel²

¹Beef Cattle Institute, Kansas State University, Manhattan, KS 66506, USA

²Machine Learning Global Black Belt Team, Microsoft Corporation, Edine, MN 55424, USA

*Corresponding author: Brad J. White

This research is published in the Journal of Translational Animal Science and is reproduced with permission from publisher as formatted in referenced publication. Lilli Heinen, Phillip A. Lancaster, Brad J. White, Eric Zwiefel. Evaluation of predictive models to determine total morbidity outcome of feedlot cattle based on pen-level feed delivery data during the first 15 days on feed, Translational Animal Science, Volume 6, Issue 3, August 2022, txac121, <https://doi.org/10.1093/tas/txac121>

Abstract

Changes in feeding behavior and intake have been used to predict the onset of bovine respiratory disease in individual animals but have not been applied to cohort-level data. Correctly identifying high morbidity cohorts of cattle early in the feeding period could facilitate the administration of interventions to improve health and economic outcomes. The study objective was to determine the ability of feed delivery data from the first 15 days of feed to predict total feeding period morbidity. Data consisted of 518 cohorts (10 feedlots, 56,796 animals) of cattle of varying sex, age, arrival weight, and arrival time of year over a 2-year period. Overall cohort-level morbidity was classified into high ($\geq 15\%$ total morbidity) or low categories with 18.5% of cohorts having high morbidity. Five predictive models (advanced perceptron, decision forest, logistic regression, neural network, and boosted decision tree) were created to predict overall morbidity given cattle characteristics at arrival and feeding characteristics from the first 15 days. The dataset was split into training and testing subsets (75% and 25% of original, respectively),

stratified by the outcome of interest. Predictive models were generated in Microsoft Azure using the training set and overall predictive performance was evaluated using the testing set. Performance in the testing set ($n = 130$) was measured based on final accuracy, sensitivity (Sn, the ability to accurately detect high morbidity cohorts), and specificity (Sp, the ability to accurately detect low morbidity cohorts). The decision forest had the highest Sp (97%) with the greatest ability to accurately identify low morbidity lots (103 of 106 identified correctly), but this model had low Sn (33%). The logistic regression and neural network had similar Sn (both 63%) and Sp (69% and 72%, respectively) with the best ability to correctly identify high morbidity cohorts (15 of 24 correctly identified). Predictor variables with the greatest importance in the predictive models included percent change in feed delivery between days and 4-day moving averages. The most frequent variable with a high level of importance among models was the percent change in feed delivered from d 2 to 3 after arrival. In conclusion, feed delivery data during the first 15 days on feed was a significant predictor of total cohort-level morbidity over the entire feeding period with changes in feed delivery providing important information.

Keywords: feeding patterns, machine learning, predictive analytics

Introduction

Bovine respiratory disease (BRD) is one of the costliest diseases in the feedlot industry (Salman et al., 1991). Early identification and prompt treatment of the disease can reduce the costs associated with BRD (Booker et al., 2004). Additionally, gastrointestinal diseases, lameness, and other adverse health events negatively impact animal performance, but the prevalence of these other diseases is relatively low. Machine learning has been applied to various fields in the agricultural industry to decrease costs and increase outputs. Predictive modeling techniques have not been heavily studied for use in the beef industry (White et al., 2018). Feedlots collect large volumes of data daily, including feed calls, antimicrobials, and other treatments administered, and various arrival characteristics. These data and more are potential inputs for predictive models that could be used to predict the health outcome of cohorts of cattle. Therefore, the study objective was to evaluate the diagnostic ability of predictive models to determine whether a cohort would have a health outcome of $\geq 15\%$ total morbidity during the feeding period based on data collected at arrival and feed delivery data for the first 15 days of feed. The hypothesis was that predictive models can accurately predict total feeding period morbidity, and that feed delivery patterns during the first 15 days on feed are important

predictors. The objective was to determine the accuracy, sensitivity, and specificity of five predictive models using arrival characteristics and feed delivery data during the first 15 days on feed to correctly identify high ($\geq 15\%$) morbidity cohorts.

Materials and Methods

Animal Care and Use Committee approval was unnecessary as data were obtained from an existing database of feedlot operational data.

Data Collection

Daily records for 12,657 cohorts of cattle (1,005,320 animals) were obtained from 10 U.S. feedlots spanning 2018 to 2020. Health event records on an individual animal basis were tied to cohort-level data. A cohort was defined as a group of cattle purchased and managed together but not necessarily housed in the same pen for the entirety of the feeding phase.

Data Transformation

Data were transformed into the appropriate format and new variables generated before use in predictive models. First, inclusion criteria were applied to the data. Cohorts needed to have complete arrival and feeding data. Feed delivery data beyond 15 days on feed were not included in the dataset, and remaining feed delivery data were adjusted for arrival body weight (percentage of arrival body weight) on a dry matter basis. Any cohorts with missing data were removed. This resulted in removal of many rows of data as several cohorts were rearranged early in the feeding period. Cohorts with extreme values for some arrival characteristics or dry matter intake were also removed to reduce data entry errors. Cohorts were removed if the cohort was designated as a hospital pen, the average arrival weight was less than 182 kg or greater than 545 kg, days on feed was less than 0 or greater than 300, the cohort had not yet been closed out, or dry matter intake as a percentage of average arrival body weight was greater than 5% or less than 0.1% on any given day within the first 15 days on feed. After applying these exclusion criteria, 12,139 cohorts were removed from the dataset. Secondly, data wrangling techniques applied in R software (R Core Team, 2021) were used to create a dataset that consisted of one row per cohort.

Several additional variables were created using the existing data to describe feed delivery characteristics such as day-to-day changes in feed delivery and rolling averages of feed delivery using 2-to-7-day time spans. Figure 2.1 outlines the feeding variables and which days on feed were accounted for in each feeding characteristic variable. Although each variable indicated in

the figure is noted only once, feeding variables were calculated for their respective increments throughout the 15-day feeding period of interest. The outcome of interest was captured in a variable called total morbidity category. This variable described the total morbidity (all diseases) of a cohort of cattle during the entire feeding period. It was expressed as a categorical variable in which high morbidity indicated a total morbidity percentage of greater than or equal to 15%. Low morbidity was a total morbidity percentage of less than 15%. Table 2.1 offers a complete overview of all the variables that populated the dataset. Following data transformation and variable creation, the final dataset consisted of 518 cohorts (56,796 animals) with 18.5% of cohorts having high morbidity.

Before the model building step occurred, data were split into separate training and testing sets, stratified by the prevalence of outcome of interest. The training set represented 75% of the original dataset while the testing set represented the other 25%. Using the Pipeline Designer function in Azure Machine Learning Studio (Microsoft, 2022), the datasets were used to create five predictive models. The models trained and tested in this study were advanced perceptron, neural network, boosted decision tree, decision forest, and logistic regression. A previous description has been given by Rojas et al. (2022) from our laboratory. Briefly, neural networks and advanced perceptron models can be useful for identifying patterns in operational data, but are difficult to utilize for describing model structure or importance of predictive variables (Rosenblatt, 1957; Zhang, 2010). Boosted decision trees and decision forests are classification models creating a series of splits in data based on attributes to minimize entropy in resulting data subsets (Breiman, 2001; Roe et al., 2005). Logistic regression models are often used for statistical analysis and these algorithms can be created, then used to estimate a probability of even occurrence in a predictive manner (Dreiseitl and Ohno-Machado, 2002). The models selected for investigation in this study are based on previous work conducted by Rojas et al. (2022) and Amrine et al. (2014), as well as the resources available on Azure.

Model Evaluation

Finally, test data were used to evaluate model performance. Adjustment of the threshold probability was done manually for each model to maximize F1 score to balance sensitivity and positive predictive value. The metrics used to evaluate the models were accuracy, sensitivity (Sn), specificity (Sp), positive and negative predictive values (PPV and NPV, respectively), and area under the receiver operating characteristics (ROC) curve (AUC). These values were

calculated using the confusion matrices produced by each model run in Azure. Using these metrics, we were able to compare models based on their ability to accurately predict the total feeding period morbidity of a cohort. Figure 2.2 demonstrates the training and testing process of the predictive models.

Results and Discussion

The ROC curves for the five predictive models are presented in Figure 2.3. A line closer to a true positive rate of 1.0 and a false positive rate of 0.0 (i.e., top left corner) has greater sensitivity (correctly identify true positives—high morbidity cohorts) and specificity (correctly identifying true negatives—low-morbidity cohorts). Logistic regression, neural network, and decision forest models have similar ROC curves, whereas the decision tree model appears to be slightly less predictive, and the advanced perceptron model has very little predictive ability. Additionally, the overall accuracy, in addition to the AUC value, support the conclusion that the advanced perceptron model has very poor performance (Table 2.2). The advanced perceptron model identified all cohorts as high morbidity based on the sensitivity of 100.0%. The logistic regression and neural network models had similar overall accuracy, as well as similar sensitivity, specificity, positive and negative predictive values. The decision tree and decision forest models had the greatest overall accuracy and specificity indicating good ability to identify low morbidity cohorts but had low sensitivity.

Previous work from our lab (Amrine et al., 2019) reported that predictive models with greater specificity had lesser sensitivity when using sale barn and arrival characteristics to predict BRD in the first 14 days on feed. The “best” model depends upon the goals of the feedlot manager. If the goal is to accurately identify low morbidity cohorts, then a model with the greatest specificity would be deemed the ‘best’ model. Conversely, if the goal is to accurately identify high morbidity cohorts, then a model with the greatest sensitivity would be deemed the ‘best’ model. Accurate identification of high morbidity cohorts could lead to actions that allow changes in risk management, frequency of disease monitoring, and other interventions. However, if the model misclassifies low morbidity cohorts as high morbidity cohorts, then valuable resources would be wasted. The decision forest model had a low sensitivity, only accurately classifying 33% of high morbidity cohorts, but based on the high PPV, this model had a high probability of being correct when it did classify a cohort as high morbidity. Thus, the decision

forest model would not identify all of the high morbidity cohorts, but if potential interventions were implemented based on this model little resources would be wasted.

Several previous studies (Sowell et al., 1999; Buhman et al., 2000; Quimby et al., 2001; Moya et al., 2015; Wolfger et al., 2015; Jackson et al., 2016) have indicated that feed intake, feeding behavior, and drinking behavior are predictive of onset of BRD in individual cattle. The onset of BRD can be predicted 4 to 7 days before clinical signs can be observed using feeding and drinking data (Buhman et al., 2000; Moya et al., 2015; Wolfger et al., 2015; Jackson et al., 2016). Feeding behavior has predicted BRD with an overall accuracy of 84% to 89% and positive predictive value of 85% to 96% (Quimby et al., 2001). Wolfger et al. (2015) indicated that feed behavior correctly predicted 81% of BRD cases and 77% of healthy animals 3 days prior to clinical signs; adding feed intake did not improve these predictions. Similarly, Moya et al. (2015), using pattern recognition techniques, reported that feeding behavior predicted BRD with good sensitivity (58% to 83%) and specificity (67% to 100%) for some of the models with the best overall accuracy.

From the Microsoft Azure platform, feature importance could be obtained for the decision tree, decision forest, and logistic regression models. In each of these models, feeding data were among the top five predictors, and the percent change in feed delivery from days on feed (DOF) 2 to 3 was one of the most important predictors. Other predictors were percent change in feed delivery from DOF 4 to 5 and from DOF 8 to 9, and rolling average in percent change in feed delivery from DOF 7 to 10. These results indicate that alterations in feeding patterns very early in the feeding period are predictive of total morbidity, which may allow interventions to mitigate disease progression in the cohort.

In commercial feedlots, cattle are managed as pens except for individual animal treatment of disease. Tracking individual animal feeding behavior is not practical nor cost-effective in a commercial feedlot. However, commercial feedlots collect real-time data that can be used to make pen-level management decisions. The results of this study and the previous discussion indicate that feeding data can be used to predict morbidity in cattle. In our study, we looked at total morbidity, but BRD accounted for 65% of the total treatments. The predictive models in this study were somewhat predictive of total cohort level morbidity indicating that feeding data can be used to predict morbidity in pens of cattle, which could lead to improved management of disease in feedlots. However, in the current study, feed delivery data were used as predictors in

the models, which do not account for feed refusals. Based on previous data (Jackson, 2016), feed intake decreases prior to clinical signs of BRD and thus in our case, a decrease in the feed delivered is likely indicative of significant feed refusals and decreased feed intake for the day prior. Although, an increase in feed refusals could also be due to several other factors such as weather, removal of cattle from the pen, etc., which cannot be ascertained from our current dataset.

Conclusion

Feed intake and feeding behavior data are predictive of BRD in individual animals, and feed delivery data are predictive of total morbidity in commercial feedlot pens. Predictive analytics is a valuable tool that can be used to convert feedlot operational data into animal health management decisions. Future research should evaluate the ability of feed delivery to predict specific diseases (BRD, bloat, etc.), and combine feed delivery data with other data types to improve the prediction of morbidity in feedlot cattle.

References

- Amrine, D. E., J. G. McLellan, B. J. White, R. L. Larson, D. G. Renter, and M. Sanderson. 2019. Evaluation of three classification models to predict risk class of cattle cohorts developing bovine respiratory disease within the first 14 days on feed using on-arrival and/or pre-arrival information. *Comp Electron Agric.* 156:439–446. doi:10.1016/j.compag.2018.11.035.
- Amrine, D. E., B. J. White, and R. L. Larson. 2014. Comparison of classification algorithms to predict outcomes of feedlot cattle identified and treated for bovine respiratory disease. *Comp Electron Agric.* 105:9–19. doi:10.1016/j.compag.2014.04.009.
- Booker, C. W., G. H. Loneragan, P. T. Guichon, G. K. Jim, O. C. Schunicht, B. K. Wildman, T. J. Pittman, R. K. Fenton, E. D. Janzen, and T. Perrett. 2004. Practical application of epidemiology in veterinary herd health/production medicine. In: *American Association of Bovine Practitioners Proceedings of the Annual Conference; September 23–25; 2004; Fort Worth, TX; p. 59–62. (vol. 37).* doi:10.21423/aabppro20044902.
- Breiman, L. 2001. Random forests. *Mach. Learn.* 45:5–32. doi:10.1023/A:1010933404324.
- Buhman, M. J., L. J. Perino, M. L. Galyean, T. E. Wittum, T. H. Montgomery, and R. S. Swingle. 2000. Association between changes in eating and drinking behaviors and

1158 respiratory tract disease in newly arrived calves at a feedlot. *Am. J. Vet. Res.* 61:1163–
1159 1168. doi:10.2460/ajvr.2000.61.1163.

1160 Dreiseital, S., and L. Ohno-Machado. 2002. Logistic regression and artificial neural network
1161 classification models: a methodology review. *J. Biomed. Inform.* 35:352–359.
1162 doi:10.1016/s1532-0464(03)00034-0.

1163 Jackson, K. S., G. E. Carstens, L. O. Tedeschi, and W. E. Pinchak. 2016. Changes in feeding
1164 behavior patterns and dry matter intake before clinical symptoms associated with bovine
1165 respiratory disease in growing bulls. *J. Anim. Sci.* 94:1644–1652. doi:10.2527/jas.2015-
1166 9993.

1167 Microsoft. 2022. Azure machine learning—ML as a service. Microsoft Azure. Available from
1168 <https://azure.microsoft.com/en-us/services/machine-learning/>

1169 Moya, D., R. Silasi, T. A. McAllister, B. Genswein, T. Crowe, S. Marti, and K. S. Schwartzkopf-
1170 Genswein. 2015. Use of pattern recognition techniques for early detection of morbidity in
1171 receiving feedlot cattle. *J. Anim. Sci.* 93:3623–3638. doi:10.2527/jas.2015-8907.

1172 Quimby, W. F., B. F. Sowell, J. G. P. Bowman, M. E. Branine, M. E. Hubbert, and H. W.
1173 Sherwood. 2001. Application of feeding behaviour to predict morbidity of newly
1174 received calves in a commercial feedlot. *Can. J. Anim. Sci.* 81:315–320.
1175 doi:10.4141/A00-098.

1176 R Core Team. 2021. R: A language and environment for statistical computing. Vienna (Austria):
1177 R Foundation for Statistical Computing. Available from <https://www.R-project.org/>

1178 Roe, B. P., H. -J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor. 2005. Boosted decision trees
1179 is an alternative to artificial neural networks for particle identification. *Nucl. Instrum.*
1180 *Methods Phys. Res., Sect. A* 543:577–584. doi:10.1016/j.nima.2004.12.018.

1181 Rojas, H. A., B. J. White, D. E. Amrine, and R. L. Larson. 2022. Predicting bovine respiratory
1182 disease risk in feedlot cattle in the first 45 days post arrival. *Pathogens.* 11:442.
1183 doi:10.3390/pathogens11040442.

1184 Rosenblatt, F. 1957. The perceptron: a perceiving and recognizing automation. Buffalo (NY):
1185 Cornell Aeronautical Laboratory, Inc. Available from chrome-
1186 extension://efaidnbmnnnibpcajpcgiclfndmkaj/[https://blogs.umass.edu/brain-](https://blogs.umass.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf)
1187 wars/files/2016/03/rosenblatt-1957.pdf

1188 Salman, M. D., M. E. King, K. G. Odde, and R. G. Mortimer. 1991. Costs of veterinary services
 1189 and vaccines/drugs used for prevention and treatment of diseases in 86 Colorado cow-calf
 1190 operations participating in the National Animal Health Monitoring System (1986–1988).
 1191 J. Am. Vet. Med. Assoc. 198:1739–1744.

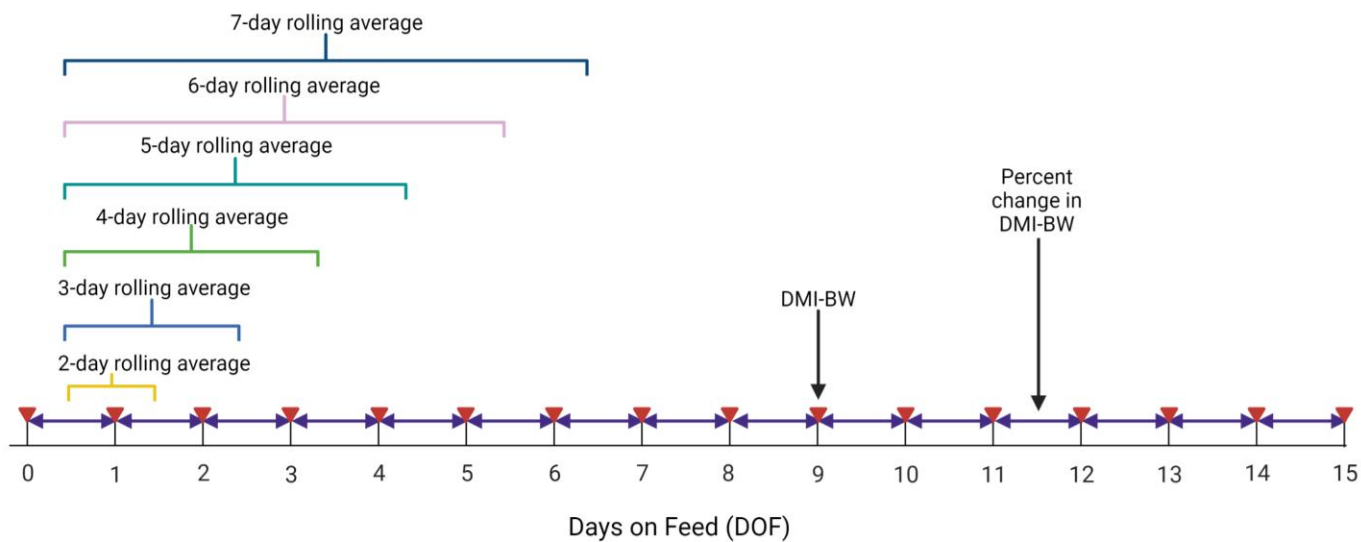
1192 Sowell, B. F., M. E. Branine, J. G. P. Bowman, M. E. Hubbert, H. E. Sherwood, and W.
 1193 Quimby. 1999. Feeding and watering behavior of healthy and morbid steers in a
 1194 commercial feedlot. J. Anim. Sci. 77:1105–1112. doi:10.2527/1999.7751105x.

1195 White, B. J., D. E. Amrine, and R. L. Larson. 2018. Big data analytics and precision animal
 1196 agriculture symposium: data to decisions. J. Anim. Sci. 96:1531–1539.
 1197 doi:10.1093/jas/skx065.

1198 Wolfger, B., K. S. Schwartzkopf-Genswein, H. W. Barkema, E. A. Pajor, M. Levy, and K. Orsel.
 1199 2015. Feeding behavior as an early predictor of bovine respiratory disease in North
 1200 American feedlot systems. J. Anim. Sci. 93:377–385. doi:10.2527/jas.2013-8030.

1201 Zhang, G. P. 2010. Neural networks for data mining. In: Maimon, O. and L. Rokach, editors.
 1202 Data mining and knowledge discovery handbook. Boston (MA): Springer US; p. 419–
 1203 444. doi:10.1007/978-0-387-09823-4_21

1204



1205
 1206 **Figure 2.1.** Timeline schematic demonstrating the feed delivery data corresponding to various
 1207 feeding predictor variables. Triangles indicate data on dry matter delivered as percentages of
 1208 arrival body weight (DMI-BW). Arrows indicate data on percent change in DMI-BW from day
 1209 to day.
 1210

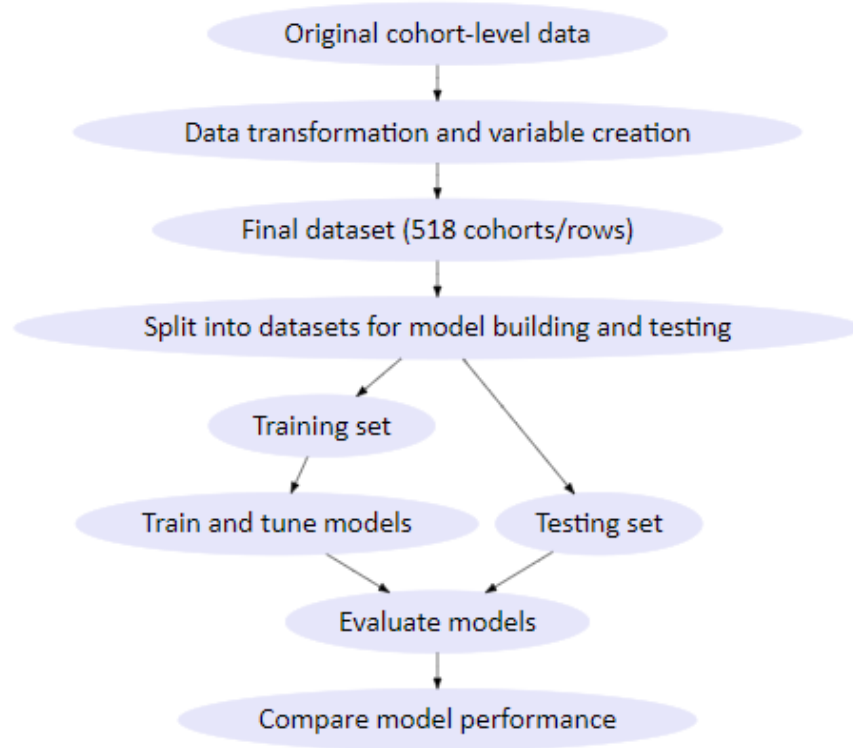
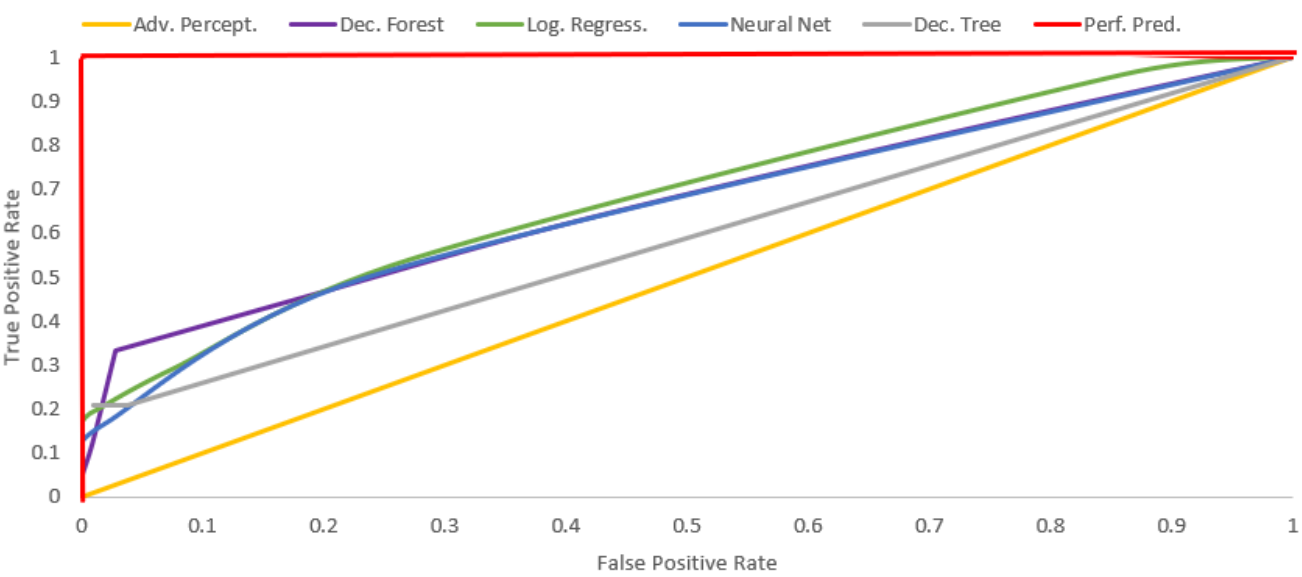


Figure 2.2. Illustration of data management, model training and evaluation process.

1214



1215

1216

1217

1218

1219

1220

Figure 2.3. Receiver operating characteristic (ROC) curves for five predictive models trained to predict high ($\geq 15\%$) morbidity cohorts of feedlot cattle. The five predictive models are Advanced Perceptron, Logistic Regression, Neural Network, Decision Tree, and Decision Forest. Perf. Pred. represents the perfect predictive model.

1221 **Table 2.1.** Complete overview of variables in the final dataset to train and test the five
 1222 predictive models
 1223

Variable category	Variable name	Description of variable
Feeding variables	DMI-BW ¹ for day 0 through 15	Feed intake measured by DMI-BW for each day starting on the day of arrival (0) to day 15 (16 total measurements)
	Percent change in DMI-BW from one day to the next for days 0 through 15	Percent change in feed intake measured by DMI-BW between sequential days (15 total measurements)
	2-day increment rolling averages of percent change in DMI-BW	Rolling averages in 2-day increments of percent change in DMI-BW (14 total measurements)
	3-day increment rolling averages of percent change in DMI-BW	Rolling averages in 3-day increments of percent change in DMI-BW (13 total measurements)
	4-day increment rolling averages of percent change in DMI-BW	Rolling averages in 4-day increments of percent change in DMI-BW (12 total measurements)
	5-day increment rolling averages of percent change in DMI-BW	Rolling averages in 5-day increments of percent change in DMI-BW (11 total measurements)
	6-day increment rolling averages of percent change in DMI-BW	Rolling averages in 6-day increments of percent change in DMI-BW (10 total measurements)
	7-day increment rolling averages of percent change in DMI-BW	Rolling averages in 7-day increments of percent change in DMI-BW (9 total measurements)
Cohort characteristics	Arrival date	Date of arrival for the cohort, format: MM/DD/YYYY
	Average arrival weight	Average weight at arrival of the cohort in pounds

	Sex	Sex of the cohort, could be heifer, steer, mixed
	Arrival head count	Head count of the cohort upon arrival
Outcome variable	Total morbidity category	High ($\geq 15\%$) or low ($< 15\%$) based on morbidity for any diagnosis as a percentage of arrival head count

1224 ¹DMI-BW indicates the feed intake on a dry matter basis given as a percentage of the average arrival weight of the
1225 cohort.

1226

Table 2.2. Model evaluation of five predictive models trained to predict high ($\geq 15\%$) morbidity cohorts of feedlot cattle

Model	AUC¹	Acc (%)	Sn (%)	Sp (%)	PPV	NPV
Advanced Perceptron	0.653	18.5	100.0	0.0	0.18	-
Logistic Regression	0.675	67.7	62.5	68.9	0.31	0.89
Neural Network	0.691	70.0	62.5	71.7	0.33	0.89
Decision Tree	0.691	78.5	29.2	89.6	0.39	0.85
Decision Forest	0.671	85.4	33.3	97.2	0.73	0.87

¹AUC = area under the receiver operator characteristic (ROC) curve; Acc = overall accuracy; Sn = sensitivity, ability to predict high morbidity cohorts; Sp = specificity, ability to predict low morbidity cohorts; PPV = positive predictive value, probability that predicted high morbidity cohorts are truly high morbidity cohorts; NPV = negative predictive value, probability that predicted low morbidity cohorts are truly low morbidity cohort

Chapter 3 - Evaluation of predictive models to determine final outcome for feedlot cattle based on information available at first treatment for bovine respiratory disease

L. Heinen¹, B.J. White¹, D.E. Amrine¹, R.L. Larson¹

¹Beef Cattle Institute, Kansas State University, Manhattan, KS 66505, USA

This research is published in the American Journal of Veterinary Research and is reproduced with permission from publisher as formatted in referenced publication. Lilli Heinen, Brad J. White, David E. Amrine, Robert L. Larson. Evaluation of predictive models to determine final outcome for feedlot cattle based on information available at first treatment for bovine respiratory disease, American Journal of Veterinary Research, Volume 84, Issue 10, August 2023, <https://doi.org/10.2460/ajvr.23.05.0094>.

Abstract

OBJECTIVE

To evaluate predictive model ability to determine whether an animal finished the feeding period using data known at first treatment for bovine respiratory disease (BRD). Additional comparisons evaluated the potential benefits of predictions by adding weather data, utilizing balancing techniques, and creating models for individual feedyards.

ANIMALS

This retrospective study included animal, pen, and feedyard data from 12 US feedyards from 2016 to 2021. The final dataset consisted of 96,382 BRD cases of which 14.2% did not finish the feeding phase.

PROCEDURES

Five predictive models were trained and underwent threshold probability adjustment to maximize F1 score. Model performance was evaluated using accuracy, sensitivity specificity, positive and negative predictive values, and area under the receiver operating characteristics curve (AUC).

RESULTS

Overall, model performance was low with a median AUC value of 0.675. The addition of weather data had little effect on AUC but resulted in more variation in sensitivity and specificity. Resampling the dataset had a limited effect on performance. Individual feedlot models had higher AUC values than others with the decision tree typically performing best in most feedyards.

CLINICAL RELEVANCE

Results indicated some utility of predictive models evaluating BRD cases to predict cattle that did not finish the feeding phase. These models could be valuable in assisting health providers making decisions on individual cases.

Keywords: bovine respiratory disease, machine learning, cattle, point-of-care diagnostics, case fatality risk

Introduction

Bovine respiratory disease (BRD) is among the costliest diseases encountered in the feedlot industry.¹ Distinguishing more severe cases from those that may require less attention could lead to more appropriate treatment and management decisions that improve treatment outcomes as well as reduce costs associated with unnecessary drug administration. Modern data analytic strategies including machine learning algorithms have been applied to various fields in the agricultural industry to decrease costs and increase outputs.²⁻⁴ Predictive modeling techniques have not been heavily studied for application in making animal health decisions in the beef industry.⁵ Feedyards collect large volumes of data daily, including but not limited to, feed calls, antimicrobials and other treatments administered, number of head per pen, and weight of individual animals and pens. These data and more are potential inputs for predictive models that could be used to predict the outcome of a singular calf at the time of first pull for BRD.

Risk factors associated with BRD morbidity and mortality have been well researched.⁶ Weight of incoming cattle to a feedyard, distance transported, time of the year, and more have been well-established markers for BRD management decisions.⁶ These markers have also been associated with disease outcome and severity.⁶ Historically, traditional statistics have been utilized in these kinds of studies to measure the association between predictor behavior and onset and outcome of BRD morbidity.^{7,8} Predictive analytics and machine learning approach the prediction of an outcome in a different way than traditional statistics. Rather than measure the

association of a predictor variable and an outcome of interest, predictive models utilize inputs to make predictions regardless of association.⁹

This study aims to evaluate the diagnostic ability of predictive models to determine whether an animal will finish its feeding period with its group or have a negative outcome based on data collected at first treatment for BRD. Additionally, we aim to determine the value of applying different data and model-tuning techniques in 3 different comparisons. One comparison will examine the difference between models created with and without weather data in addition to data from the feedyard. Another comparison aims to evaluate the impact that balancing a dataset with respect to minority outcome will have on the predictive ability of the models. Finally, the last comparison will analyze the differences between individualized models created for feedyards or the models created using a dataset consisting of many feedyards. We hypothesize that the predictive models created with such data will be able to accurately predict the outcome of an animal treated for BRD at a feedyard. In addition, we hypothesize that adding data and using techniques to increase the diagnostic ability of predictive models will have a positive impact on their performance in comparison to the original.

Methods

Animal Care and Use Committee approval was not obtained for this study due to data being obtained from existing databases provided by several feedyards.

Data collection

Treatment records were obtained from 12 US feedyards spanning 2016 to 2021 including information from 396,486 head of cattle. The feedyards were located throughout the Great Plains and were located in Kansas, Nebraska, Colorado, and Texas. Treatment records were on an individual animal basis and were tied to cohort-level data. A cohort was defined as a group of cattle purchased and managed similarly but not necessarily housed in the same pen for the entirety of the feeding phase. In addition to data obtained from feedyards, weather data were obtained using the Visual Crossing Weather application programming interface for the feedyard locations.¹⁰

Data transformation

Before data could be used to create predictive models, they were transformed into the appropriate format and all variables to be included in the models were generated. First, the

inclusion criteria were applied to the available data. To be included in the study, an animal had to be diagnosed with BRD and treated with an antimicrobial. Any data collected after the animal's first treatment for BRD were not considered. Secondly, data wrangling techniques were applied in R¹¹ to remove incomplete rows and extreme values. Criteria used to remove data are described (Table 3.1). The exclusion criteria included what the authors deemed to be extreme values, often caused by data entry errors. To minimize these, data points outside of the physiological limits of the cattle population were excluded. Additionally, animals with incomplete data were removed. A number of additional variables were created using the existing data to describe individual-, cohort-, and feedyard-level characteristics. These variables encompassed only information that would be known at the time of first treatment for BRD. A complete overview of all the variables that populated the dataset is available elsewhere (Appendix Table A.1). The outcome of interest was captured in a variable called did not finish (DNF). This variable described whether an animal finished the feeding period with its pen mates or if it died or was culled (railed) before that time. The data were then transformed so each row of data corresponded to an animal and each animal had data for each variable. Following data transformation and variable creation, the final dataset consisting of 96,382 animals with a 14.2% prevalence of DNF was used to create predictive models that would be tested on their ability to accurately predict DNF.

Model-building process

Before the model-building step occurred, the dataset to be used for the creation of predictive models was first split into separate training and testing sets, stratified by prevalence of outcome of interest, DNF. The training set represented 75% of the original dataset while the testing set represented the other 25%. Using the Pipeline Designer function in Azure Machine Learning,¹² the datasets were used to create a multitude of predictive models. Five different models, averaged perceptron, neural network, boosted decision tree, decision forest, and logistic regression, were trained with the same training dataset and evaluated with the corresponding testing dataset. These models were selected based on previous research⁵ and what was available in Pipeline Designer. Adjustment of the threshold probability was done manually to maximize F1 score to balance sensitivity (Sn) and PPV after the models were trained. The models are described in more detail below.

Neural network—Neural network models consist of units called neurons organized in layers. The input layer receives predictor data and transfers them to multiple hidden layers that

learn patterns in the data and connect them to the output layer which consists of the known outcome. Neural networks are adaptive and useful for real-world data in which patterns may be hard to identify.¹³

Averaged perceptron—Perceptron models perform similarly to neural networks but do not consist of multiple layers. Statistical measurements from large data are used to construct a system that can recognize patterns. Perceptron models can shorten the amount of it takes to categorize data into a prediction in comparison to conventional models.¹⁴ Both the neural network and averaged perceptron algorithms fall into the “black box” category of predictive analytics.

Boosted decision tree—Decision trees are classification models that split data based on discrete attributes within variables which best characterize a class to be predicted. One split is called a node. A following split creates a branch. These splits continue through all variables in the dataset. Boosting is an ensemble method that creates multiple additional trees based on the misclassification of the first tree. Boosted decision trees have been reported to be more robust and efficient than more complicated predictive models such as neural networks.¹⁵

Decision forest—Decision forests or random forest models are many decision trees built from randomly split sections of data. The large number of decision trees produced then vote for the most popular classification category.¹⁶ The decision forest model is intended for classification tasks.¹⁷

Logistic regression—Logistic regression is a well-known statistical method used to assign probability to an outcome. In machine learning, this model calculates the probability of class membership and uses it to make a prediction. Many papers have shown that logistic regression models work well on a large variety of data.¹⁸ Logistic regression models are manually built in a stepwise function in traditional statistics in which predictor variables that are not significantly associated are eliminated from the model. In predictive modeling, this process is completed by the computer without input from the researcher.

Model evaluation

Finally, the testing dataset was used to evaluate the performance of the trained and tuned models. Performance was evaluated using accuracy, Sn, specificity (Sp), positive, and negative predictive values (PPV and NPV, respectively), and area under the receiver operating characteristics (ROC) curve (AUC). These values were calculated using the confusion matrices

produced by each model run in Azure. Using these metrics, we were able to compare model sets based on their ability to accurately predict DNF for a calf treated for the first time due to BRD. All the model-building and evaluation steps are outlined (Figure 3.1). The performance of the models was then reevaluated in a series of comparisons to determine the benefit of adding weather data (Weather vs Base), balancing with regard to the minority outcome class (Balanced vs Base), and creating individualized models (Individual vs Base). The comparisons are discussed in detail below.

Weather versus Base models

The value of including weather data to predict the desired outcome was tested by comparing the base predictive models to models built with the same dataset with additional variables to describe weather. Weather predictive models received 80 more variables than the Base models. These included temperature, humidity, precipitation, and wind speed among others. The new weather variables described the weather on the same day as first treatment for BRD, in addition to up to 7 days before the initial BRD treatment occurred. This included weather variables from the day previous to the day of first treatment for BRD and moving averages of all weather variables from 3, 5, and 7 days previous to the treatment. This was done to capture weather information on the day of treatment and days leading up to treatment.

Balanced versus Base models

Next, models trained with varying degrees of prevalence of outcome were compared to the Base models (naïve prevalence of outcome). The purpose of this comparison was to test the value of “balancing” a dataset with respect to outcome to increase predictive model performance.¹⁹ To balance a dataset means to increase the prevalence of the minority outcome (in the case of our study, the percentage of cattle that DNF) to provide the model with more data to predict that outcome. In this comparison, the training dataset used to train the predictive models proceeded to an additional balancing step that either randomly duplicated minority class observations (oversampling) or randomly deleted majority class observations (undersampling) until the prevalence of DNF was 50%. This step was performed in R using the tidymodels package.²⁰ The Base models were trained on a dataset with a DNF prevalence of 14.2%. The same models were retrained using datasets with a DNF prevalence of 50% created by

oversampling and undersampling, respectively. This made for a total of 3 sets of models that we compared to each other to determine the benefit of oversampling the minority class and undersampling the majority class. The flow of data for the creation of the datasets used to train the Balanced models is demonstrated (Figure 3.2).

Individual versus Base models

Finally, the Base models were compared to models built for individual feedyards. Each feedyard's data were separated from the rest and used to build the same models as the base. This resulted in the creation of 12 new sets of predictive models whose performance was compared to that of the Base models. The purpose of this comparison was to determine the value of creating individualized models for a feedyard instead of using models built using combined data.

Results

Evaluation of Base models

Five models (averaged perceptron, decision forest, logistic regression, neural network, boosted decision tree) were trained and evaluated. Table 3.2 provides a summary of the metrics used to evaluate performance for the Base models and all comparisons. The boosted decision tree model had the highest AUC. Overall, all algorithms possessed high NPVs and low PPVs. Despite low PPVs, the ability to identify cattle that DNF using predictive models is greater than simply guessing at baseline prevalence. This is because the baseline prevalence of animals that DNF was 14.2% and 4 out of 5 of the models had a PPV above this value. High NPVs give high confidence in the animals that the models predicted to finish. It is important to keep in mind that the prevalence of animals that finished the feeding period was much higher than those that DNF, which would naturally increase the NPV.

Comparison of weather and Base models

The 5 base models (Base) were compared to the same 5 models created using additional weather data (Weather). AUC values decreased by very little and other evaluation metrics between the model sets did not vary greatly. The ranges in Sn and Sp values increased numerically in the Weather models when compared to the Base models. For both model sets, the boosted decision tree algorithm ranked highest with respect to AUC (0.735 and 0.742 for

Weather and Base, respectively). Overall, Weather models demonstrated very similar performance to the Base models with respect to the metrics evaluated in this study.

Comparison of balanced versus Base models

The Base models were compared to models that oversampled and undersampled data for cattle that DNF so that the number of cattle that DNF accounted for 50% of the animals in the training dataset. Again, 5 models were created for each sampling target and compared to the Base models. Models performed similarly across sets with regard to AUC. Numerically, AUC values tended to increase with both sampling strategies in the logistic regression and decision forest models. The boosted decision tree model received the highest AUC for each model set (Table 3.2). The Balanced models demonstrate a tighter spread of Sn, Sp, and Accuracy values when compared to the Base models. PPV and NPV for the Balanced models resembled what was observed for the Base models. Overall, balancing a dataset with respect to the outcome of interest resulted in marginal performance benefits regardless of the balancing strategy.

Comparison of individual versus Base models

Each yard's dataset was used to train 5 models, which were then compared to the Base models. Median AUC values were more varied, with a range from 0.548 to 0.868 between yards (Figure 3.3). Boosted decision tree models led again with regard to the highest AUC value. Accuracy, Sn, and Sp values followed the pattern observed for the previous comparisons. The model set for yard 6 was the only one to demonstrate a different pattern compared to other yards with very high Sp values (average of 90.3%) and low Sn values (average of 37.8%). A notable difference from the Base models was several yards producing models with high PPVs (maximum: 0.71 for the boosted decision tree model from yard 8). The results for NPV remained similar to the pattern of the Base models.

Discussion

Accurate and early diagnosis of BRD in cattle in feedyards can reduce costs associated with morbidity and improve the health outcome for the animal.²⁰ Prudent antimicrobial use is also of importance in the industry with efforts made to reduce metaphylactic use of antimicrobials to reduce both morbidity and mortality associated with BRD. Diagnostic strategies at the time of treatment in combination with visual cues contribute to the diagnosis of BRD in the feedyard today. These include rectal temperature and lung auscultation. Many

studies have been conducted to compare new diagnostic strategies such as computer-aided lung auscultation (Whisper) and serum acute phase protein concentrations.²¹ Our study does not attempt to diagnose BRD but instead aims to diagnose DNF, using evaluation metrics similar to ones used in a previous study.⁵ Predicting whether or not an animal will finish the feeding period could have a longer term effect, as an accurate prediction provides information on the calf's final outcome, thus offering an opportunity to make interventions to prevent a negative outcome. However, the predictive ability of our models is not yet good enough to be used in a real-time, feedyard setting. Further development and research could make this possible.

The Whisper Stethoscope automatically assigns a respiratory score and has been evaluated on its ability to detect BRD at the time of treatment. By including a fever cut-point, its AUC (0.686) is similar to the median seen in our Base model set of 0.675.²² Using acute phase proteins, such as haptoglobin and lipopolysaccharide binding protein, to predict BRD morbidity can yield AUCs as high as 0.967.²³ The Whisper Stethoscope²² can be used at the time the animal removed from the pen for treatment. The sample for acute phase protein analysis kits is collected at the time of treatment but the test must be performed in a lab.²² Both methods of predicting BRD morbidity require a significant monetary investment, either up front in the case of the Whisper Stethoscope, or on a per head basis, as with the acute phase protein test kit. The predictive models evaluated in our study were created on the Azure Machine Learning platform by Microsoft, which allows users to train and deploy predictive models for use in the field.¹² Such a tool would allow for immediate results at the time of treatment with little input from employees on the ground, given that most of the data are already collected by the feedyard and can automatically populate the model of choice. Additionally, utilizing predictive analytic strategies in place of traditional statistics could help investigate outcomes using data that traditional statistical methods have difficulty handling. For example, neural network models are designed to work as “black box” models for use on real-world data in which patterns are hard to identify.¹³

Despite the relative ease of use of the predictive models investigated in this study, their predictive power was lacking. Additionally, predictive models, unlike explanatory models, do not provide associations between risk factors and outcomes. Traditional statistical methods provide the mathematical equations used to assess causal relationships between predictor

variables and the outcome of interest. Predictive models focus on providing the most accurate prediction, without regard for interpretability, which makes their outputs difficult to assess.

Predictive models require large amounts of data for training. A common rule of thumb is that the training set should have more than 10 times as many rows as columns to ensure adequate variability in the features of interest to make accurate predictions. Our training set surpassed this minimum threshold, and thus, we feel confident that the amount of data used for training the predictive models investigated in this study was sufficient. However, the data used to predict DNF were extremely variable, which is often the case in biology. Additionally, data were collected at multiple different operations, and while many of the measures used were objective, the data collection methods and accuracy may have varied by operation. This could have limited the performance of the predictive models. Methods to decrease the variability of the data used in this study were not investigated but could be an interesting area for future research.

A common finding among the models tested in this study was an increased ability to accurately classify the majority class (animals that finished the feeding period with their pen mates) in comparison to the minority class (animals that DNF the feeding period with their pen mates). This finding is demonstrated by the overall lower Sn values compared to the higher Sp values. This is a ubiquitous finding in algorithms attempting to predict a rare or minority outcome.²⁴ Often the minority outcome is most interesting to the user, as is the case in this study. In such cases, evaluating predictive models based on accuracy alone can be misleading.²⁴ The models will have a bias toward the majority class. Hiaxiang et al²⁵ explain that the evaluation of such models should utilize other metrics besides accuracy alone. Among others, the AUC of a ROC curve is adapted for use in imbalanced data problems because they account for the distribution of classes. For this reason, AUC as well as accuracy, Sn, Sp, PPV, and NPV were used to characterize the performance of all models evaluated in this study.

Many techniques are employed to improve the performance of predictive models across many fields. For example, ensemble methods, such as boosting and stacking, improve the performance of classification models trained with an imbalanced dataset.²⁵ This likely explains the greater AUC values, and thus greater predictive power, of the decision tree models across multiple comparisons demonstrated in this study. This feature was built into the decision tree model but none of the others used in this study. These findings indicate the potential benefit of adding ensemble methods to other models to increase performance.

Other common strategies, utilized especially in biomedical engineering models, are resampling techniques.²⁵ Oversampling and undersampling are both popular techniques applied to the training dataset before model building. However, using these strategies does not always increase performance. Oversampling techniques are prone to creating overgeneralized models and can be problematic for predicting a very rare outcome.²⁴ An overgeneralized model is one that can make accurate predictions with the training data but cannot generalize its model to a different dataset. This makes the model less accurate when exposed to naïve data. Undersampling has been found to be superior to oversampling in datasets with hundreds of minority observations.^{5,25} Additionally, combining the 2 strategies to resample the training dataset has shown promise²⁵ but was not investigated in this study.

Our results demonstrated little benefit in utilizing a resampling strategy on the training dataset. This could be due to overgeneralization of our models. This has been described with an oversampling strategy but not an undersampling strategy.²⁵ Small numerical increases in the performance metrics of the models investigated in this study indicate that further investigation into balancing techniques may be warranted. Only preprocessing methods were used in this study. Learning methods applied to the predictive models themselves, and postprocessing methods may contribute a greater benefit to the model performance.

High dimensionality is another issue commonly seen in predictive analytics. It is intuitive to believe that the addition of more data can only have a beneficial effect on the predictive ability of a model. However, research has shown that high dimensionality in a training dataset can lead to noisy data contributing to the prediction of an outcome.²⁴ This can have a degradative effect on the evaluation metrics. High dimensionality could have contributed to the underwhelming performance of the predictive models in this study.

In the case of our Weather versus No Weather comparison, we saw virtually no difference in the median AUC between the 2 model sets (0.674 vs 0.675). This comparison showed neither a deleterious effect nor a beneficial effect with the addition of weather data. Dimensionality reduction methods may have been useful in this comparison but were not investigated. Previous research has shown that weather variables are significantly associated with increased BRD morbidity and mortality.^{26,27} The current study cannot evaluate the associations between the weather data included and DNF. However, the lack of increased performance suggests weather data do not increase the predictive power of the models evaluated in this study.

Management and data collection vary from feedyard to feedyard. For this reason, models created for individual yards were compared to the Base models. Amrine et al made a similar comparison in their 2012 paper. In that study, the COMBO dataset created models that ranged in accuracy from 52% to 77%.⁵ Our study found a wider range of accuracies (14.2% to 75.9%). This is likely due to differences in model selection and parameter tuning. A similar finding in both studies was the occurrence of several models classifying every animal as DNF, resulting in high Sn values and low Sp values. Our research shows that the creation of models for individual feedyards can produce models that perform better than models created using the base dataset. However, some individual feedyard models were not as accurate as the Base models. This indicates some differences in quality of data collected regarding predictive ability. Research investigating the differences in feedyard data is needed to better understand the variation in model performance across feedyards.

The objective of this study was to evaluate the diagnostic ability of various predictive models to determine final outcome of an animal at its first treatment for BRD. This study also aimed to determine the benefit of including weather data, utilizing data balancing techniques, and creating individualized models. Models were tuned by maximizing Sn and PPV in relation to each other using F1 score. We evaluated model performance using various metrics including AUC of the ROC curve, Sn, Sp, PPV, NPV, and overall accuracy. AUC value was used to evaluate overall predictive ability of each of the models.

Wide ranges for all evaluation metrics were noted across all models trained and tested. This could be due to ranges in data quality as well as ranges in the predictor variables. The model algorithms themselves may contribute to this range as some models are better suited to the type of data used in this study than others. Overall, the boosted decision tree model outperformed other models consistently across comparisons with regard to AUC. The addition of weather data had very little effect on AUC, likely due to its contribution to high dimensionality which can negatively impact model performance. Balancing the training dataset did not benefit model performance. This is likely caused by overgeneralization of the models trained with an oversampled dataset, which can lead to a decreased predictive performance when subsequently exposed to the testing dataset. Creation of individual models for feedyards can lead to increased performance when compared to the Base models for some feedyards. However, other feedyards produced models that performed worse in comparison to the Base models. This finding is likely

due to the vast difference in information quality and warrants further research. At this time, the predictive models evaluated in this paper are not likely to greatly improve detection of animals that will not finish. More research could lead to the improvement of predictive models applied to this problem which could lead to application to a modern-day feedyard treatment protocol to assist in decision-making in real time. Early culling, application of more aggressive treatment, or moving that animal into a convalescent pen are all possible interventions that could be investigated using models that predict DNF. Future research is needed to elucidate the economic and animal health benefits of these options.

References

1. Salman M, King M, Odde K, et al. Costs of veterinary services and vaccines/drugs used for prevention and treatment of diseases in 86 Colorado cow-calf operations participating in the National Animal Health Monitoring System (1986 - 1988). *JAVMA* 1991;198:1739–1744.
2. Cantor MC, Costa JHC. Daily behavioral measures recorded by precision technology devices may indicated bovine respiratory disease status in preweaned dairy calves. *Int J Dairy Sci* 2022;105:6070–6082.
3. Feuz R, Feuz K, Johnson M, et al. Improving Feedlot Profitability Using Operational Data in Mortality Prediction Modeling. *J Agric Econ* 2021;46:242–255.
4. Wisnieski L, Norby B, Pierce SJ, et al. Predictive models for early lactation diseases in transition dairy cattle at dry-off. *Prev Vet Med* 2019;163:68–78.
5. Amrine DE, White BJ, Larson RL. Comparison of classification algorithms to predict outcomes of feedlot cattle identified and treated for bovine respiratory disease. *Comput Electron Agric* 2014;105:9–19.
6. Avra TD, Abell KM, Shane DD, et al. A retrospective analysis of risk factors associated with bovine respiratory disease treatment failure in feedlot cattle. *J Anim Sci* 2017;95:1521–1527.
7. Kayser WC, Carstens GE, Jackson KS, et al. Evaluation of statistical process control procedures to monitor feeding behavior patterns and detect onset of bovine respiratory disease in growing bulls. *J Anim Sci* 2019;97:1158–1170.

8. Wottlin LR, Carstens GE, Kayser WC, et al. Efficacy of statistical process control procedures to monitor deviations in physical behavior for preclinical detection of bovine respiratory disease in feedlot cattle. *Livest Sci* 2021;248.
9. Wisnieski L, Amrine DE, Renter DG. Predictive modeling of bovine respiratory disease outcomes in feedlot cattle: a narrative review. *Livest Sci* 2021;251.
10. Anon. Timeline Weather API. Visual Crossing Available at: <https://www.visualcrossing.com/>. Accessed August 10, 2021.
11. Anon. R: a language and environment for statistical computing. R Foundation for Statistical Computing Available at: <https://www.R-project.org/>. Accessed May 28, 2021.
12. Anon. Microsoft Azure Machine Learning. Microsoft Available at: <https://azure.microsoft.com/en-us/services/machine-learning/>. Accessed May 24, 2021.
13. Zhang GP. Neural networks for data mining. In: *Soft Computing for Knowledge Discovery and Data Mining*. Springer, Boston, MA, 2008;17–44.
14. Rosenblatt F. The perceptron: a perceiving and recognizing automaton. Cornell Aeronautical Laboratory, Inc.: Cornell University; 1957.
15. Roe BP, Yang H, Zhu J, et al. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nucl Instrum, Methods Phys Res A: Accel Spectrom Detect Assoc Equip* 2005;543:577–584.
16. Breiman L. Random Forests. *Mach Learn* 2001;45:5–32.
17. Anon. Two-Class Decision Forest component. Microsoft Docs Azure Machine Learning 2021. Available at: <https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-decision-forest>. Accessed May 22, 2022.
18. Dreiseital S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform* 2002;35:352–359.
19. Japkowicz N. Learning from imbalanced data sets: a comparison of various strategies. 2000.
20. Kuhm M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.; 2022. Available at: <https://www.tidymodels.org/>. Accessed May 18, 2022.

21. Wolfger B, Manns BJ, Barkema HW, et al. Evaluating the cost implications of a radio frequency identification feeding system for early detection of bovine respiratory disease in feedlot cattle. *Prev Vet Med* 2015;118:285–292.
22. Wolfger B, Timsit E, White BJ, et al. A systematic review of bovine respiratory disease diagnosis focused on diagnostic confirmation, early detection, and prediction of unfavorable outcomes in feedlot cattle. *Vet Clin N Am – Food Anim Pract* 2015;31:351–365.
23. Noffsinger T, Brattain K, Quakenbush G, et al. Field results from Whisper stethoscope studies. *Anim Health Res Rev* 2014;15:142–144.
24. Idoate I, Vander Ley B, Schultz L, et al. Acute phase proteins in naturally occurring respiratory disease of feedlot cattle. *Veterinary Immunology and Immunopathology* 2015;163:221–226. Available at: <https://linkinghub.elsevier.com/retrieve/pii/S0165242714003080>. Accessed April 12, 2023.
25. Branco P, Torgo L, Ribeiro RP. A survey of predictive modeling on imbalanced domains. *J ACM* 2016;49:1–50.
26. Hiaxiang G, Yiing L, Shang J, et al. Learning for class-imbalanced data: review of methods and applications. *Expert Syst Appl* 2017;73:220–239.
27. Cernicchiaro N, Renter DG, White BJ, et al. Associations between weather conditions during the first 45 days after feedlot arrival and daily respiratory disease risks in autumn-placed feeder cattle in the United States¹. *Journal of Animal Science* 2012;90:1328–1337. Available at: <https://academic.oup.com/jas/article/90/4/1328-1337/4764639>.
28. Wisnieski LC, Amrine DE, Cernicchiaro N, et al. Weather conditions associated with death attributed to bovine respiratory disease complex in high-risk auction market-sourced male beef calves. *ajvr* 2021;82:644–652. Available at: <https://avmajournals.avma.org/view/journals/ajvr/82/8/ajvr.82.8.644.xml>.

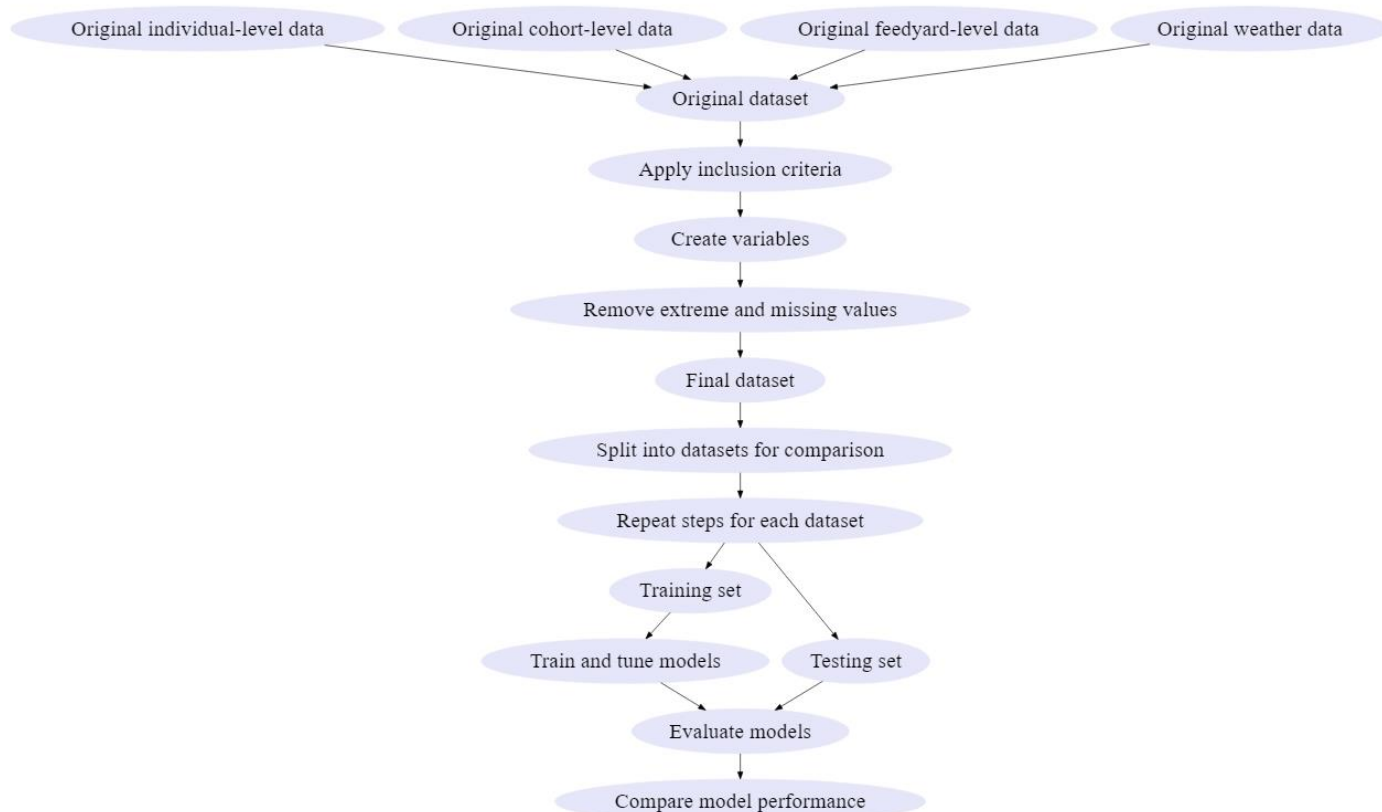
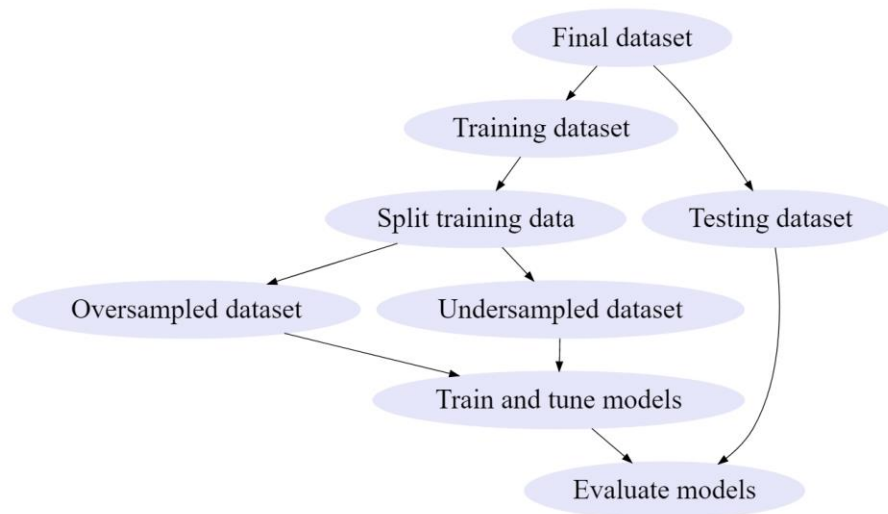


Figure 3.1. Flowchart describing the steps followed to create the base predictive models investigated in this trial. Inclusion criteria include a first-time treatment for bovine respiratory disease with an antimicrobial.

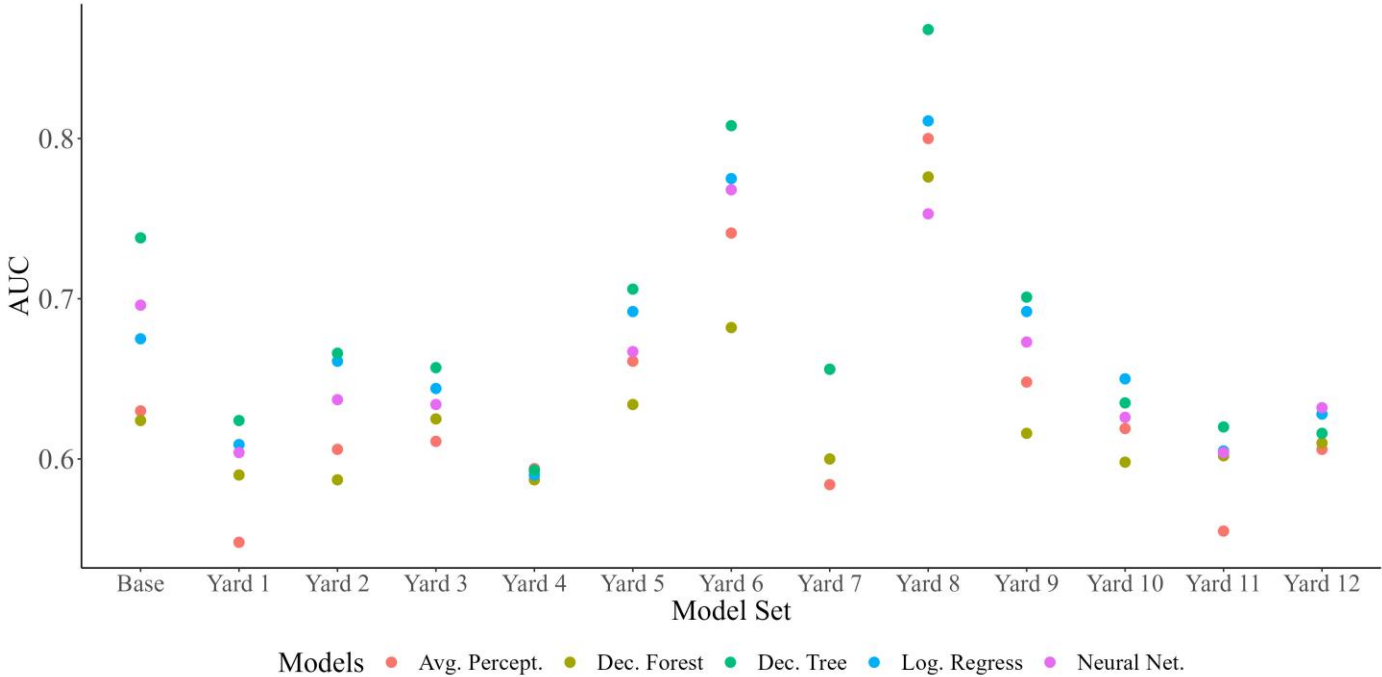
1684
1685



1686
1687
1688
1689
1690
1691

Figure 3.2. Flowchart describing the extra steps followed to balance the training dataset with respect to minority outcome class before model training. This process allowed for the creation of 3 model sets built with dataset of varying prevalence of did not finish: 14.2% (Base) and 50% (Oversample and Undersampled).

1692



1693

1694

1695

1696

1697

Figure 3.3. AUC in Base predictive models including all feedyards compared to individual feedyard models. Each point demonstrates the AUC value for 5 predictive models (specific model designated by color).

1698 **Table 3.1.** Summary of exclusion criteria and the rows of data removed from the study. Rows
 1699 were removed in order of the criteria listed. 300,104 animals were removed from the study.

Criteria to remove data	Rows removed
Day of pull less than 0	714
Sex of the lot other than mixed, steer, or heifer	3,324
Weight at time of pull greater than 1360.8 kg (3000 lbs.)	15
Lot's average arrival weight greater than 1360.8 kg (3000 lbs.)	1,884
Lot's head in greater than 600	1,679
Lot's arrival data after 02/01/2021	15
Animals with incomplete data	292,473

1700

1701

1702

1703 **Table 3.2.** Summary of Base model metrics and results.

Model Results						
Model	AUC	Sn (%)	Sp (%)	Accuracy (%)	PPV	NPV
Base						
Avg. Percept.	0.630	100.0%	0.0%	14.2%	0.14	DBZ*
Dec. Forest	0.624	70.3%	47.6%	50.8%	0.18	0.91
Log. Regress.	0.675	55.5%	68.8%	66.9%	0.23	0.90
Neural Net.	0.696	50.0%	76.2%	72.5%	0.26	0.90
Dec. Tree	0.738	57.0%	76.3%	73.6%	0.28	0.91
Weather						
Avg. Percept.	0.623	100.0%	0.0%	14.2%	0.14	DBZ
Dec. Forest	0.612	47.8%	68.8%	65.9%	0.20	0.89
Log. Regress.	0.674	50.8%	72.8%	69.7%	0.24	0.90
Neural Net.	0.69	59.8%	67.0%	65.9%	0.23	0.91
Dec. Tree	0.735	51.2%	80.0%	75.9%	0.30	0.91
Undersampled						
Avg. Percept.	0.613	47.0%	68.4%	65.4%	0.20	0.89
Dec. Forest	0.664	52.2%	70.2%	67.6%	0.22	0.90
Log. Regress.	0.678	46.1%	77.5%	73.1%	0.25	0.90
Neural Net.	0.680	46.5%	77.2%	72.9%	0.25	0.90
Dec. Tree	0.726	63.3%	69.3%	68.4%	0.25	0.92
Oversampled						
Avg. Percept.	0.619	68.8%	59.4%	59.2%	0.19	0.90
Dec. Forest	0.652	53.4%	68.8%	66.7%	0.22	0.90
Log. Regress.	0.678	60.7%	64.9%	64.3%	0.22	0.91
Neural Net.	0.673	56.6%	68.6%	66.9%	0.23	0.91
Dec. Tree	0.732	50.3%	80.8%	76.5%	0.30	0.91

1704 *DBZ indicates a division by zero

Chapter 4 - Economic impact of mortality prediction by predictive model at first and second treatment for bovine respiratory disease

L. Heinen¹, B.J. White¹, R.L. Larson¹, D. Kopp¹, D.L. Pendell²

¹Beef Cattle Institute, Kansas State University, Manhattan, KS 66506, USA

²Department of Agriculture Economics, Kansas State University, Manhattan, KS 66502, USA

This research is published in the American Journal of Veterinary Research and is reproduced with permission from publisher as formatted in referenced publication. Lilli Heinen, Brad J. White, Robert L. Larson, Dannell Kopp, Dustin L. Pendell. Economic impact of mortality prediction by predictive model at first and second treatment for bovine respiratory disease, American Journal of Veterinary Research, Volume 85, Issue 12, September 2024, <https://doi.org/10.2460/ajvr.24.06.0169>.

Abstract

OBJECTIVE

To evaluate a predictive model's ability to determine cattle mortality following first and second treatment for bovine respiratory disease and to understand the differences in net returns comparing predictive models to the status quo.

METHODS

2 boosted decision tree models were constructed, 1 using data known at first treatment and 1 with data known at second treatment. Then, the economic impact of each outcome (true positive, true negative, false positive, and false negative) was estimated using various market values to determine the net return per head of using the predictive model to determine which animals should be culled at treatment. This was compared to the status quo to determine the difference in net return.

RESULTS

The models constructed for the prediction of mortality performed with moderate accuracy (areas under the curve > 0.7). The economic analysis found that the models at a high specificity (> 90%) could generate a positive net return in comparison to status quo.

CONCLUSIONS

This study showed that predictive models may be a useful tool to make culling decisions and could result in positive net returns.

CLINICAL RELEVANCE

Bovine respiratory disease is the costliest health condition experienced by cattle on feed. Feedyard record-keeping systems generate vast amounts of data that could be used in predictive models to make management decisions. It is essential to understand the accuracy of predictions made via machine learning. However, the economic impact of implementing predictive models in a feedyard will influence adoption.

Keywords: bovine respiratory disease, diagnosis, feedlot, machine learning, net returns

Introduction

Bovine respiratory disease (BRD) is among the costliest diseases encountered in the feedlot industry.¹ Few diagnostics can be efficiently applied at the time of treatment to predict the final outcome of an animal diagnosed with BRD in the feedyard. Economic loss due to mortality is a concern for feedyard managers, and early shipments of chronically sick or poor-doing cattle at a discounted price (commonly called railing or realizing) is used to ensure some return on investment in these cattle. Modern data analytic strategies, including machine learning algorithms and predictive models, are now being researched for use in dairy and beef production.²⁻⁴ The vast amount of data produced by feedyards today make them an environment ideal for testing the application of predictive models to assist management decisions. Predictive analytics and machine learning approach the prediction of an outcome differently than traditional statistics. Rather than measure the association of a predictor variable and an outcome of interest, predictive models utilize inputs to make predictions regardless of association.⁵

Risk factors associated with BRD morbidity and mortality have been well-researched using traditional statistical methods. Weight of incoming cattle to a feedyard, distance transported, time of the year, and more have been well-established inputs for BRD management decisions.⁶ These data are readily available in a feedyard's data management system and could easily be incorporated into a predictive model. The performance of a predictive model is measured using several metrics that describe its accuracy, but when applying this tool to make management decisions in a feedyard, it is also important to understand the economic implications.

The objective of this study is to determine the ability of a boosted decision tree predictive model to predict mortality within 15 days of an animal's first or second treatment for BRD and to quantify the net return of using the predictive model compared to status quo at different probability thresholds as a management tool to decide which animals should be culled (railed) ahead of their pen mates. We hypothesize that the boosted decision tree models will perform with moderate accuracy when asked to determine the outcome of an animal in the next 15 days. We anticipate that by using the models to determine which animals should be railed, feedyard managers may have the ability to return an economic benefit depending on which probability threshold is utilized.

Methods

IACUC approval was not obtained for this study due to data being obtained from existing databases provided by several feedyards.

Data collection

Treatment records were obtained from 13 US feedyards from 2018 through 2021 and included information from 396,486 head of treated cattle. Feedyards included in this dataset were located in the central High Plains region of the US. Records obtained from each yard were on an individual animal basis and were tied to cohort-level data. A cohort was defined as a group of cattle purchased and managed similarly but not necessarily housed in the same pen for the entirety of the feeding phase. In addition to data collected from feedyards, relevant economic data were acquired from a variety of sources. Cull cow slaughter prices were obtained from the USDA Agricultural Marketing Service database.⁷ These records were obtained using an application programming interface call to the system to obtain cull cow slaughter prices from each state represented in the dataset. The cull cow slaughter prices were later used in the economic analysis. Monthly fed cattle prices from January 2018 through December 2021 were obtained from the CattleFax online database.⁸ Historic commodity prices from October 2018 through May 2019 were used to determine the price per pound of a feed ration. Alfalfa hay and corn prices were acquired from *Focus on Feedlots Monthly Reports*.⁹ Wet distillers grain price was from Agriculture Marketing Resource Center's *Weekly Ethanol, Distillers Grain, and Corn Prices* and was an average of the monthly prices from Iowa, Illinois, Nebraska, South Dakota, and Wisconsin.¹⁰

Feedyard data transformation

Before feedyard data could be used to create predictive models, they were transformed into the appropriate format, and feature engineering was used to generate additional variables to be included in the models. These steps were all performed in R.¹¹ First, inclusion criteria were applied to the dataset, including validating records, only cohorts with heifers, steers or mixed genders, cohort and treatment weights less than 3,000 pounds, cohort arrival head counts less than 600, and cattle meeting case definition for each model (Figure 4.1). These inclusion criteria were applied to minimize the inclusion of data entry errors and to ensure that only animals with complete data were included in the dataset.

Two different datasets were created: 1 included all data known at cattle's first treatment for BRD (BRD1), and the other included all data known at cattle's second treatment for BRD (BRD2). All cattle in the second treatment dataset were included in the first treatment dataset. A case definition was created for each dataset. The diagnosis of BRD was based on standard operational procedures for each feedyard, which involved identifying individual animals within each cohort that illustrated signs typical of BRD, which included depression, anorexia, and increased respiration. Following diagnosis in the pen, animals were taken to a hospital facility where diagnosis was confirmed. To be considered a first treatment for BRD, an animal was administered an antimicrobial for a diagnosis of BRD for the first time. To be considered a second treatment for BRD, an animal was administered an antimicrobial for a diagnosis of BRD for the second time. These case definitions further refined our datasets and are also represented in Figure 3.1.

Feature engineering and variable generation consisted of creating new variables using the existing data to describe individual-, cohort-, and feedyard level characteristics. These variables encompassed only information that would be known at the time of first or second treatment for BRD for each respective dataset (Appendix Tables B.1 and B.2). The outcome of interest was captured in a variable called DNF_15 (did not finish at 15 days post treatment event). This variable described whether an animal experienced mortality or was culled (railed) within the next 15 days following the BRD treatment of interest. The cutoff of 15 days was used based on an assessment of the range of days post treatment event, which demonstrated that approximately 40% of mortalities and railing events occurred by 15 days after treatment. Data were then transformed so that each row of data corresponded to 1 animal and each animal had data for all

variables. The final BRD1 dataset consisted of 60 variables, and the final BRD2 dataset consisted of 102 variables, with 60 of those repeated from the BRD1 dataset. These datasets were used to construct predictive models capable of determining the DNF_15 of individual animals at first or second treatment for BRD.

Model building process

Before the model building step occurred, each dataset was first split into separate training and testing sets, stratified by prevalence of outcome of interest, DNF_15. The training set represented 75% of the original dataset, whereas the testing set represented the other 25%. Using the Pipeline Designer function in Azure Machine Learning, datasets were used to train, test, and, eventually, validate a boosted decision tree model.¹² This algorithm was chosen out of the many possible options due to prior experience and research done within our group.^{2,13} Decision trees are classification models that split data based on discrete attributes within variables that best characterize a class to be predicted. One split is called a node. A following split creates a branch. These splits continue through all variables in the dataset. Boosting is an ensemble method that creates multiple additional trees based on the misclassification of the first tree. Boosted decision trees have been reported to be more robust and efficient than more complicated predictive models, such as neural networks.¹⁴

Before model training occurred, a 2-way chi-squared feature scoring method was applied to the training dataset in order to remove redundant variables in each dataset. Feature selection allows for better performance of a predictive model and the removal of unnecessary variables.¹⁵ Following 2-way chi-squared analysis, features were selected for inclusion in machine learning models.

Hyperparameters are used in machine learning to adjust the model training process and optimize performance. In this study, a module provided by Azure Machine Learning performed a parameter sweep to identify the hyperparameters that would maximize the area under the receiver operating characteristics curve (AUC).¹⁶ This sweep was performed with the training dataset following feature selection. Hyperparameter tuning led to adjustments made in the default architecture. The tuning process indicated that the same parameters should be used for the BRD1 and BRD2 models to maximize the predictive power of the models as indicated by the highest AUC value. The maximum number of leaves per tree was set to 32, resulting in a maximum of 32 nodes that could be created by any tree. A larger number would put the model at

risk of overfitting the data. A minimum of 50 samples per leaf node was set. This means that at least 50 cases were needed in order to create a rule. The number of trees constructed was set to 500. Creating more trees results in better coverage of the data but increases training time. Lastly, the learning rate was set to 0.025. Adjustment of the learning rate defines how fast the learner converges on the optimal solution. Decreasing the learning rate results in more precision but increases training time.¹⁷

The boosted decision tree model was now tested to predict the outcome of interest, DNF_15, using the testing dataset that had been withheld from the previous training steps. This was done using a technique called crossfold validation. This technique is commonly applied in machine learning to assess the variability and reliability of the dataset used to train a model.¹⁸ The testing datasets were each divided randomly into 10 folds, with data from 1 fold retained for validation. The model is trained on 9 folds, and final performance metrics are calculated using the validation fold. The process is repeated, with each fold undergoing a turn as the validation fold. The end result is 10 sets of performance metrics, allowing for evaluation of the dataset's variability and the model's reliability. To simplify the analysis, the fold resulting in the highest AUC value was selected for further evaluation.

Model performance was evaluated using accuracy, sensitivity, specificity, positive and negative predictive values, and AUC. These values were calculated using the confusion matrices produced by each model run in Azure. Figure 4.2 outlines all the model building and evaluation steps. The usefulness of the results needed to be put into context, so an economic analysis was performed using the scored datasets obtained from the crossvalidation fold of each predictive model with the highest AUC. The economic analysis is described in further detail below.

Economic analysis

In order to address the economic implications of applying predictive models to assist in feedyard post-treatment culling decisions, an economic model was created to estimate net returns. The predictive models described above encompassed the 15-day post-treatment interval; however, the economic model was designed to account for the remainder of the feeding period to capture the long-term economic effects of using a predictive model to make a culling decision. Cost and income calculations were estimated for various processes of production and culling.

The net returns for each individual diagnostic outcome were estimated using unique formulae for each diagnostic outcome at first and second treatment for BRD. Diagnostic

outcomes fell into 1 of 4 categories: true negatives, false positives, true positives, or false negatives. Cattle that truly lived and remained with the group beyond 15 days were either predicted to live by the model (true negatives) or predicted to experience mortality (false positives). Cattle that truly died or were culled within 15 days post-treatment were either predicted to experience mortality by the model (true positives) or predicted to live by the model (false negatives). A summary of each outcome and their respective net return formulae for the BRD1 data are described in Figure 4.3. The BRD2 formulae were similar but included an additional treatment cost in each outcome formula.

Detailed descriptions of the calculations used to estimate the individual elements of the net return equations for each model outcome can be found in Table 4.1. This economic model did not include purchase cost, processing cost, yardage, or feed cost up until the time of treatment as we were interested in determining the differences between model outcomes for net returns following first or second treatment for BRD. Feed costs were adjusted for the true negative and false positive animals based on expected changes in consumption for animals in each category. False positive animals would be shipped early, thus sparing the feedyard this feed cost, so it was included as cost savings in the calculation for net return of this outcome. Treatment costs were applied to negative prognostic model outcomes (true and false negatives) and were based on the cost of generic tulathromycin.¹⁹

A feed ration consisting of 10% alfalfa hay, 30% wet distillers' grain, and 60% corn was developed. Prices for alfalfa, distillers' grains, and corn were as follows⁹: \$170.24/ton, \$46.12/ton, and \$137.5/ ton. These values were used to set a feed price of \$0.0803/lb of feed. The national average fed cattle price was calculated to be \$116.10/cwt.⁸ This was used to determine the sale price of animals that finished the feeding period. The animals predicted by the model to experience mortality in the next 15 days (true positives and false positives) would be shipped early (railed/culled) at a discounted price. Railer price was estimated using the following equation from Horton et al²⁰:

$$74.6\% \times \text{average breaker cull cow price}$$

Cull cow prices were derived from CattleFax data from the states represented in our dataset from 2018 through 2021.⁸ This method of estimating railer prices was recently described by Horton et al.²⁰ This yielded a railer price of \$45.04/cwt. Sale price was determined based on the average out-weight of the lots included in each dataset and calculated using 6-state averages

from CattleFax.⁸ The risk of did not finish ($Risk_{DNF}$) later in the feeding period needed to be accounted for in the economic model for the false positive and true negative model outcomes. The median $Risk_{DNF}$ was calculated using data from Babcock et al²¹ by dividing the median cumulative mortality rate (0.013) by the median days on feed (178 days), resulting in a $Risk_{DNF}$ of 0.007%. The $Risk_{DNF}$ for each dataset was estimated by multiplying the estimated daily risk of mortality by the average days between treatment and end of the feeding period. The respective sale prices for each dataset were multiplied by the $Risk_{DNF}$ later to estimate the cost associated with DNF later (Figure 4.3).

Next, the overall net return for using the predictive model (either at first or second treatment for BRD) was based on the frequency of each outcome and their individual net returns. A status quo was calculated based on no change in action from normal management procedures performed following treatment (cattle remained in pen with no culling but may have experienced mortality within the 15-day post-treatment observational period). The cost associated with the status quo was estimated in order to make a comparison to the economic results from utilizing the predictive model. To do this, a cost for animals that DNF within 15 days of treatment for BRD was determined to be a loss equivalent to the money lost from the potential sale of that animal (sale price) and the money lost on administering a treatment to that animal. The animals that survived also received an estimated cost. This was equivalent to the calculation for a true negative model outcome (the net return from the sale price after the subtraction of feed cost and treatment cost and the adjustment for $Risk_{DNF}$ later in the feeding period). These 2 costs were calculated for both the BRD1 and BRD2 datasets and were used to calculate a status quo cost per head. The status quo was set to zero and compared to the overall net returns of using the predictive models to make a culling decision at first and second treatment.

Each predictive model run produced a scored dataset that included the true outcome, predicted outcome, and the probability of that predicted outcome for each observation. A scored dataset displays the predicted outcome and a measure of certainty of the prediction, usually given as a probability. The decision of scoring a predicted outcome is based on the probability threshold, the default of which is usually 0.5. For example, a predicted outcome of DNF_15 with a 0.7 probability will be scored as a DNF_15 positive (would not finish) by the model, whereas a predicted outcome of DNF_15 with a 0.4 probability will be scored as an animal that will survive. The probability threshold can be increased or decreased to adjust sensitivity and

specificity, thus altering the total numbers of true positives, true negatives, false positives, and false negatives. In this study, 9 probability thresholds from > 0.1 to > 0.9 in 0.1 increments were used for each model. The results for the net return calculation for each prognostic outcome were multiplied by the total number of true positives, true negatives, false positives, and false negatives at each probability threshold for each model (BRD1 and BRD2). The total costs were then divided by the number of observations (head of cattle) in each dataset to return an approximate net return of running the model per head. The difference of these values from the status quo per head calculations were then determined to understand the difference in net return between using the predictive model to make an early culling decision and the status quo for the feedyard.

Results

Following data transformation and variable creation, the final BRD1 dataset consisted of 133,249 animals, with a 6.1% prevalence of DNF_15. The final BRD2 dataset consisted of 27,726 animals, with a 14.0% prevalence of DNF_15. For the BRD1 dataset, the top 30 variables were kept, and the top 50 variables were kept for the BRD2 dataset following ranking from the chi-squared filtering approach. Appendix Tables B.3 and B.4 indicate the features selected for inclusion in both datasets along with the associated chi-square correlation coefficients.

Evaluation of models

Two separate boosted decision tree models were trained, tested, and validated to predict DNF_15 based on information known at first (BRD1) and second treatment (BRD2) for BRD. Table 4.2 provides a summary of performance metrics for these models. Overall, both models demonstrated high specificities and high negative predictive values but low positive predictive values and low sensitivities. The AUCs for both models were between 0.7 and 0.8, demonstrating moderate predictive power.

Economic analysis results

The results from the economic model illustrated different net returns for each of the outcomes as classified by model prediction and actual cattle outcome. The net returns for true positives were \$393.75 and \$388.58/head for the BRD1 and BRD2, respectively. True negative net returns were profits of \$1,377.12 and \$1,371.11/head for BRD1 and BRD2, respectively. False positives resulted in the largest losses, with net returns of $-\$1,000.94$ and $-\$998.94$ / head

for BRD1 and BRD2, respectively. False negatives were also losses, with $-\$411.32/\text{head}$ for BRD1 and $-\$404.99/\text{head}$ for BRD2. The results for status quo economic estimates for BRD1 and BRD2 were $\$1,186.36$ and $\$933.32$, respectively.

Using the equations for net return described above, the difference from the status quo was calculated for each model (BRD1 and BRD2) across all probability thresholds, from > 0.1 to > 0.9 . Figure 4.4 graphs the economic results from the BRD1 and BRD2 models. In general, as the probability threshold increased, the net return per head increased as well. For the BRD1 model, net return peaked at the threshold of > 0.6 and then decreased slightly. In the BRD2 model, net return continued to increase up to the probability threshold of > 0.9 .

Discussion

With the models constructed in this study, we attempted to predict mortality within the next 15 days of both first and second treatment for BRD. Our models included variables previously identified to be associated with mortality due to BRD, such as weight and body temperature at the time of treatment, quarter of arrival, and days on feed. Often the variables collected for these kinds of studies are involved in 2- and potentially 3-way interactions.⁶ These complex interactions may be difficult to interpret and thus make machine learning an attractive application to predict mortality. Additionally, predictive models are well-suited to handling large amounts of data, which is often the case when conducting a retrospective analysis of feedyard data.²² Predictive models are able to handle large amounts of data, and algorithms exist to handle collinearity, mitigating another concern with using traditional statistical approaches.²³ The algorithm used in this study was a boosted decision tree. Boosting allows for correction of misclassification and has been shown to be superior to other algorithms, such as neural networks, when used on datasets with large numbers of variables.¹⁴ Decision trees are known to be applicable to a wide range of problems in addition to their ability to handle data with high dimensionality.²⁴

Based on our AUC values of 0.81 for the BRD1 model and 0.78 for the BRD2 model, we determined the models investigated in this study to have moderate accuracy. Recent studies^{2,3,21,25} have investigated the prediction of BRD morbidity or case fatality using machine learning. Feuz et al²⁵ reports higher sensitivities for the algorithms tested in their study but do not report AUC values. Amrine et al² report a maximum AUC of 0.62, which is considerably lower than our reported AUC values. Babcock et al²¹ reports as high as 91% of lots correctly classified

for cumulative morbidity risk; however, they do not report AUC values, making their models difficult to directly compare to ours. Rojas et al³ constructed models to determine BRD morbidity on a pen-level basis. The highest AUC reported in that study was 0.79. Overall, our models perform similarly to other published classification model studies; however, our models demonstrate lower sensitivities, likely due to the imbalanced nature of the outcome class. Additionally, direct comparisons are difficult to make as the aforementioned studies made predictions on a lot-level basis, whereas the current study investigated predictions on an individual basis. Predictions made on an individual basis are likely subject to higher variability and thus are less accurate than those made on a lot level.

We utilized a pooled dataset consisting of feedyard data from multiple operations in different geographic locations over several years. This allowed for a robust but highly variable dataset. The predictive models built from this dataset are therefore quite general and may result in poorer performance in comparison to a predictive model built using a specific feedyard's dataset. Previous research has demonstrated that models built using individual feedyard data can perform either worse or better in comparison to models built from a pooled dataset.^{2,13} The reason for this finding likely lies in the data quality differences between feedyards that wash out in the pooled dataset. However, consistent between the aforementioned studies is the finding that the variability in model performance decreases when creating models for individual feedyards.

Our economic analysis found that increasing the specificity of the models increased the net return per head compared to the status quo, with an even further increased net return found with the BRD2 model. A couple of papers have performed similar economic analyses using classification models on treated cattle in a feedlot. In 2021, Feuz et al²⁵ reported an average increase of \$14.01 net return per head compared to a simulated standard culling procedure when using the logistic regression model described in their study. They simulated the net return per head using various predictive model algorithms to model mortality in feedlot cattle removed from their pen for any health-related reason. In 2022, Feuz et al²⁶ conducted a similar study with more feedyards included in the dataset. This time they reported an average increase in net return of \$6.31/head. These are far lower than our economic results. Feuz et al investigated the prediction of mortality to the end of the feeding period in both papers,^{25,26} whereas we focused on the period of 15 days post treatment for BRD. Other work has shown a decrease in predictive ability over longer periods of time.¹³ Finally, Feuz et al, in both papers,^{25,26} included calves

pulled and treated multiple times for the same or different disease process as the first pull. This means that each health incident and prediction in an individual calf was considered independent from any others, allowing for that animal to be newly classified unless it had been previously categorized as DNF.

Despite the differences in net returns, their findings support that increased specificity results in higher net returns, indicating that false positives cost more than false negatives. This same finding is supported by other studies that have investigated changes in net returns associated with changing diagnostic test characteristics.²⁷ Thus, when using economic analysis to consider predictive models for implementation, specificity (the ability to determine the true negatives) will drive the selection. This consequently sacrifices sensitivity, resulting in an increased number of false negatives. In situations where detecting truly diseased individuals outweighs economic advantages, consideration should be given to a model with an increased sensitivity.

The models designed in this study have never been implemented to make culling decisions in real time, and thus the results reported are simply expectations. This paper investigated the economic implications of making a culling decision assuming the animal in question would be shipped the same day the decision was made. This is unlikely to happen in actuality as animals that are culled early are not shipped daily and often must remain at the feedyard in order to wait out their withdrawal period. Therefore, we are likely overestimating the net returns in our economic analysis as we did not account for continued loss that would inevitably occur as animals wait to get shipped. The outcome DNF_15 incorporated both animals lost to mortality and those that were culled early. There is no guarantee that the animals that were culled in the original dataset would have died before the end of the feeding phase. However, they were included in our outcome because this population of cattle serves as a source of loss comparable to mortalities at the feedyard.

Additionally, the predictive models were built on data limited to the feedyards we had access to. The formulae created for the economic analysis were based largely on estimates of prices that are constantly changing. Although beyond the scope of the current study, a sensitivity analysis or stochastic model of the economic analysis would be useful in capturing the impact the volatile market has on outcome. Further research and additional data are needed to create more robust analyses.

In conclusion, the intersection between machine learning and livestock production has occurred in recent years. More studies are exploring the application of predictive models to health and productivity challenges in various livestock systems. With this paper, we hope to contribute to the growing body of knowledge on the use of predictive models to make management decisions on US feedyards by creating models that determined whether or not an animal would die in the next 15 days after first and second treatment for BRD. Additionally, we sought to investigate the estimated net returns of using a predictive model to make a culling decision in comparison to the status quo. The boosted decision tree models we constructed performed with moderate accuracy, and we found that by increasing the specificity of those models we were able to generate high net returns in comparison to a feedyard's status quo. These models and the economic analysis performed in this study cannot capture the complexity and volatility of cattle health and the market. Therefore, analysis with a stochastic predictive and economic model could help more accurately represent what the expected economic benefit would be when using machine learning to implement an early culling strategy at a feedyard.

References

1. USDA APHIS National Animal Health Monitoring System. Feedlot 2011 Part IV: Health and Health Management on U.S. Feedlots with a Capacity of 1,000 or More Head. Available at: https://www.aphis.usda.gov/sites/default/files/feed11_dr_partiv.pdf.
2. Amrine DE, White BJ, Larson RL. Comparison of classification algorithms to predict outcomes of feedlot cattle identified and treated for bovine respiratory disease. *Comput Electron Agric* 2009;105:9–19.
3. Rojas HA, White BJ, Amrine DE, et al. Predicting Bovine Respiratory Disease Risk in Feedlot Cattle in the First 45 Days Post Arrival. *Pathogens* 2022;11:442. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9029152/>.
4. Wisnieski L, Norby B, Pierce SJ, et al. Predictive models for early lactation disease in transition dairy cattle at dry-off. *Prev Vet Med* 2019;163:68–78.
5. Wisnieski L, Amrine DE, Renter DG. Predictive modeling of bovine respiratory disease outcomes in feedlot cattle: a narrative review. *Livest Sci* 2021;251.
6. Avra TD, Abell KM, Shane DD, et al. A retrospective analysis of risk factors associated with bovine respiratory disease treatment failure in feedlot cattle. *J Anim Sci* 2017;95:1521–1527.

7. Anon. Agricultural Marketing Service. US Department of Agriculture 2023. Available at: <https://www.ams.usda.gov/>.
8. Anon. CattleFax Six State Fed Steer Price. CattleFax 2022. Available at: <https://www.cattlefax.com/#!/data/cattle/other-prices/six-state-fed-steer/>.
9. Anon. Focus on Feedlots Monthly Reports. Animal Sciences and Industry - Kansas State University. Available at: <https://www.asi.k-state.edu/about/newsletters/focus-on-feedlots/monthly-reports.html>. Accessed June 1, 2022.
10. Johanns A. Prices and Profitability Models. Agricultural Marketing Resource Center 2022. Available at: <https://www.agmrc.org/renewable-energy/prices-and-profitability-models>.
11. Anon. R: a language and environment for statistical computing. R Foundation for Statistical Computing Available at: <https://www.R-project.org/>. Accessed May 28, 2021.
12. Anon. Microsoft Azure Machine Learning. Microsoft Available at: <https://azure.microsoft.com/en-us/services/machine-learning/>. Accessed May 24, 2021.
13. Heinen L, White B, Amrine D, et al. Evaluation of predictive models to determine final outcome for feedlot cattle based on information available at first treatment for bovine respiratory disease. *Am J Vet Res* 2023;84.
14. Roe BP, Yang H, Zhu J, et al. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nucl Instrum, Methods Phys Res A: Accel Spectrom Detect Assoc Equip* 2005;543:577–584.
15. likebupt, PeterCLu, v-chmccl. Filter Based Feature Selection. Microsoft Docs Azure Machine Learning 2021. Available at: <https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/filter-based-feature-selection?view=azureml-api-2>.
16. likebupt, Bowen-Guo-Microsoft, PeterCLu, et al. Tune Model Hyperparameters. Microsoft Docs Azure Machine Learning 2021. Available at: <https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/tune-model-hyperparameters?view=azureml-api-2>.
17. Anon. Two-Class Boosted Decision Tree component. Microsoft Docs Azure Machine Learning 2021. Available at: <https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-boosted-decision-tree?view=azureml-api-2>.

18. Anon. Cross Validate Model. Microsoft Docs Azure Machine Learning 2021. Available at: <https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/cross-validate-model?view=azureml-api-2>.
19. Horton L, Depenbusch B, Pendell D, et al. Description of feedlot animals culled for slaughter, revenue received, and associations with reported US beef market prices. *The Bovine Practitioner* 2021;55:65–77. Available at: <https://bovine-ojs-tamu.tdl.org/bovine/article/view/8202>.
20. Anon. Macrosyn Tulathromycin by Bimeda. ValleyVet 2023. Available at: https://www.valleyvet.com/ct_detail.html?pgguid=a1dd674f-99e9-41e7-bd16-5cc0fc3e0c8b. Accessed June 1, 2022.
21. Babcock A, White B, Renter D, et al. Predicting cumulative risk of bovine respiratory disease complex using feedlot arrival data and daily morbidity and mortality counts. *CJVR* 2013;77:33–44.
22. Hudson C, Kaler J, Down P. Using big data in cattle practice. In *Pract* 2018;40:368–423.
23. Schooling C, Jones H. Clarifying question about “risk factors”: predictors versus explanation. *Emerg Themes Epidemiol* 2018;15.
24. Swain P, Hauska H. The decision tree classifier: design and potential. *IEEE Trans Geosci Remote Sens* 1977;15:142–147.
25. Feuz R, Feuz K, Johnson M, et al. Improving Feedlot Profitability Using Operational Data in Mortality Prediction Modeling. *J Agric Econ* 2021;46:242–255.
26. Feuz R, Feuz K, Gradner J, et al. Scalability and robustness of feed yard mortality prediction modeling to improve profitability. *Agric Resour Econom Rev* 2022:1–23.
27. Theurer M, White B, Larson R, et al. A stochastic model to determine the economic value of changing diagnostic test characteristics for identification of cattle for treatment of bovine respiratory disease. *JAS* 2015;93:1398–1410.

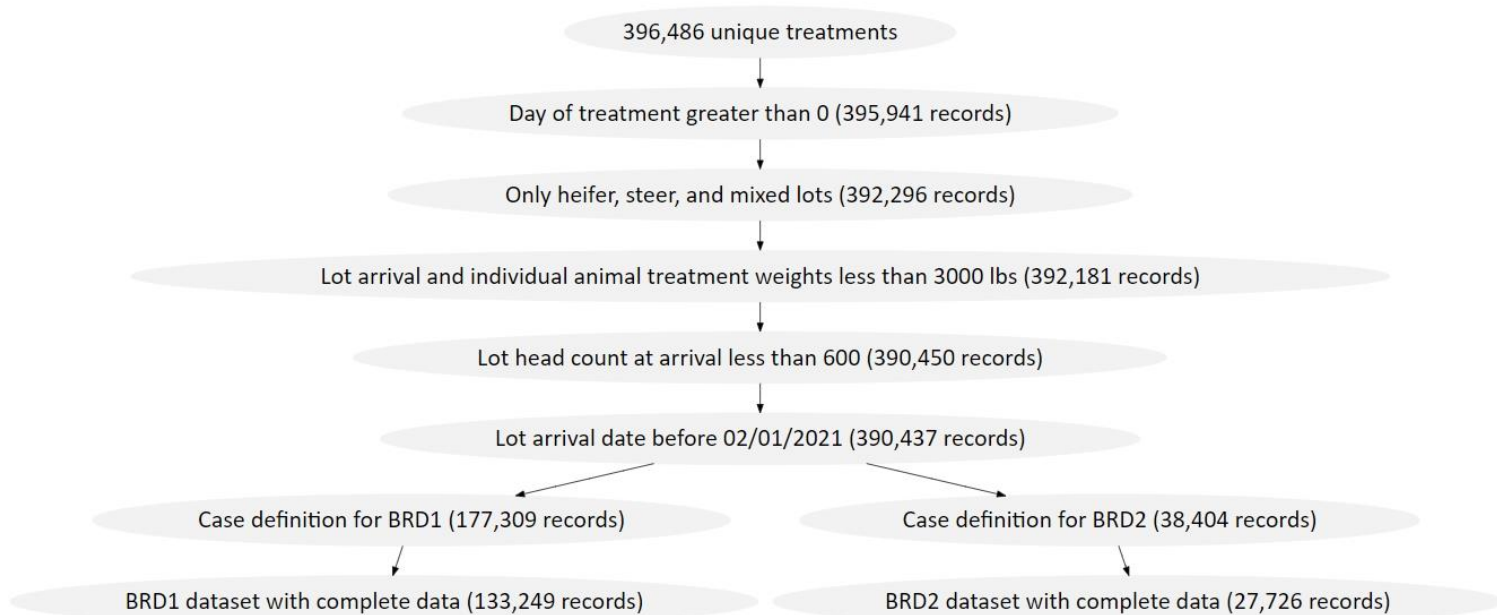
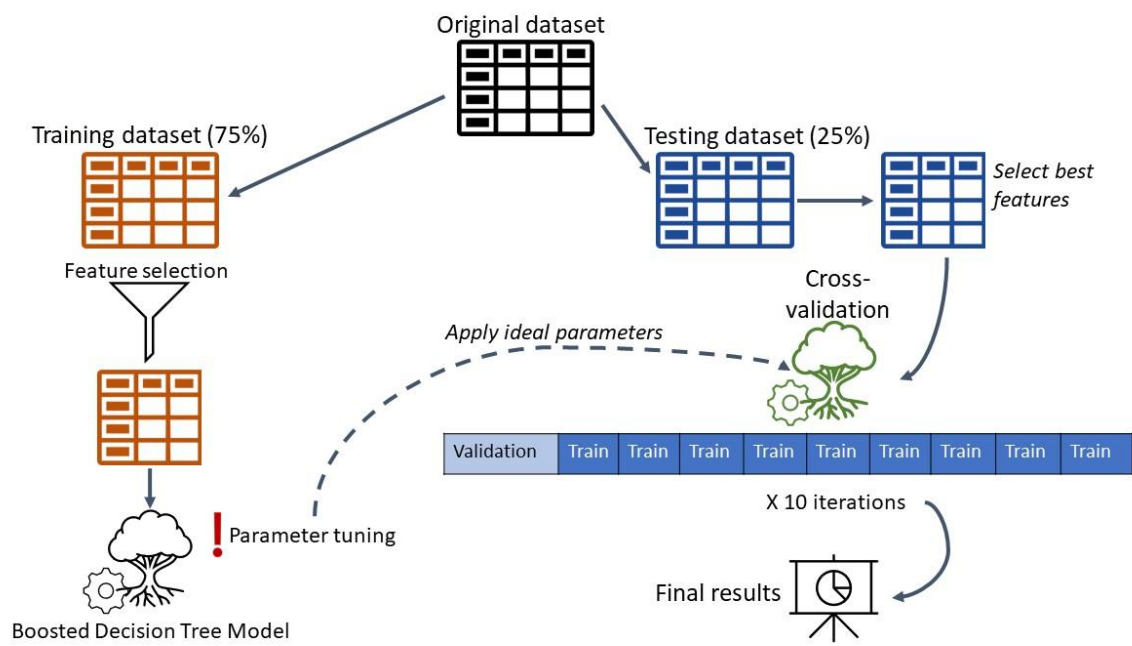


Figure 4.1. Flowchart summarizing exclusion criteria and dataflow. BRD1 refers to first treatment for BRD. BRD2 refers to second treatment for BRD.

2162



2163

2164

2165

2166

Figure 4.2. Illustration describing the steps followed to train, test, and validate the boosted decision tree models. These steps were performed for each dataset (BRD1 and BRD2).

2167

Model Predicted Result	True Result		
			
		The animal was correctly railed early (TP) RP	The animal was incorrectly railed early (FP) $RP + FC_{TN} + Risk_{DNF} - SP$
		The animal was incorrectly not railed early (FN) $-TC - RP$	The animal was correctly not railed early (TN) $SP - FC_{TN} - TC - Risk_{DNF}$

2168

2169

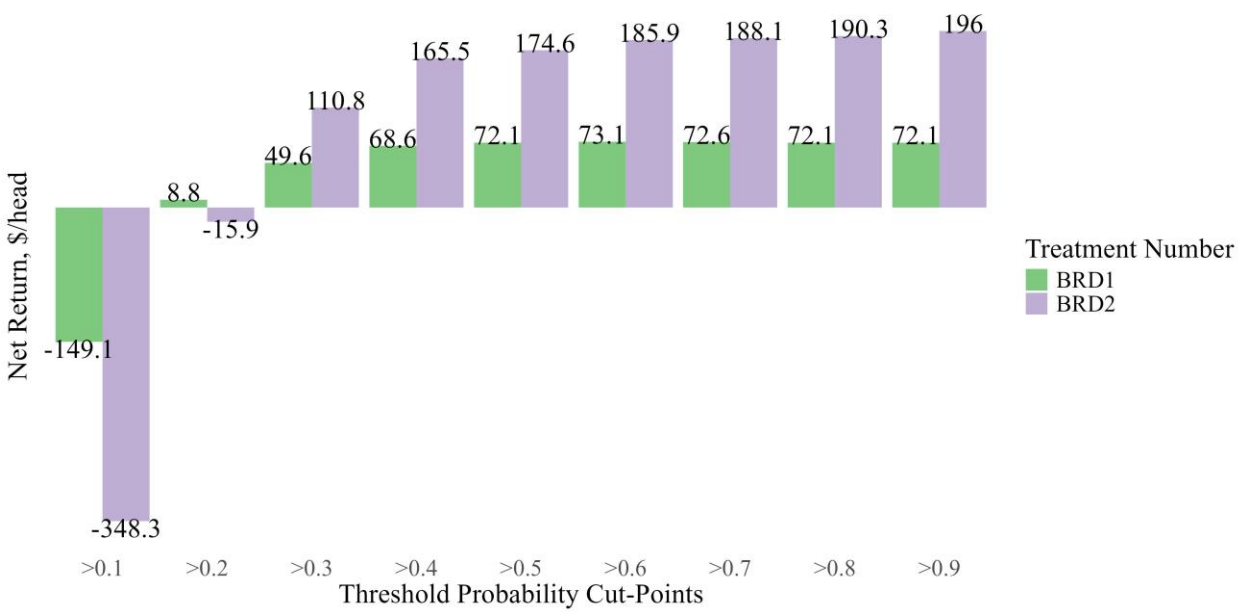
2170

2171

2172

Figure 4.3. Confidence matrix describing the costs or profits associated with each possible model outcome for the BRD1 model. RP = railer price, FC_{TN} = adjusted feed cost for true negative, SP = sale price, TC = treatment cost, Risk_{DNF} = risk of DNF adjustment

2173



2174

2175 **Figure 4.4.** Difference in net returns using the model compared to status quo for BRD1 and
2176 BRD2

2177

Table 4.1. Summary of calculations for estimation of cost and prices associated with model results.

Cost/Income	Calculation	Result (\$)
Railer price (<i>RP</i>)	Average weight at treatment * estimated railer price per cwt	BRD1 = (817 lbs) * (\$45.04/100 lbs) = \$367.78 BRD2 = (863 lbs) * (\$45.04/100 lbs) = \$388.58
Sale price (<i>SP</i>)	Average out weight * national average sale price per cwt	BRD1 = (1,499 lbs) * (\$116.10/100 lbs) = \$1,739.92 BRD2 = (1,490 lbs) * (\$116.10/100 lbs) = \$1,730.26
Adjusted feed cost for true negative (<i>FC_{TN}</i>)	(Average feed consumed by lot/average head in the lot) * average cost of feed per pound	BRD1 = (613,544 lbs/151 head) * (\$0.0803/lb) = \$326.87 BRD2 = (644,290 lbs/160 head) * (\$0.0803/lb) = \$323.36
Cost of first BRD treatment (<i>TC</i>)	Average weight at first BRD treatment * cost of tulathromycin per cwt	(874 lbs) * (\$2.01/100 lbs) = \$17.57
Cost of second BRD treatment (<i>TC</i>)	Average weight at second BRD treatment * cost of tulathromycin per cwt	(863 lbs) * (\$2.01/100 lbs) = \$16.41
Risk of DNF adjustment (<i>Risk_{DNF}</i>)	[(median cumulative mortality / median final days on feed) * average days between treatment and end of feeding period] * sale price	BRD1 = [(0.013/178 days) * 145 days] * \$1,739.92 = \$18.36 BRD2 = [(0.013/178 days) * 153 days] * \$1,730.26 = \$19.38

Table 4.2. Summary of model performance predicting DNF_15 at first and second treatment for BRD.

Metric	BRD1 Model	BRD2 Model
AUC-ROC	0.811	0.783
True positives	8	14
True negatives	3100	586
False positives	10	11
False negatives	213	82
Sensitivity	3.6%	14.6%
Specificity	99.7%	98.2%
PPV	0.44	0.56
NPV	0.94	0.88
Accuracy	93.3%	86.6%

Chapter 5 - Predictive models to determine best strategy for metaphylaxis application in cattle at arrival to a feedyard

L. Heinen¹, P. A. Lancaster, R.L. Larson¹, D.L. Pendell², D. Kopp¹, B.J. White¹

¹Beef Cattle Institute, Kansas State University, Manhattan, KS 66506, USA

²Department of Agriculture Economics, Kansas State University, Manhattan, KS 66502, USA

This research is published in the Journal of Preventive Veterinary Medicine and is reproduced with permission from publisher as formatted in referenced publication. Lilli Heinen, Phillip A. Lancaster, Robert L. Larson, Dustin L. Pendell, Dannell Kopp, Brad J. White, Predictive models to determine best strategy for metaphylaxis application in cattle at arrival to a feedyard, Preventive Veterinary Medicine Volume 242, May 2025, <https://doi.org/10.1016/j.prevetmed.2025.106572>.

Abstract

Metaphylaxis, the application of an antimicrobial to a cohort of cattle at arrival to a feedyard, is an important bovine respiratory disease (BRD) control strategy for certain cattle populations. Predictive modeling techniques could be used to assist in determining which cohorts should receive metaphylaxis based on a desired economic outcome instead of subjectively. The study objective was to evaluate predictive models trained with cattle demographic variables to determine which cohorts should receive metaphylaxis based on an economic evaluation of highest net returns and to elucidate the benefit to model performance with the addition of origin and external economic variables. Data from 16,368 cattle cohorts were used to build four predictive models: boosted decision tree, logistic regression, neural network, and random forest. Area under the Receiver Operating Characteristics curve (AUC-ROC) was used to evaluate model performance. The same algorithms were used to compare adding origin and external economic data to the baseline models. Overall, model performance was high with AUC-ROC values ranging from 0.80 to 0.93 in the baseline models. Adding external economic variables such as commodity futures prices increased performance (AUC-ROC=0.92 – 0.94). Adding origin data, such as city and state, resulted in poorer performance (AUC-ROC=0.79 – 0.89). The combination of external economic data and origin resulted in intermediate AUC-ROC values

(AUC-ROC=0.87 – 0.91). The study demonstrated that predictive models can be used successfully to select an optimal metaphylaxis strategy as determined by economic evaluation for cattle arriving at the feedyard.

Keywords: machine learning, antibiotic, bovine respiratory disease, feedlot

Introduction

Bovine respiratory disease (BRD) is a complex disease condition of cattle affecting the respiratory tract and is associated with a myriad of host, environmental, and microbial factors (Fulton, 2009; O'Neill et al., 2014). Environmental risk factors are well-researched and include long-distance transportation, commingling, and others (Cernicchiaro et al., 2011; Cernicchiaro et al., 2012; Sanderson et al., 2008). Economic loss due to BRD mortality was estimated to be about \$643 million in 2010 (USDA NASS, 2011). Perhaps more important is loss attributed to BRD morbidity which includes negative effects on average daily gain (ADG) and feed conversion in morbid animals, treatment and labor costs, and lower carcass grades at slaughter (Gardner et al., 1999; Irsik et al., 2006; Schneider et al., 2009). To mitigate these losses, feedyards implement antibiotic metaphylaxis, the application of an antibiotic to an entire group of cattle at or near the time of feedyard arrival in order to decrease BRD incidence (Ives and Richeson, 2015). The decision to apply metaphylaxis to a group of cattle at arrival is often based on risk classification informed by the various group-level risk factors including, but not limited to, weight as a proxy for age, source (ranch or auction), distance transported, presence of commingling, and pre-arrival conditioning such as vaccination, dehorning, and castration procedures (Ives and Richeson, 2015; Nickell and White, 2010; Seeger et al., 2008).

Implementation of metaphylaxis has been shown to decrease both BRD morbidity and mortality (Baptiste and Kyvsgaard, 2017; Horton et al., 2023; Tennant et al., 2014). Additionally, positive net returns have been demonstrated when administering metaphylaxis at arrival versus no metaphylaxis although this effect is modified by antibiotic class (Dennis et al., 2020). Despite the evidence demonstrating the animal health and economic benefits of metaphylaxis administration, calls to reduce antimicrobial use in the livestock industry from organizations like the European Union and World Health Organization have led to efforts to reduce this practice (European Union, 2015; World Health Organization, 2017). In addition to researching antibiotic alternatives, industry interest has also begun to incorporate technology which would help specifically target antibiotic application. A potential tool to encourage strategic antimicrobial use

could be a predictive model trained to determine which cohorts of cattle arriving at the feedyard will benefit from metaphylaxis based on estimated economic returns. Using economic returns as a metric allows for the incorporation of several outcome possibilities including morbidity, mortality, and finishing to slaughter.

Predictive models are algorithms with statistical backbones trained to estimate a future outcome based on retrospective data. Depending on the algorithm, models can analyze complex relationships between input variables. Predictive models are trained on a dataset with input variables and a known outcome. Following training, the model can be introduced to a dataset where the outcome is unknown and make a prediction with some margin of error. As artificial intelligence technology advances and human and computational machines become further connected via the internet, more industries are utilizing predictive modeling techniques to expand production and improve consumer satisfaction (The Pecan Team, 2023). However, predictive models cannot consider all aspects of a decision-making process like a human can and thus cannot fully replace all human-decision making.

The popularity of machine learning and predictive models has reached production agriculture. The swine and poultry industries utilize predictive models to solve both animal health and productivity issues (Astill et al., 2018; Lee et al., 2019; Magalhaes et al., 2023; Omomule et al., 2020). Within the cattle industry, much research exists in the dairy sector investigating application of predictive models to anticipate calving, thermal stress, and post-partum diseases (Rutten et al., 2017; Vergara et al., 2014; Wang et al., 2018). Similar research exists in the beef sector with focus on carcass growth, heat stress, and BRD (Heinen et al., 2023; Heinen et al., 2024; Huuskonen et al., 2013; Vieira de Sousa et al., 2018; Wisnieski et al., 2021). Utilizing predictive models to target antibiotic treatment and metaphylactic application is an important area for development in the beef cattle industry. In addition to reducing the cost of drug application, the strain on labor demand could be eased and public perception could be improved via the reduction of antibiotic use (Drouillard, 2018). Thus, the objective of the current study is to evaluate the ability of predictive models to determine if on-arrival metaphylaxis application is optimal for groups of cattle given cattle demographic data. Additionally, a secondary objective is to compare performance of models given demographic data alone with models given origin and external economic data. We hypothesize that models built with origin

and economic data will perform with greater accuracy than models built with cattle demographic data alone.

Materials and methods

Ethical statement

Institutional Animal Care and Use Committee (IACUC) approval was not required as all data used for analysis were obtained from existing operation databases and no procedures specific for this study were performed on cattle.

Data sources

Feedyard data

Retrospective data from 20 feedyards were collected over an eight-year period (2015-2023). These data encompass cohort-level information where a cohort is defined as a group of cattle purchased, managed and sold together, but not necessarily housed together for the entirety of the feeding period. Cohort-level information in this study included head count at arrival, arrival and exit dates, days on feed, cohort average entry and exit weights, sex, morbidity and mortality counts, and origin information when available. Origin data incorporated: city, state, latitude, and longitude. Morbidity and mortality counts included those animals diagnosed with BRD at the time of treatment or death. Diagnosis of BRD was made by trained personnel at each feedyard and based on typical clinical signs such as elevated rectal temperature, elevated respiratory rate, increased respiratory effort, anorexia/lethargy, cough, and ocular or nasal discharge. Necropsy to confirm cause of death due to BRD was not consistently performed among all feedyards included in the dataset. In addition to including mortality counts, cohort data also included a “railer count” which captured the number of animals in each cohort that were culled from the group to be shipped early at a discount. An animal is typically culled in this manner due to failure to respond to treatment. Finally, a variable called “meta_yard” described whether or not each cohort received metaphylaxis at arrival. Cohorts were considered to have received metaphylaxis if more than 50% of the cohort received an antibiotic at arrival. For clarity, cohorts that originally received metaphylaxis will be referred to from this point forward as META cohorts whereas, cohorts that did not originally receive metaphylaxis will be called NO META cohorts.

Records for 22,789 cohorts of cattle were collected. Only heifer and steer unisex cohorts were kept, all mixed sex cohorts or cohorts without sex information were removed ($n = 437$ and $n = 1,142$, respectively). To ensure adequate and accurate data, cohorts with less than 50 or greater than 250 days on feed (DOF) and cohorts with less than 50 head received were removed ($n = 1,047$ and $n = 275$, respectively). Cohorts with 100% morbidity were eliminated ($n = 3$) as well as cohorts with average daily gain (ADG) calculations of over 6 lbs (2.7 kg) ($n = 21$). To eliminate cohorts with data entry mistakes, groups with average exit weights lower than average arrival weights ($n = 43$) and cohorts with less than 0 head out ($n = 4$) were removed. These criteria is similar to criteria used in similar studies (Heinen et al., 2023, Heinen et al., 2024). Due to adjusted ADG calculations for the economic analysis, cohorts with BRD death counts higher than BRD treatment counts ($n = 1,974$), and cohorts with BRD treatment counts greater than head received ($n = 101$) were removed as these cohorts would result in a nonsensical values for adjusted ADG (Formula 4). Cohorts with average arrival weights below 300 lbs (136.1 kg) or greater than 1,100 lbs (499.0 kg) were removed due to lack of price availability outside this range. Additionally, cohorts with exit weights below 1,000 (453.6 kg) and over 1,800 (816.5 kg) lbs were removed for the same reason ($n = 1,347$). Finally, cohorts with any missing data were removed ($n = 27$). This left a total of 16,368 cohorts for analysis.

Several variables were estimated or calculated for the subsequent economic analysis using the existing feedyard data. These variables and economic analysis are similar to a study performed by Kopp et al. (2024). No variable existed to capture the number of mortalities due to causes other than BRD so one was calculated by subtracting the BRD mortalities and railed animals from the difference between head shipped to harvest and head count at arrival. ADG was calculated by dividing the difference between average exit weight and average arrival weight by the total DOF.

Feed delivery data were not available for some feedyards in the dataset. Dry matter intake (DMI) per head was calculated using a linear regression equation based on a different feedyard's dataset and then applied to the calculated ADG from the current study's average arrival weight (Formula 1). The dataset to build the DMI estimate formula did not contain the same cohorts as those included in this analysis, but cattle type, management, and nutrition were similar between the two datasets. Three models were created to estimate DMI and tested against actual intake information. Each model included ADG and either average arrival weight, average exit weight,

or average weight throughout the feeding period as the additional variable. The dataset consisted of 6,289 individual cohorts of cattle with 70% (n = 4,402) used for training and 30% (n = 1,887) set aside for validating. Each of the three models were tested against the actual intake data and the model with the highest correlation to the actual DMI was selected to calculate DMI for the present dataset. The selected model's DMI calculation had a correlation coefficient of 0.91 when compared to the actual DMI provided by the feedyard.

$$DMI \text{ per head} = 4.0993 + (2.5944 * ADG) + (0.0101 * avg. arrival weight)$$

Formula 1. A formula for dry matter intake (DMI) per head from a linear regression equation where 4.0993 represents the intercept, average daily gain (ADG) is multiplied by a coefficient of 2.5944 and average arrival weight is multiplied by a coefficient of 0.0101. Correlation coefficient = 0.91 with actual DMI

Head-days were used to calculate expenses that are typically accrued on a daily basis by the number of cattle present during each DOF. The formula accounts for the days on feed spent by each animal that arrives at the feedyard and includes estimates for the days on feed spent by animals that die during the feeding phase (Formula 2). BRD mortalities were estimated to occur at 50 DOF and all other mortalities were estimated to occur half-way through the feeding phase. A case fatality risk (CFR) was calculated as the BRD mortality count divided by the BRD morbidity count for each lot.

$$Head \text{ Days} = (BRD \text{ mortalities} * 50) + (Other \text{ mortalities} * 0.5 * DOF) + (exit \text{ head count} * DOF)$$

Formula 2. A formula for head days. BRD mortalities are estimated to occur at 50 days on feed, and other mortalities are estimated to occur halfway to the end of their feeding phase. All other animals in the cohort live to the end of the feeding phase. BRD = bovine respiratory disease, DOF = days on feed.

Cattle price data

Feeder cattle prices for the same eight-year period (2015-2023) were acquired from the United States Department of Agriculture's Agriculture Marketing Service Livestock, Poultry & Grain Market News (USDA AMS Livestock, Poultry, & Grain Market News, 2015a) and were given as a price per hundredweight and organized by sex, hundredweight category, month, and

year. The purchase price for each cohort was based on average entry weight and feeder cattle price relative to corresponding weight category in the month and year that cohort of cattle entered the feedyard. Fat cattle price was obtained from the same database with a similar structure (USDA AMS Livestock, Poultry & Grain Market News, 2015a). Sale price for each cohort was calculated based on average exit weight and fat cattle price relative to the corresponding weight category in the month and year that cohort of cattle exited the feedyard.

Feed data and cost calculation

The feed ration price per pound was calculated using historical feedstuffs data for the same eight-year period and a hypothetical diet based on a survey of consulting feedyard nutrition recommendations (Samuelson et al., 2016) was created. Historical alfalfa hay prices were collected from Kansas State University's Focus on Feedlots Monthly Report (K-State Research and Extension, 2015). These data provided average alfalfa hay prices on a monthly basis for each year in the dataset. USDA's Agriculture Marketing Service's database was used to collect monthly prices for wet distillers' grains and corn for the years 2015-2023 (2015b). All feedstuff prices were acquired for the state of Kansas as most feedyard in the dataset were in Kansas or in the greater High Plains region. Prices were matched to each cohort's arrival month and year. A supplement cost of \$700 per ton was also included and this number was based on expert opinion and internal data.

A hypothetical ration consisting of 10% alfalfa hay, 67% corn, 20% wet distillers' grains and 3% supplement was created based on a study by Samuelson et al. (2016) who used conducted a survey of feedyard consulting nutritionists' recommendations. Using this hypothetical ration, feedstuff prices mentioned above, and each diet component's dry matter as a proportion a formula was constructed to calculate the total feed cost for each cohort (Formula 3). Each feedstuff price was converted to dollars per ton before inclusion in the calculation for feed cost per pound.

Feed cost per lb

$$= \frac{\frac{0.67 * \text{corn price}}{0.88} + \frac{0.20 * \text{wdg price}}{0.35} + \frac{0.10 * \text{hay price}}{0.85} + \frac{0.03 * \text{supp price}}{0.93}}{2000}$$

Formula 3. A formula to calculate feed cost on a per pound basis. A diet component's price is expressed in dollars per ton (as fed) and multiplied by the ingredient's proportion in the diet to then divided by that diet component's dry matter expressed as a proportion. Following the

2397 addition of all terms, the sum is divided by 2000 to obtain the feed cost in dollars per pound.
2398 WDG = wet distillers' grains, hay = alfalfa, supp = supplement.

2399 **Treatment cost**

2400 The cost of a dose of metaphylaxis and the cost of treatment were determined by the cost
2401 of generic tulathromycin at the time of the economic analysis and was measured in dollars per
2402 hundredweight (Macrosyn, dd). The total cost of metaphylaxis for a cohort was determined by
2403 the average arrival weight divided by 100 and multiplied by the cost of tulathromycin in
2404 \$/hundredweight, and the total head count at arrival. The total cost of treatment for a cohort was
2405 determined by the average arrival weight divided by 100 and multiplied by the cost of
2406 tulathromycin per hundredweight and BRD morbidity count.

2407 **Economic analysis to calculate returns**

2408 **Economic analysis for existing metaphylaxis status**

2409 Each cohort's total expenses and revenue were calculated to simulate net returns based on
2410 existing metaphylaxis status captured in the variable, meta_yard. The calculation of sale price
2411 was described in the Cattle Price section above and this value was the only source of revenue for
2412 each cohort. No income from railer cattle, or cattle shipped early at a discount, was included.
2413 Total expenses were calculated using purchase price, cost of processing, cost of treatment,
2414 yardage, feed costs, and cost of metaphylactic therapy for the META cohorts. Purchase price was
2415 described in the Cattle Price section. The cost of processing was set at \$28.60 per head based on
2416 internal data. The per head processing cost was multiplied by the total head count entering the
2417 feedyard. Treatment costs contributed to the total expenses and included a chute charge per head
2418 treated of \$1.50 based on internal data added to the antimicrobial treatment cost described in the
2419 previous section. Total cohort treatment cost was calculated by summing the cost of treatments
2420 within each cohort over the entire period. Yardage was charged at a rate of \$0.35 per head per
2421 day. Head-days were multiplied by the yardage charge to calculate the total yardage for each
2422 cohort. The cost of feed was calculated by multiplying the DMI per head calculated using
2423 Formula 1 by head-days calculated using Formula 2 and cost of feed per pound using Formula 3.
2424 For META cohorts, the cost of metaphylactic treatment was included in the total expenses. There
2425 was no charge for metaphylactic treatment included in the total expenses for the NO META
2426 groups. Net returns (total revenue – total expenses) were calculated for each cohort.

Economic simulation to determine returns based on opposite metaphylaxis status

Economic returns for each cohort were calculated twice: once to represent the actual performance of the cohort given the original decision of META or NO META at the feedyard, and once more to simulate the cohort's performance if metaphylaxis status at arrival was opposite from reality. For example, a META cohort would be modeled as a NO META cohort and vice versa. This was done by first determining the efficacy of metaphylaxis with tulathromycin using a systematic review and mixed treatment meta-analysis as previously reported by Kopp et al. (2024). This informed a metaphylactic efficacy odds ratio of 0.35 selected based on research from Abell et al. (2017).

For NO META cohorts, a reduced morbidity risk was calculated with this odds ratio. The original CFR from NO META cohorts was then multiplied by the new morbidity risk to determine a new number of BRD mortalities. Mortalities due to other causes were assumed to remain the same. No adjustments were made for railed animals. Subsequent adjustments in head days, ADG, income, and costs were made. A new ADG for healthy cattle was calculated using the original ADG, total arrival and exit counts, BRD morbidity counts, and mortality counts (Formula 4). This healthy ADG was calculated using the original ADG and an adjustment factor of 5% decrease in weight gain in cattle treated for BRD (Schneider et al., 2009). A new sick ADG was calculated by multiplying the healthy ADG by the ADG adjustment factor. The new healthy and sick ADGs were used to adjust exit weight of the cohort which influenced income for each metaphylaxis modeled cohort. Additionally, the new healthy and sick ADG informed how much feed was consumed therefore influencing feed cost.

For META cohorts, a new morbidity risk was calculated by dividing the cohort's existing odds of a BRD morbidity by the metaphylactic efficacy odds ratio from above. A new CFR was then calculated using the new morbidity risk. Similar to the simulated NO META cohorts, ADG needed to be adjusted along with head-days which in turn impacted income and total expense.

$$\text{Healthy ADG} = \frac{\text{ADG}}{\frac{\text{head received} - \text{head treated BRD} - \text{other mortalities}}{\text{head exited} + [(1 - \text{ADG adjustment}) \times (\frac{\text{head treated BRD} - \text{BRD mortalities}}{\text{head exited}})]}}$$

Formula 4. A formula to adjust ADG for healthy animals. ADG = average daily gain. ADG adjustment = 0.05 based on data from Schneider et al. (2009). BRD = bovine respiratory disease.

Following modeling of each cohort's opposite scenario, the resulting dataset included two values for net return based on the status of application of metaphylaxis on arrival. A new variable called Best Option was created to determine which metaphylaxis status yielded the highest net return, Metaphylaxis or No Metaphylaxis. All economic calculations and metaphylaxis scenario adjustments were performed in R Studio (2024). The dataset was reduced to include only variables that would be known at the time of arrival to the feedyard and was used to create the Baseline predictive models. Additional data were added to the Baseline dataset in order to determine if this would improve predictive ability. This process is described in greater detail below.

Predictive modeling

Creation of datasets for predictive models

A Baseline dataset was built to represent the cattle demographic information known at arrival. These included variables describing sex, head count at arrival, arrival month and year, average arrival weight, and the variable to be predicted: Best Option (Metaphylaxis or No Metaphylaxis). The Baseline dataset did not include origin information as these data were not available for each cohort. To determine the impact of origin information on predictive ability of machine learning algorithms, a dataset was created with only cohorts that included this information. The Baseline + Origin dataset, therefore, included the same demographic variables from above in addition to variables describing the city, state, latitude, and longitude of the cohort's origin. The Baseline + Origin dataset had fewer cohorts ($n = 2406$); so, to compare to the Baseline using the same number of cohorts, a subset of the Baseline dataset called Baseline Subset was created and used to build models.

One objective of this study was to understand the influence of various predictor variables describing the relevant market prices on the predictive impact of models trained to determine whether a cohort of cattle entering the feedyard should receive metaphylaxis. To do this, additional data were captured from Livestock Marketing Information Center to describe futures prices of various commodities for each month from 2015 to 2023 (Livestock Marketing Information Center, 2024a; 2024b). This included feeder cattle, fat cattle, soybean, soybean meal, and corn futures prices. Brent oil futures prices were obtained from a different database

(Investing.com, 2005). The futures prices data represented the nearby futures contracts for each month within each year.

In addition to the relevant eight-year span (2015-2023), historical data were collected up to a decade prior to 2015. This was done to capture monthly trends in commodity futures prices. Ten-year averages of nearby futures prices for each month were calculated for each commodity. Each cohort's current months nearby futures price was determined and additional variables then captured the difference between the nearby futures prices of each commodity and the ten-year average of the same commodity. Monthly national interest rate data were captured from the Federal Reserve and similar variables as described for futures prices data were created (FedPrimeRate.com, 2005). Combined with the original demographic data, these new predictor variables created a Baseline + Econ dataset.

The combination of the origin and additional economic variables with the Baseline was also created to determine the effect of both sets of new variables on predictive ability. In total, five different datasets (Baseline, Baseline + Econ, Baseline Subset, Baseline + Origin, Baseline + Origin + Econ) were generated to create predictive models that were trained to determine whether or not a cohort should receive metaphylaxis on arrival (Figure 5.1). Each of the five datasets were used to train four different algorithms using the same methodology. This methodology and the metrics used to evaluate each set of algorithms are described below.

Predictive modeling

All predictive models were built in Microsoft Azure Machine Learning Studio (Microsoft, 2024). The procedure used to build, optimize, and test the predictive algorithms in this study is similar to what has been described previously by Heinen et al. (2024). To summarize, the dataset is split stratified by outcome of interest (Best Option, in this case) into a training and testing set. The training set, accounting for 75 % of the whole dataset, is further split. Another 75% of the training split is used to train an algorithm while the remaining 25 % is used in procedures called feature importance and hyperparameter tuning.

Feature importance computes importance scores for each predictor variable of a dataset and ranks them based on a specified evaluation metric. For this study, accuracy (the proportion of true results to total cases) was selected. Importance scores are calculated by computing the sensitivity of the selected metric to random permutations of the features. The idea being that if a variable is not important to the prediction of the outcome of interest, that random permutations of

that variable will have little effect on the selected metrics. In this manner, a list of the variables is generated from most important to least important (Bleik, 2015). This can then be used to reduce the number of predictor variables to just the 10 to 20 most important. This reduces run time and is a commonly used technique in machine learning to optimize generalizability of models (Bleik, 2015). Our datasets contained less than 30 variables so, although feature selection was performed and results will be reported, no filtering of features was performed on these models and all variables in the dataset were included in model building.

Hyperparameter tuning also optimizes predictive model performance. The processing of tuning involves performing a parameter sweep to determine an optimal set of hyperparameters for each specific trained model (“Tune Model Hyperparameters,” 2024). Parameters are unique to each algorithm and can be adjusted manually to improve performance. Performing a sweep and selecting a metric to evaluate parameters allows us to easily determine which parameters result in optimized performance. Area Under the Receiver Operating Characteristics (AUC-ROC) was selected as the primary metric for evaluating performance. This measure represents the ability of the model to distinguish between the two classes of outcome of interest. Following determination of optimum hyperparameters, these values were adjusted in the final training step. The testing subset (25% of the whole dataset) was then used to train and test the final model in a process called crossfold validation. This is standardized practice in machine learning and is used to evaluate the variability and thus, reliability of the dataset of interest to train and test a model.

The testing subset was divided into 10 equal folds at random. Of those 10 folds, the model is trained on 9 and validated with the final fold. The process is repeated so that each fold takes a turn as the validation fold. At the end of the process, there are 10 sets of predictive model performance values (“Cross Validate Model,” 2024). The metrics used to evaluate performance of predictive models include AUC-ROC, sensitivity, specificity, accuracy, positive predictive value, and negative predictive value. Each metric is described in more detail below.

Area under the Receiver Operating Characteristics curve (AUC-ROC) is used to evaluate the predictive ability of a model. This outcome gives a good overall impression of the model’s ability to differentiate between two outcome classes, and it is measured as a probability with values ranging from 0 to 1. At a 0.5 level, the model is guessing at the outcome class, with values closer to 1 representing a better model. Sensitivity is a value commonly represented as a percentage that describes the proportion of the META cohorts that the model correctly identifies.

On the opposite side, specificity describes the proportion of the NO META cohorts that the model correctly identifies. Positive predictive value (PPV) and negative predictive value (NPV) are both probabilities. PPV determines how likely the cases called META by the model are truly in that class. NPV determines how likely the cases called NO META by the model are truly in that class.

Four different predictive modeling algorithms were trained and tested for each dataset described above (Baseline, Baseline + Econ, Baseline Subset, Baseline + Origin, Baseline + Origin + Econ). These were chosen based on experience among the authors (Amrine, 2014; Amrine, 2019; Heinen, 2023; Rojas, 2022). The differences between algorithms stem from the math working in the background to generate probabilities and assign classes. The algorithms are explained in greater detail below.

Boosted decision tree (BDT)

The decision tree algorithm is a popular method for classification prediction (Song and Lu, 2015). It can be widely adapted to many different data classes and problems. It is visualized as a tree with a root node (representing the whole population of interest) and decision nodes (representing predictor variables split at specific criteria) with branches connecting the two nodes to each other and to leaf nodes (indicating the final outcome) (Jijo and Abdulazeez, 2021). Boosting allows for improved performance of a decision tree by correcting the mistakes made by previous iterations of the decision tree (“Two-Class Boosted Decision Tree Component,” 2024).

Several parameters can be adjusted to improve performance. Details about each parameter and the consequences of adjustment are found on the model’s accompanying help page (“Two-Class Boosted Decision Tree Component,” 2024). Adjustment of the maximum number of leaves per tree allows one to control the number of terminal nodes created. Large numbers of terminal nodes result in greater precision but poorer generalizability. Similarly, the number of cases required to create a terminal node can also be changed. By increasing this number, the threshold for creating a new rule is stricter (i.e. requires more cases) and therefore makes that rule more specific. Learning rate determines the rate at which the learner takes to determine the optimal solution. By making this parameter smaller, you increase the time it takes the model to train. The number of trees constructed can also be changed. By creating more trees, the boosting method for increasing performance has more trees for which to correct errors but this can greatly increase run time.

Logistic regression (LR)

Logistic regression is a widely used and well understood statistical model. Used to model binary outcomes, it can be applied to complex, multivariate analysis to determine associations with predictor variables. In the context of predictive modeling, data are fit into a logistic function and the probability of experiencing the outcome is used by the algorithm to determine its prediction (“Two-Class Logistic Regression component,” 2024). Although a simple algorithm, logistic regression predictive models are capable of handling large amounts of complex data (Shipe et al., 2019).

Azure Machine Learning Studio allows the adjustment of a few parameters for logistic regression models (“Two-Class Logistic Regression component,” 2024). Regularization prevents overfitting by assigning a penalty to models with extreme coefficient values regardless of accuracy. Specifically for the models in the present study, L2 regularization was adjusted. This value should be adjusted if large amounts of data are used to train the model. Optimization tolerance can also be changed. This specifies a threshold for improvement between iterations that triggers the algorithm to converge on a solution. The closer this value is to zero, the longer the training time and the greater the risk of overfitting.

Neural network (NN)

Inspired by the human brain, neural network algorithms recognize patterns in data by assigning weights to the inputs to predict the outcome (Dreiseitl and Ohno-Machado, 2002). This process is repeated in layers and the number of layers varies depending on the complexity of relationships between outcomes and inputs (“Two-Class Neural Network component,” 2024). Neural networks are similar to logistic regression models in the way model parameters are determined by maximum likelihood estimation; however, while the coefficients and intercepts supplied by a logistic regression algorithm can be interpreted in the context of the data, this is not the case with parameters provided by neural network models. These types of algorithms are being increasingly applied to the medical field and have already been well-established in artificial intelligence software and image recognition technology (Nicholson, 2023).

A few parameters are set to default and cannot be changed manually by the user in the Azure software (“Two-Class Neural Network component,” 2024). Only one hidden layer is between the output and input layers and is fully connected to both. The number of nodes in input

layer is the same as the number of input variables. The number of nodes in the hidden layer is set at 100. The number of nodes in the output layer is set to the number of classes of the outcome variable (two in the present study. Learning rate may be adjusted to determine the size of step taken at each iterative model run before correction. Larger values here make for faster training time but can result in less accurate results. The number of learning iterations allows the user to specify the node weights during training (“Two-Class Neural Network component,” 2024).

Random forest (RF)

Random forest algorithms are derived from decision tree classifiers. Multiple decision trees are constructed during the training phase, each on a random subset of data. The algorithm then votes for a final prediction based on the results of all the trees (Liu et al., 2012). Traditionally, the trees are built using only subsets of input variables, again selected at random. However, the Azure Machine Learning Studio’s random forest model utilizes the entire selection of input variables in the dataset but starts building the trees using different variables (“Two-Class Decision Forest component,” 2024). Advantages of this algorithm include its ability to manage non-linearly correlated and noisy data and reduce overfitting (Sarica et al., 2017).

Using Azure Machine Learning Studio, individual trees are created using a bagging method, otherwise known as bootstrap aggregating (“Two-Class Decision Forest component,” 2024). Using this method, each tree is constructed on a random sample of data which is continually replaced until the dataset is the size of the original. Various parameters may be adjusted (“Two-Class Decision Forest component,” 2024). The number of decision trees can be adjusted. Increasing this number results in greater coverage but can dramatically increase run time. The maximum depth may also be adjusted to increase or decrease precision with the trade-off of increased or decreased overfitting risk. The minimum sample per leaf node can also be altered, similar to the BDT model described above.

Parameters

Baseline

For the Baseline BDT model, the learning rate was set to 0.05. The minimum number of samples per leaf node was set to 50 with a maximum number of leaves per tree set to 32. The total number of trees constructed was set to 100. The Baseline Subset BDT model had a learning rate which was set to 0.2, a minimum of 10 samples per leaf node, a maximum of 8 leaves per

tree, and a total of 20 trees constructed. The LR model regularization weight was set to 1.0 with an optimization tolerance of 0.00001 in the Baseline model. The L2 regularization weight in Baseline Subset model was changed to 0.1. The same NN model parameters were used in the Baseline and Baseline Subset series were the same. The learning rate was set to 0.4 with a number of learning iterations set to 160. The RF model parameters in the Baseline and Baseline Subset series included the decision tree maximum depth set to 16 with a total number of decision trees set at 32. The minimum number of samples per leaf node was set to 4.

Baseline + Econ

For the BDT model, a learning rate of 0.05 was selected. The number of trees was set at 100 with a maximum number of leaves per tree at 50, and a minimum number of samples per leaf node set at 32. Parameters for the LR model included an L2 regularization rate of 0.01 and the same optimization tolerance for the Baseline and Baseline Subset models. The NN parameters were same as previously written and did not change for any of the following model runs. The maximum depth of decision trees was set at 16 for the RF model. The number of decision trees was set at 32 with a minimum of 16 samples needed per leaf node.

Baseline + Origin

The BDT learning rate was set to 0.025. The minimum number of samples per leaf node was set to 50 with a maximum number of leaves per tree set to 8. The total number of trees constructed was set at 500. The parameters for LR were identical to those above described for the Baseline model. The RF parameters were as follows: maximum depth at 64, minimum number of samples per leaf node at 1, and number of decision trees at 32.

Baseline + Origin + Econ

A learning rate of 0.05 was set for the BDT model. There were 100 trees constructed with a maximum of 8 leaves per tree with a minimum of one sample needed per leaf node. The LR parameters remained identical to those reported for the Baseline Subset model. The RF model consisted of 32 decision trees each with a maximum depth of 16 and a minimum of one sample per leaf node required.

Results

Descriptive statistics

16,368 cohorts of cattle were included in this study. The average cohort size was 172 head. The overall average arrival weight was 762 lbs (345 kgs) and the average exist weight was 1,322 lbs (600 kgs). Approximately 44% of the cohorts were heifers (7,264/16,368) with the remaining 56% being steer cohorts. Average cohort level BRD mortality risk in the study population was 1.4%. Cohort-level BRD morbidity risk was 12.6% on average. Complete origin information (city, state, latitude, and longitude) was only available for 2,406 cohorts. In total, 33 states were represented in the origins with greater than 75% of cattle coming from Kansas, Oklahoma, Texas, Nebraska, and Missouri. Table 5.1 provides descriptive statistics for the dataset split by economic model-assigned outcome and metaphylaxis choice made by the feedyard.

The outcome of interest, Best Option, captures the results of the economic simulation described above which determines whether or not a cohort of cattle arriving at the feedyard should receive a metaphylactic antibiotic. In the study dataset ($n = 16,368$), 28.5% (4672/16,368) of the cohorts had a best option of “Metaphylaxis.”

Baseline

The Baseline dataset contained 16,368 cohorts. The prevalence of Best Option = Metaphylaxis is 28.5 % (4672/16,368). Baseline models included head received, average arrival weight, sex, arrival year, and arrival month as predictor variables. Average AUCs ranged from 0.81 to 0.93. Standard deviations of AUC-ROCs were small at 0.01 to 0.02. See Figure 5.2. Overall, sensitivities were high (87.9 – 91.3 %) indicating the models identified around 87 to 91 % of the cohorts whose Best Option was identified as “Metaphylaxis”. Specificities were moderate (36 – 74 %) meaning the models identified 36 to 74 % of the cohorts whose Best Option was “No Metaphylaxis.” PPV ranged from 0.78-0.90 and NPVs ranged from 0.55 – 0.77 for Baseline models. The BDT model demonstrated the highest metrics across all the models. However, the NN model had a higher sensitivity than the BDT model. The LR model had the lowest performance metrics.

Baseline + Econ

The Baseline + Econ dataset contained the original number of cohorts ($n = 16,368$) and the same prevalence of Best Option = Metaphylaxis (28.5 %). The Baseline + Econ models were given data to incorporate the 10-year monthly averages of soybean, soybean meal, corn, brent oil, feeder cattle, and fat cattle futures prices along with interest rates. Additionally, the difference between the current arrival month's commodity futures prices and interest rate to the comparable 10-year average was also calculated and added to the dataset as a predictor variable. In total, ten new predictor variables were added to the original Baseline dataset to build the Baseline + Econ dataset. AUC-ROC values ranged from 0.92 to 0.94 with small standard deviations (0.01) (Figure 5.2). Sensitivity values ranged from 89.7 % to 91.2% and specificity values ranged from 71 % to 75.9%. The PPVs and NPVs were also high at 0.89 to 0.90 and 0.74 to 0.77, respectively. For each performance metric, the BDT model ranked above the other models and the LR ranked the lowest. See Table 5.2 for a full summary of Baseline + Econ model performance metrics.

Baseline Subset

The Baseline Subset models were used to compare the models built with the Baseline + Origin and Baseline + Origin + Econ datasets. Only 2406 cohorts from the original 16,368 cohort dataset had origin data available. The prevalence of Best Option = Metaphylaxis was 31.0 % (747/2406). The mean AUC-ROC values of the Baseline Subset ranged from 0.80 to 0.89 with standard deviations of 0.03 to 0.05). See Figure 5.3 for visualization of AUC-ROC values. Sensitivity ranged from 85.8 to 89.3. Specificity ranged from 38.2 to 67.2. PPVs ranged from 0.87 to 0.90 and NPVs ranged from 0.55 to 0.77. The BDT model performed with the highest metrics while the LR model possessed the lowest metrics. See Table 5.2 for a complete summary of the metrics for the Baseline Subset models.

Baseline + Origin

The Baseline + Origin dataset consisted of 2406 cohorts with a prevalence of Best Option = Metaphylaxis of 31.0 % (747/2406). The Baseline + Origin models were given city, state, latitude, and longitude values corresponding to the origin of each of the cohorts. Mean AUC-ROC values for the Baseline + Origin models ranged from 0.79 to 0.89 with standard deviations ranging from 0.03 to 0.04 (Figure 5.3). The sensitivity values ranged from 80.0 % to 88.7 % and

the specificity values ranged from 39.1 % to 66.3 %. The range of PPVs was 0.76 to 0.85. The range of NPVs was 0.55 to 0.71. The BDT model again performed the best according to all metrics. The LR had the lowest performance metrics.

Baseline + Origin + Econ

The Baseline + Origin + Econ dataset has origin information as well as additional economic predictor variables in addition to the same cohort demographic variables included in the Baseline dataset resulting in an additional 14 variables. The dataset contains 2,406 cohorts and a prevalence of Best Option = Metaphylaxis of 31.0 %. The average AUC-ROC values for the Baseline + Origin + Econ models demonstrated a range from 0.87 to 0.91 and standard deviations of 0.03 to 0.04 (Figure 5.3). Sensitivities were high with a range of 88.1 to 90.1 % with lower specificities of 61.2 to 69.4 %. PPVs had a narrow range from 0.83 to 0.87. NPVs had a range of 0.71 to 0.75. The BDT outperformed all models with respect to the performance metrics calculated. The LR model had the lowest values for each performance metric apart from NPV. The NN model had the lowest NPV value.

Feature importance

Feature importance rankings were performed for each model run for each dataset. All feature importance data could not be displayed. The BDT model displayed the highest AUC-ROC values for each dataset so only the top three variables are presented for each model set in Table 3. A full ranking of the variables for each dataset are displayed in the supplementary material (Appendix Table C.1).

Discussion

The study objective was to evaluate the ability of predictive models to determine if on-arrival metaphylaxis application is optimal for groups of cattle given cattle demographic data. The Baseline models were trained with cattle demographic data to predict Best Option. The AUC-ROC values ranged from 0.81 to 0.93 demonstrating a good ability to distinguish between the outcome classes, Metaphylaxis and No Metaphylaxis. Specifically, the high sensitivity values (87.9-91.0 %) demonstrate that the Baseline models are able to detect high proportions of the Metaphylaxis class.

The secondary objective was to compare the performance of the Baseline models to models built with additional origin and external economic data. The Baseline + Econ models

resulted in higher AUC-ROC values than those of the Baseline (0.92 to 0.94). The Baseline + Econ sensitivity values are similar to Baseline sensitivity values; however, the specificity values were higher (71-75.9 % for the Baseline + Econ models, 36-73.9 % for the Baseline models) indicating an improved ability to predict the No Metaphylaxis class in the Baseline + Econ models. Adding origin data resulted in slightly lower AUC-ROC values as compared to the Baseline Subset (0.79- 0.89 vs. 0.80- 0.89, respectively). Other performance metrics remained similar between the Baseline + Origin and Baseline Subset models. The Baseline + Origin + Econ models possessed higher AUC-ROC values as compared to the Baseline Subset (0.87-0.91 vs 0.80-0.89, respectively). Similar to the pattern seen in the Baseline vs. Baseline + Econ comparison, sensitivity values were comparable between the Baseline Subset and Baseline + Origin + Econ models, but the specificity values increased in the Baseline + Origin + New Econ models. The addition of origin data appears to provide no benefit to model performance whereas the addition of commodity futures prices variables results in higher AUC-ROC values and increased specificity.

The overall rate of cohorts receiving metaphylaxis (META) at the feedyards in the dataset was 15.8 % (2594/16,368). When using the economic model to determine whether metaphylaxis is the best option, the number of cohorts receiving metaphylaxis increases to 24.8 % (4071/16,368). This indicates that using net returns as the deciding factor for metaphylactic treatment, results in a greater prevalence of metaphylaxis application than what is currently performed. This could be due to pressures on feedyard management to reduce metaphylaxis application to cohort arriving at feedyards. Estimates for net return of using metaphylaxis ranges from \$90 to \$118 per head in high risk, yearling animals when compared to no administration of metaphylaxis (Dennis et al., 2020). Another study reports a mean profit per head ranging from \$45 to \$84 in metaphylaxis-treated, high risk calves compared to control calves which suffered from an average loss of \$41 per head (Corbin et al., 2009). The findings of different net returns between metaphylaxis-treated calves and control calves were not repeated in a study on medium-risk calves (Horton et al., 2023). Overall, it appears that metaphylactic application of an antimicrobial at the time of arrival improves morbidity and mortality risks as well as performance outcomes, leading to high net returns although this finding is not consistent in groups of cattle that aren't high risk. These positive net returns seen with metaphylaxis likely drove our economic model to assign metaphylaxis as the best option to cohorts more frequently

than the feedyards were actually administering. It is important to acknowledge that economic returns should not be the sole driver of a decision to apply metaphylaxis or not. Public health concerns, such as antimicrobial resistance, and animal welfare are difficult, if not impossible, to measure objectively but are important factors to consider.

The models in this study possessed AUC-ROC values between 0.79 to 0.94 indicating good performance. This range is greater than values from other research in our group investigating different outcomes with similar datasets (Amrine, 2014; Amrine, 2019; Heinen, 2023; Heinen, 2024; Rojas, 2022). Across model series tested in the present study, the BDT and RF models often demonstrated higher AUC-ROC values. Decision tree algorithms have consistently performed well in other research done within our group (Heinen, 2023; Heinen, 2024). Surprisingly, the LR models demonstrated lower AUC-ROC values than the NN models in several comparisons. This differs from findings previously reported where NN models had the lowest AUC-ROC values (Heinen, 2023).

Higher AUC-ROC values in the current study compared to previous by our group are likely due to the nature of the data (Amrine, 2014; Amrine, 2019; Heinen, 2023; Heinen, 2024; Rojas, 2022). Previous papers investigated the use of predictive models to make predictions on health outcomes (either morbidity or mortality due to BRD). These biological processes are likely very complex and difficult to predict. The current study focused on a prediction of an economic outcome. The data used to inform the economic outcome came from historical databases with easily observed trends. Additionally, trends in the market are driven by consumer behavior and not subject to the randomness of biology. Furthermore, the predictions made in the current study were on a cohort-level where several previous papers focused on predictions made on an individual animal-level. Previous investigations often used group-level averages as predictor variables for an individual animal outcome which likely introduced much error. In comparison, this experiment applied group-level averages to a group-level prediction potentially lessening error and minimizing noise compared to previous investigations (Amrine, 2014; Amrine, 2019; Heinen, 2023; Heinen, 2024; Rojas, 2022).

The Baseline + Econ models containing external economic data such as corn and feeder cattle futures prices at the time of arrival of a cohort had higher AUC-ROC values than the Baseline (0.92-0.94 vs. 0.81-0.93). This could be due to the information captured in the futures prices variables. As futures contracts reach their expiration date, the prices converge on the

market price for that asset (Phung, 2024). This convergence is a common phenomenon in a well-functioning market. The data captured in the futures prices variables in our models could be highly informative of the economic outcome variable, which was calculated using cattle prices as well as feedstuff prices, leading to our high AUC-ROC values in the Baseline + Econ models. Additionally, the cattle market likely influences cattle buying behavior and vice versa which in turn could change the risk tolerance of cattle buyers. Feeder cattle prices specifically could drive cattle buyers to purchase groups of cattle at higher or lower risk of BRD outbreaks. This could be the reason feeder cattle futures was ranked as an important feature; however, using feature importance ranking does not allow one to make specific assumptions about the direction and magnitude an effect has on an outcome of interest.

The addition of economic data in the form of average and current commodity futures prices resulted in increased model performance. Subsequently, the new economic variables appeared high in the feature importance ranking while still ranking below the cattle demographic variables. Among the higher-ranking variables, fat cattle and feeder cattle futures prices frequently made an appearance. This makes sense as these futures prices represent the projected value of cattle in the market, making metaphylaxis to reduce morbidity and mortality a more valuable management tool. While predictive models do not offer specific associations to be made between variables, it can be assumed that as futures prices increased and cattle were worth more, it would become more profitable to apply metaphylaxis to cohorts of cattle entering the feedyard in order to protect that investment.

The Baseline + Origin models displayed similar AUC-ROC values to the Baseline Subset (0.79 – 0.89 vs 0.80 – 0.89) demonstrating that the addition of origin data may not be useful in the prediction of whether an incoming cohort of cattle to the feedyard should receive metaphylaxis. This could be because origin data received for a subset of the cohort included in this study reflected the last origin before cattle arrived. This information did not necessarily reflect where cattle were born or raised, which may be more pertinent information. Additionally, both ranges of the Baseline + Origin and Baseline Subset AUC-ROCs were slightly lower than the range reported for the Baseline (0.81 – 0.93). This finding highlights an important consideration for predictive modeling. Large datasets are necessary for the iterative process needed to generate predictive models that are reliable and generalizable to a greater population (Wisnieski et al., 2021). Due to the limited origin data available, only a subset of the total dataset

(2406/16,368) was able to be used to build both the Baseline + Origin and Baseline Subset models likely resulting in greater variability in performance and decreased generalizability when exposed to the novel testing data.

Origin variables appeared to rank lower than both cattle demographic and economic variables. Overall, Baseline + Origin models performed with similar AUC-ROC values to the Baseline subset. In feature importance rankings, the origin variables tended not to rank high, and their ranking order appeared random. This is surprising because cattle origin often informs management decisions upon feedyard arrival. Biases exist with cattle buyers for the quality of calves from different regions of the United States. For example, cattle buyers in the Midwest may believe that calves from ranches in the southeastern states will experience higher BRD morbidity and mortality than calves from northern states. Perhaps, the distribution of our data sources and overall rarity of origin information created noise in the data that were not informative for the predictive models. Perhaps, the bias among cattle buyers against southeastern cattle is unfounded. Further investigation is prompted to determine if origin variables can improve the prediction of the benefit of metaphylaxis application to cohorts of cattle at feedyard arrival.

The results of our models were characterized by overall high sensitivity values (85.8-91.3%) and moderate to low specificity values (36.0-75.9%). These are opposite values of what has been reported by other studies applying predictive models to feedyard management decisions (Amrine, 2014; Amrine, 2019; Heinen, 2023; Heinen, 2024; Rojas, 2022). These papers reported overall high specificities and low sensitivities. This is most likely due to the differences in outcome class distributions. Amrine et al. (2014) and Heinen et al. (2023) used BRD case fatality as their outcome of interest. Case fatality risks in their datasets ranged from 8.5% to 14.2%, resulting in a more unbalanced dataset with regards to the outcome classes than in the current study. Generally, predictive models perform better with balanced data (Haixiang, 2017). Additionally, extreme cases of unbalanced outcome classes result in biased selection of the modeling algorithm to the majority class, resulting in much higher sensitivity or specificity (depending on the majority outcome class). Our dataset did not demonstrate highly imbalanced outcome classes which likely contributed to the more balanced nature of our sensitivity, specificity, PPV, and NPV when compared to previous work (Amrine, 2014; Amrine, 2019; Heinen, 2023; Heinen, 2024; Rojas, 2022).

Feature importance ranking was performed in the current study to evaluate the contribution of various input variables to the prediction of metaphylaxis on arrival. Cattle demographic variables consistently ranked high. This makes sense as these variables are often used by cattle managers to inform the susceptibility to BRD outbreaks in cohorts of cattle arriving at the feedyard. Arrival month and year often ranked above other cattle demographic variables. Month was likely a proxy for the type of cattle entering the feedyard along with average arrival weight, which also ranked high. The combination of these three variables likely informed the predictive models in this study of the relative age. Younger, lighter weight cattle placed in the fall months are typically seen as higher risk for a BRD outbreak due to the high likelihood of recent weaning. Young cattle at weaning are more susceptible to illness due to risk of an underdeveloped immune system (Booker and Lubbers, 2020).

As with all economic models, the decision to administer metaphylaxis to a cohort of cattle based on the calculations made in the current study is based on many assumptions. These assumptions are very likely not representative of every cohort of cattle on feed in North America. However, we believe that this economic model is useful as it incorporates the most important inputs and costs associated with placing cattle on feed such as feeder and fed cattle price, corn prices, and interest rates (Mark, 2000). Additionally, the net returns of a decision may not be the only determining factor. Other motives for administering metaphylaxis or not include animal welfare concerns, a desire to reduce antimicrobial resistance, labor or facility constraints reducing the ability of an operation to respond adequately to a large health challenge, or the marketing of a “never treated with antibiotic” product. Using these outcomes, the proportion of cohorts receiving metaphylaxis would likely be different than what we modeled. Furthermore, by basing the outcome variable on a sole, objective measurement (net returns), this study could support the variety of subjective assessments made by feedyard personnel. In addition to subjective evaluations made by people, this model does not consider additional aspects that are likely to motivate application or not of metaphylaxis to a group of cattle such as desire to reduce antimicrobial use and the public’s perception of antimicrobial use in cattle.

The results of the current study demonstrate that certain types of data are important for the prediction of Best Option. The origin data collected for the Baseline + Origin and Baseline + Origin + Econ models did not improve model performance when compared to the Baseline Subset. Based on this and its limited availability, it is likely not useful to collect and include in

the prediction of Best Option. Perhaps, if more accurate information was available about a group of cattle's ranch origin, there would be improved model performance. The addition of external economic data improved predictive ability in the current study. The cattle market is heavily influenced by other commodity markets, so it is likely that it is important to incorporate these data in prediction of Best Option. Further investigation is needed to establish which external economic data would be most useful to the prediction of Best Option.

A limitation of the current study is the use of retrospective data collected from participating feedyards. One aspect of the study that is greatly impacted by this limitation is the diagnosis of BRD made by feedyard personnel. Clinical diagnosis of BRD in cattle has been demonstrated to have low sensitivity (Churchill et. al, 2025). By using a subjective evaluation to diagnose disease in feedyard cattle, this study may have been limited in its ability to diagnose true cases of BRD. Despite this limitation, using the data this way provides high external validity as subjective evaluation of cattle is the industry standard for diagnosis of disease with very few in the feedyard industry utilizing additional diagnostics.

Perhaps the biggest limitation of predictive modeling is the inability to draw conclusions about associations between predictor variables and the outcome of interest. Numerical associations from traditional statistical models often help inform decision making in the real world. This is not possible with predictive modeling. However, using predictive models in the manner described in the current study could be advantageous to the cattle industry. Machine learning algorithms, like the models used in this study, can handle large amounts of complex data and run in real time to assist with management decisions.

The goal of this research was to determine the ability of predictive models to predict the best option for metaphylaxis application to groups of cattle given cattle demographic information, origin data, and external economic data. Different algorithms were evaluated, and comparisons were made to determine which models and data would result in the greatest performance. Overall, cattle demographic data alone resulted in high-performing predictive models. Performance improved with external economic data describing market conditions around the time of arrival and trends over several years. The addition of origin data, however, resulted in poorer performance as measured with AUC-ROC values. Cattle demographic variables consistently ranked high in variable rankings based on contribution to prediction. Several economic variables also ranked highly when included. The Best Option variable in this

study was based on a comparison of net returns between administering metaphylaxis or not. In reality, the decision to administer metaphylaxis to a cohort of cattle at arrival is based on a variety of subjective and objective factors. Future research should incorporate these factors into the outcome to further refine predictive models as a potential tool for making a management decision regarding metaphylactic application to cohorts of cattle at arrival to a feedyard.

References

- Abell, K., Theurer, M., Larson, R., White, B., Apley, M., 2017. A mixed treatment comparison meta-analysis of metaphylaxis treatments for bovine respiratory disease in beef cattle. *J. Anim. Sci* 95, 626–635. <https://doi.org/10.2527/jas.2016.1062>
- Amrine, D.E., White, B.J., Larson, R.L., 2014. Comparison of classification algorithms to predict outcomes of feedlot cattle identified and treated for bovine respiratory disease. *Comput. Electron. Agr.* 105, 9–19. <https://doi.org/10.1016/j.compag.2014.04.009>
- Amrine, D.E., McLellan, J.G., White, B.J., Larson, R.L., Renter, D.G., Sanderson, M., 2019. Evaluation of three classification models to predict risk class of cattle cohorts developing bovine respiratory disease within the first 14 days on feed using on-arrival and/or pre-arrival information. *Comput. Electron. Agr.* 156, 439–446. <https://doi.org/10.1016/j.compag.2018.11.035>
- Astill, J., Dara, R., Fraser, E., Sharif, S., 2018. Detecting and predicting emerging disease in poultry with the implementation of new technologies and big data: a focus on Avian Influenza Virus. *Front. Vet. Sci.* 5. <https://doi.org/10.3389/fvets.2018.00263>
- Baptiste, K., Kyvsgaard, N., 2017. Do antimicrobial mass medications work? A systematic review and meta-analysis of randomised clinical trials investigating antimicrobial prophylaxis or metaphylaxis against naturally occurring bovine respiratory disease. *Pathog. Dis.* 75, 1–12. <https://doi.org/10.1093/femspd/ftx083>
- Bleik, S., 2015. Permutation Feature Importance. Microsoft Machine Learning Blog. URL <https://learn.microsoft.com/en-us/archive/blogs/machinelearning/permutation-feature-importance> (accessed 11.22.24).
- Booker, C., Lubbers, B., 2020. Bovine respiratory disease treatment failure: impact and potential causes. *Vet Clin North Am Food Anim Pract* 36, 487–496. <https://doi.org/10.1016/j.cvfa.2020.03.007>

2968 Cernicchiaro, N., Renter, D., White, B., Babcock, A., Fox, J., 2011. Associations between
 2969 weather conditions during the first 45 days after feedlot arrival and daily respiratory
 2970 disease risks in autumn-placed feeder cattle in the United States. *J. Anim. Sci.* 90, 1328–
 2971 1337. <https://doi.org/10.2527/jas.2011-4657>
 2972 Cernicchiaro, N., White, B., Renter, D., Babcock, A., Kelly, L., Slattery, R., 2012. Associations
 2973 between the distance traveled from sale barns to commercial feedlots in the United States
 2974 and overall performance, risk of respiratory disease, and cumulative mortality in feeder
 2975 cattle during 1997 to 2009. *J. Anim. Sci.* 90, 1929–1939.
 2976 <https://doi.org/10.2527/jas.2011-4599>
 2977 Churchill, K.J., Winder, C.B., Goetz, H.M., Wilson, D., Uyama, T., Pardon, B., Renaud, D.L.,
 2978 2025. Evaluating case definitions of respiratory disease in dairy calves: a scoping review.
 2979 *J. Dairy Sci.* 108 (4), 4030-4048. <https://doi.org/10.3168/jds.2024-25827>
 2980 Corbin, M.J., Gould, J.A., Carter, B.L., McClary, D.G., Portillo, T.A., 2009. Effects and
 2981 economic implications of metaphylactic treatment of feeder cattle with two different
 2982 dosages of tilmicosin on the incidence of bovine respiratory disease (BRD) - a summary
 2983 of two studies. *Bov. Pract.* 43, 140–152. [https://doi.org/10.21423/bovine-vol43no2p140-](https://doi.org/10.21423/bovine-vol43no2p140-152)
 2984 152
 2985 Cross Validate Model, 2024. . Microsoft Machine Learning Blog. URL
 2986 [https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/cross-](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/cross-validate-model?view=azureml-api-2)
 2987 [validate-model?view=azureml-api-2](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/cross-validate-model?view=azureml-api-2) (accessed 11.25.24).
 2988 Dennis, E., Schroeder, T., Renter, D., 2020. Net return distributions when metaphylaxis is used
 2989 to control bovine respiratory disease in high health-risk cattle. *Trans. Anim. Sci* 4, 1091–
 2990 1102. <https://doi.org/10.1093/tas/txaa020>
 2991 Dreiseitl, S., Ohno-Machado, L., 2002. Logistic regression and artificial neural network
 2992 classification models: a methodology review. *J. Biomed. Inform.* 35, 352–359.
 2993 [https://doi.org/10.1016/S1532-0464\(03\)00034-0](https://doi.org/10.1016/S1532-0464(03)00034-0)
 2994 Drouillard, J., 2018. Current situation and future trends for beef production in the United States
 2995 of America - a review. *AJAS* 31, 1007–1016. <https://doi.org/10.5713/ajas.18.0428>
 2996 European Union, 2015. Guidelines for the prudent use of antimicrobials in veterinary medicine
 2997 [WWW Document]. European Union. URL

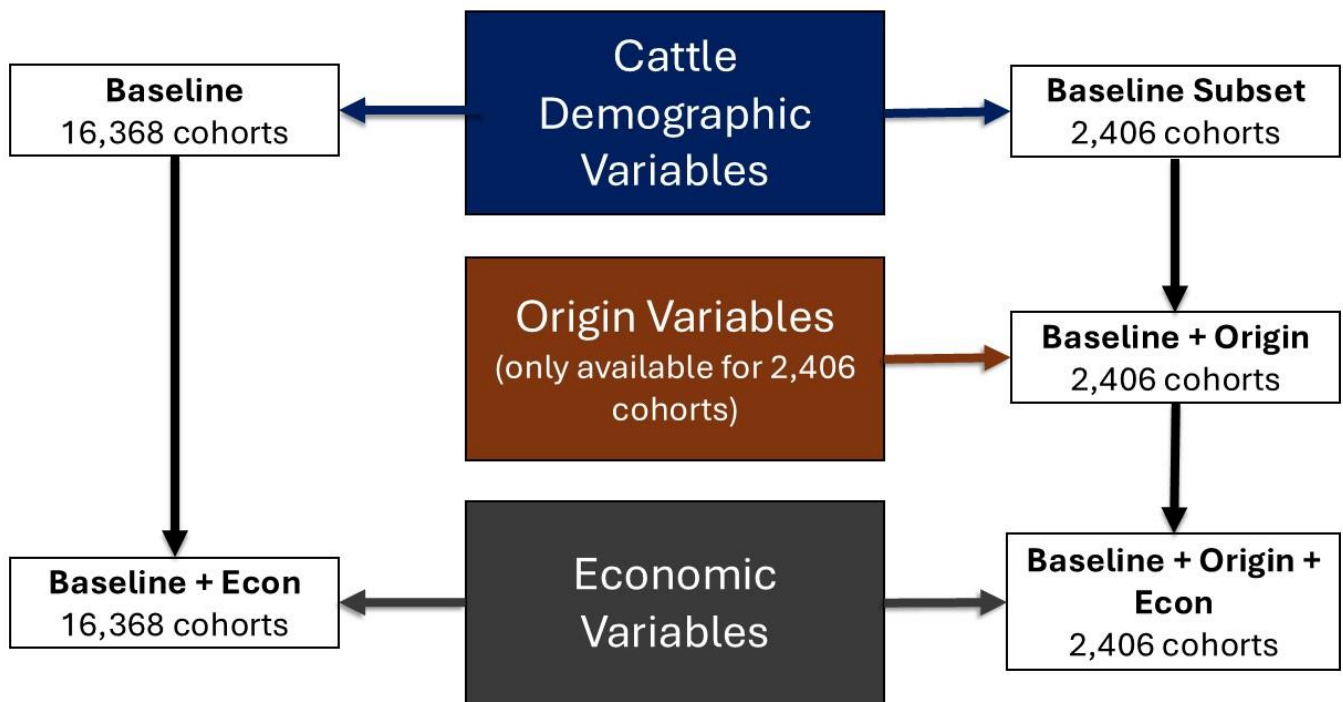
2998 [https://health.ec.europa.eu/document/download/190841e8-5975-4390-a304-](https://health.ec.europa.eu/document/download/190841e8-5975-4390-a304-908c259592ab_en)
 2999 [908c259592ab_en](https://health.ec.europa.eu/document/download/190841e8-5975-4390-a304-908c259592ab_en) (accessed 12.10.24).
 3000 FedPrimeRate.com, 2005. Prime Rate History – Monthly [dataset].
 3001 Fulton, R., 2009. Bovine respiratory disease research (1983-2009). *Anim. Health Res. Rev.* 10,
 3002 131–139. <https://doi.org/10.1017/S146625230999017X>
 3003 Gardner, B., Dolezal, H., Bryant, L., Owens, F., Smith, R., 1999. Health of finishing steers:
 3004 effects on performance, carcass traits, and meat tenderness. *J. Anim. Sci* 77, 3168–3175.
 3005 <https://doi.org/10.2527/1999.77123168x>
 3006 Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G., 2017. Learning from
 3007 class-imbalanced data: review of methods and applications. *Expert Syst. Appl.* 73, 220–
 3008 239. <https://doi.org/10.1016/j.eswa.2016.12.035>
 3009 Heinen, L., White, B., Amrine, D., Larson, R., 2023. Evaluation of predictive models to
 3010 determine final outcome for feedlot cattle based on information available at first
 3011 treatment for bovine respiratory disease. *Am. J. Vet. Res.* 84.
 3012 <https://doi.org/10.2460/ajvr.23.05.0094>
 3013 Heinen, L., White, B., Larson, R., Kopp, D., Pendell, D., 2024. Economic impact of mortality
 3014 prediction by predictive model at first and second treatment for bovine respiratory
 3015 disease. *Am. J. Vet. Res.* <https://doi.org/10.2460/ajvr.24.06.0169>
 3016 Horton, L., Depenbusch, B., Dewsbury, D., McAtee, T., Betts, N., Renter, D., 2023.
 3017 Comprehensive outcomes affected by antimicrobial metaphylaxis of feedlot calves at
 3018 medium-risk for bovine respiratory disease from a randomized controlled trial. *Vet. Sci.*
 3019 10, 67–90. <https://doi.org/10.3390/vetsci10020067>
 3020 Huuskonen, A., Huhtanen, P., Joki-Tokola, E., 2013. The development of a model to predict
 3021 intake by growing cattle. *Livest. Sci.* 158, 74–83.
 3022 <https://doi.org/10.1016/j.livsci.2013.10.005>
 3023 Investing.com, 2005. Brent Oil Futures Historical Data [dataset].
 3024 Irsik, M., Langemeier, M., Schroeder, T., Spire, M., Roder, J., 2006. Estimating the effects of
 3025 animal health on the performance of feedlot cattle. *Bov. Pract.* 40, 65–74.
 3026 <https://doi.org/10.21423/bovine-vol40no2p65-74>

Ives, S., Richeson, J., 2015. Use of antimicrobial metaphylaxis for the control of bovine
 respiratory disease in high-risk cattle. *Vet. Clin. North Am. Food Anim. Pract.* 31, 341–
 350. <https://doi.org/10.1016/j.cvfa.2015.05.008>
 Jijo, B.T., Abdulazeez, A.M., 2021. Classification based on decision tree algorithm for machine
 learning. *Journal of Applied Science and Technology Trends* 2, 20–28.
<https://doi.org/10.38094/jastt20165>.
 Kopp, D.J., Larson, R.L., Lancaster, P.A., White, B.J., Smith, K.J., Pendell, D.L., 2024.
 Determining the Economically Optimum Metaphylactic Strategy for Cattle Cohorts of
 Varied Demographic Characteristics. *Animals* 14, 1423–1435.
<https://doi.org/10.3390/ani14101423>
 K-State Research and Extension, 2015. Kansas Feedlot Performance and Feed Cost Summary
 [dataset].
 Lee, W., Ham, Y., Ban, T., Jo, O., 2019. Analysis of growth performance in swine based on
 machine learning. *IEEE Access* 7, 161716–161724.
<https://doi.org/10.1109/ACCESS.2019.2951522>
 Liu, Y., Wang, Y., Zhang, J., 2012. New machine learning algorithm: random forest, in:
 Information Computing and Applications. Presented at the International Conference on
 Information Computing and Applications, Springer-Verlag Berlin Heidelberg, Chengde,
 China, pp. 246–252. https://doi.org/10.1007/978-3-642-34062-8_32
 Livestock Marketing Information Center, 2024b. Weekly Commodity Futures [dataset].
 Livestock Marketing Information Center, 2024a. Weekly Slaughter Cattle Price. [dataset]
 Macrosyn Tulathromycin for Cattle and Swine [WWW Document], n.d. . ValleyVet.com. URL
[https://www.valleyvet.com/ct_detail.html?pgguid=A1DD674F-99E9-41E7-BD16-
 5CC0FC3E0C8B](https://www.valleyvet.com/ct_detail.html?pgguid=A1DD674F-99E9-41E7-BD16-5CC0FC3E0C8B) (accessed 6.1.22).
 Magalhaes, E., Zhang, D., Wang, C., Thomas, P., Moura, C., Holtkamp, D., Trevisan, G.,
 Rademacher, C., Silva, G., Linhares, D., 2023. Field implementation of forecasting
 models for predicting nursery mortality in a Midwestern US swine production system.
Animals 13, 2412–2423. <https://doi.org/10.3390/ani13152412>
 Mark, D., Schroeder, T. C., Jones, R. D., 2000. Identifying economic risk in cattle feeding. *J.*
Agribus. 18(3), 331–344. <http://dx.doi.org/10.22004/ag.econ.14712>.
 Microsoft, 2024. Azure Machine Learning. Microsoft, Washington, USA.

3058 Nicholson, C., 2023. A beginner's guide to neural networks and deep learning [WWW
 3059 Document]. Pathmind. URL <https://wiki.pathmind.com/neural-network> (accessed
 3060 11.25.24).
 3061 Nickell, J., White, B., 2010. Metaphylactic antimicrobial therapy for bovine respiratory disease
 3062 in stocker and feedlot cattle. *Vet. Clin. North Am. Food Anim. Pract.* 26, 285–301.
 3063 <https://doi.org/10.1016/j.cvfa.2010.04.006>
 3064 Omomule, T., Ajayi, O., Orogun, A., 2020. Fuzzy prediction and pattern analysis in poultry egg
 3065 production. *Comput. Electron. Agric.* 171. <https://doi.org/10.1016/j.compag.2020.105301>
 3066 O'Neill, R., Mooney, J., Connaghan, E., Furphy, C., Graham, D., 2014. Patterns of detection of
 3067 respiratory viruses in nasal swabs from calves in Ireland: a retrospective study. *Vet.*
 3068 *Record* 175, 351–351. <https://doi.org/10.1136/vr.102574>
 3069 The Pecan Team, 2023. The complete guide to predictive modeling [WWW Document]. Pecan
 3070 AI. URL <https://www.pecan.ai/blog/predictive-modeling/> (accessed 12.10.24).
 3071 Phung, A., 2024. Futures Prices Converge Upon Spot Prices [WWW Document]. Investopedia.
 3072 URL
 3073 [https://www.investopedia.com/ask/answers/06/futuresconvergespot.asp#:~:text=Futures](https://www.investopedia.com/ask/answers/06/futuresconvergespot.asp#:~:text=Futures%20prices%20converging%20upon%20spot,contract%20approaches%20its%20expiration%20date)
 3074 [%20prices%20converging%20upon%20spot,contract%20approaches%20its%20expiratio](https://www.investopedia.com/ask/answers/06/futuresconvergespot.asp#:~:text=Futures%20prices%20converging%20upon%20spot,contract%20approaches%20its%20expiration%20date)
 3075 [n%20date](https://www.investopedia.com/ask/answers/06/futuresconvergespot.asp#:~:text=Futures%20prices%20converging%20upon%20spot,contract%20approaches%20its%20expiration%20date). (accessed 1.21.25).
 3076 Rojas, H.A., White, B.J., Amrine, D.E., Larson, R.L., 2022. Predicting bovine respiratory disease
 3077 risk in feedlot cattle in the first 45 days post arrival. *Pathogens* 11, 442–456.
 3078 <https://doi.org/10.3390/pathogens11040442>
 3079 RStudio Team, 2024. RStudio: integrated development for R. RStudio, Boston, MA.
 3080 Rutten, C., Kamphuis, C., Hogeveen, H., Huijps, K., Nielen, M., Steeneveld, W., 2017. Sensor
 3081 data on cow activity, rumination, and ear temperature improve prediction of the start of
 3082 calving in dairy cows. *Comput. Electron. Agric.* 132, 108–118.
 3083 <https://doi.org/10.1016/j.compag.2016.11.009>
 3084 Samuelson, K., Hubbert, M., Galyean, M., Loest, C., 2016. Nutritional recommendations of
 3085 feedlot consulting nutritionists: the 2015 New Mexico State and Texas Tech University
 3086 survey. *J. Anim. Sci.* 94, 2648–2663. <https://doi.org/10.2527/jas.2016-0282>

3087 Sanderson, M., Dargatz, D., Wagner, B., 2008. Risk factors for initial respiratory disease in
 3088 United States' feedlots based on producer-collected daily morbidity counts. *Can. Vet. J.*
 3089 49, 373–378. PMID: PMC2275341
 3090 Sarica, A., Cerasa, A., Quattrone, A., 2017. Random forest algorithm for the classification of
 3091 neuroimaging data in Alzheimer's Disease: a systematic review. *Front. Aging Neurosci.*
 3092 9. <https://doi.org/10.3389/fnagi.2017.00329>
 3093 Schneider, M., Tait Jr., R., Busby, W., Reecy, J., 2009. An evaluation of bovine respiratory
 3094 disease complex in feedlot cattle: impact on performance and carcass traits using
 3095 treatment records and lung lesion scores. *J. Anim. Sci.* 87, 1821–1827.
 3096 <https://doi.org/10.2527/jas.2008-1283>
 3097 Seeger, J., Grotelueschen, D., Stokka, G., Sides, G., 2008. Comparison of feedlot health,
 3098 nutritional performance, carcass characteristics and economic value of unweaned beef
 3099 calves with an unknown health history and weaned beef calves receiving various herd-of-
 3100 origin health protocols. *Bov. Pract.* 42, 27–40. <https://doi.org/10.21423/bovine->
 3101 [vol42no1p27-40](https://doi.org/10.21423/bovine-vol42no1p27-40)
 3102 Shipe, M.E., Deppen, S.A., Farjah, F., Grogan, E.L., 2019. Developing prediction models for
 3103 clinical use using logistic regression: an overview. *J. Thorac. Dis.* 11, S574–S584.
 3104 <https://doi.org/10.21037/jtd.2019.01.25>
 3105 Song, Y., Lu, Y., 2015. Decision tree methods: applications for classification and prediction.
 3106 *Shanghai Archives of Psychiatry* 27, 130–135. <https://doi.org/10.11919/j.issn.1002->
 3107 [0829.215044](https://doi.org/10.11919/j.issn.1002-0829.215044)
 3108 Tennant, T., Ives, S., Harper, L., Renter, D., Lawrence, T., 2014. Comparison of tulathromycin
 3109 and tilmicosin on the prevalence and severity of bovine respiratory disease in feedlot
 3110 cattle in association with feedlot performance, carcass characteristics, and economic
 3111 factors. *J. Anim. Sci.* 92, 5203–5213. <https://doi.org/10.2527/jas.2014-7814>
 3112 Tune Model Hyperparameters, 2024. . Microsoft Machine Learning Blog. URL
 3113 [https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/tune-](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/tune-model-hyperparameters?view=azureml-api-2)
 3114 [model-hyperparameters?view=azureml-api-2](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/tune-model-hyperparameters?view=azureml-api-2) (accessed 11.22.24).
 3115 Two-Class Boosted Decision Tree Component, 2024. . Microsoft Machine Learning Blog. URL
 3116 [https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-boosted-decision-tree?view=azureml-api-2)
 3117 [class-boosted-decision-tree?view=azureml-api-2](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-boosted-decision-tree?view=azureml-api-2) (accessed 11.25.24).

3118 Two-Class Decision Forest component, 2024. . Microsoft Machine Learning Blog. URL
 3119 [https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-decision-forest?view=azureml-api-2)
 3120 [class-decision-forest?view=azureml-api-2](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-decision-forest?view=azureml-api-2) (accessed 11.25.24).
 3121 Two-Class Logistic Regression component, 2024. . Microsoft Machine Learning Blog. URL
 3122 [https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-logistic-regression?view=azureml-api-2)
 3123 [class-logistic-regression?view=azureml-api-2](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-logistic-regression?view=azureml-api-2) (accessed 12.2.24).
 3124 Two-Class Neural Network component, 2024. . Microsoft Machine Learning Blog. URL
 3125 [https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-neural-network?view=azureml-api-2)
 3126 [class-neural-network?view=azureml-api-2](https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-neural-network?view=azureml-api-2) (accessed 11.25.24).
 3127 USDA AMS Livestock, Poultry & Grain Market News, 2015a. Kansas Weekly Cattle Auction
 3128 Summary [dataset].
 3129 USDA AMS Livestock, Poultry & Grain Market News, 2015b. National Weekly Grain Co-
 3130 Products Report [dataset].
 3131 USDA NASS, 2011. Cattle Death Loss. USDA, Washington D.C., USA.
 3132 Vergara, C., Doepfer, D., Cook, N., Nordlund, K., McArt, J., Nydam, D., Oetzel, G., 2014. Risk
 3133 factors for postpartum problems in dairy cows: explanatory and predictive modeling. J.
 3134 Dairy Sci. 97, 4127–4140. <https://doi.org/10.3168/jds.2012-6440>
 3135 Vieira de Sousa, R., Vinicius da Silva Rodrigues, A., Gomes de Abreu, M., Tabile, R., Martello,
 3136 L., 2018. Predictive model based on artificial neural network for assessing beef cattle
 3137 thermal stress using weather and physiologic variables. Comput. Electron. Agric. 144,
 3138 37–43. <https://doi.org/10.1016/j.compag.2017.11.033>
 3139 Wang, X., Gao, H., Gebremedhin, K., Bjerg, B., Van Os, J., Tucker, C., Zhang, G., 2018. A
 3140 predictive model of equivalent temperature index for dairy cattle (ETIC). J. Therm. Biol.
 3141 76, 165–170. <https://doi.org/10.1016/j.jtherbio.2018.07.013>
 3142 Wisnieski L., Amrine D., Renter D., 2021. Predictive modeling of bovine respiratory disease
 3143 outcomes in feedlot cattle: a narrative review. Livest. Sci. 251, 104–117.
 3144 <https://doi.org/10.1016/j.livsci.2021.104666>
 3145 World Health Organization, 2017. WHO guidelines on use of medically important antimicrobials
 3146 in food-producing animals [WWW Document]. World Health Organization. URL
 3147 <https://www.who.int/publications/i/item/9789241550130> (accessed 12.10.24).
 3148



3

3150 **Figure 5.1.** Flowchart demonstrating the different variables, datasets, and number of cohorts
 3151 included in each dataset.

3152

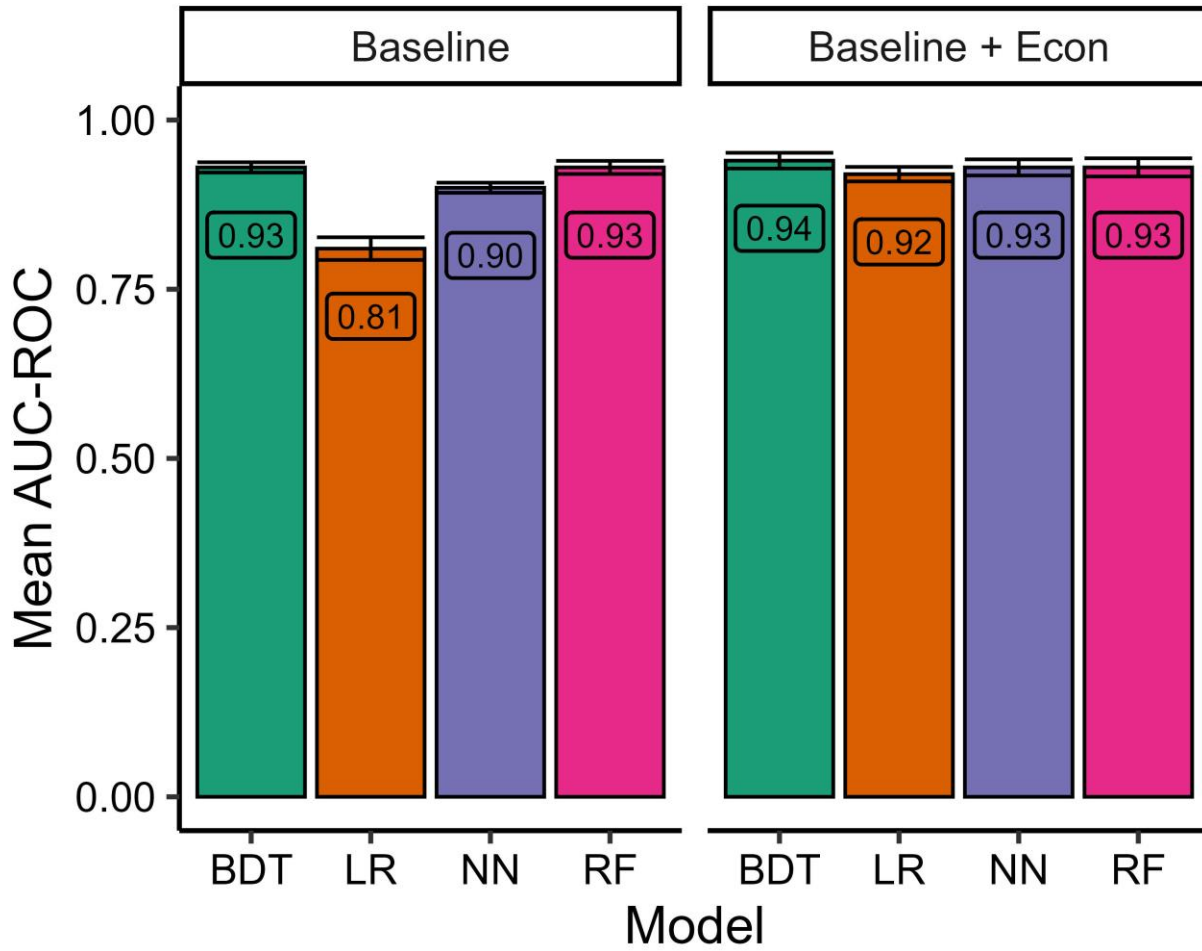


Figure 5.2. Bar chart demonstrating the average AUC-ROC values for the Baseline and the Baseline + Econ models. Error bars represent the standard deviation. BDT = boosted decision tree, LR = logistic regression, NN = neural network, RF = random forest.

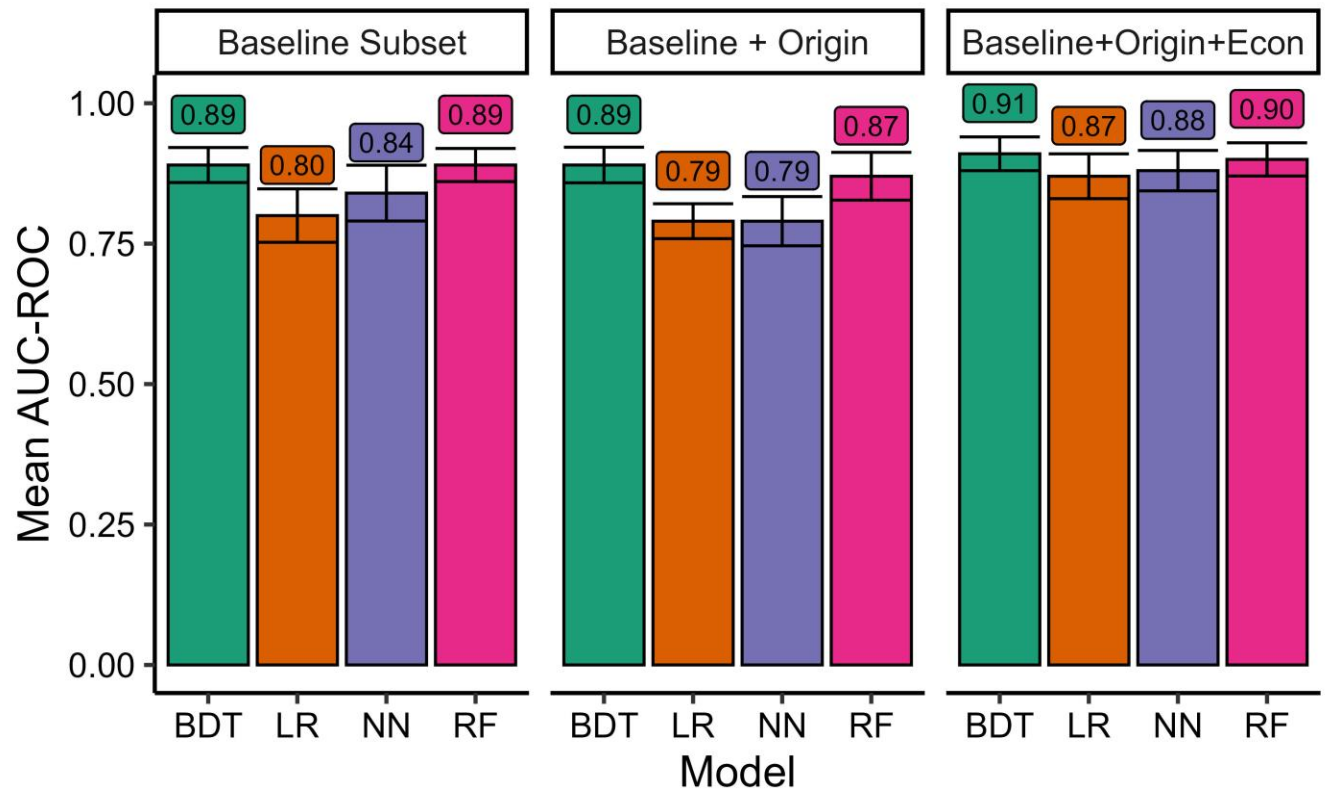


Figure 5.3. Bar chart demonstrating the average AUC-ROC values for the Baseline Subset, Baseline + Origin, and Baseline + Origin + Econ models. Error bars represent the standard deviation. BDT = boosted decision tree, LR = logistic regression, NN = neural network, RF = random forest

3164 **Table 5.1.** Table of descriptive statistics of selected variables split by metaphylaxis choice made
3165 by feedyard and economic model.

Variable	Best Option Economic Model		Feedyard Actual	
	<i>Meta-phylaxis</i>	<i>No Metaphylaxis</i>	<i>META¹</i>	<i>NO META²</i>
Total Cohorts	4,672	12,297	2,594	13,774
Sex: % Heifer Cohorts (% Steer Cohorts)	49 (51)	43 (57)	43 (57)	45 (55)
Average Head Received	140	183	172	173
Average Arrival Weight, lbs (kg)	724 (328)	774 (351)	681 (309)	777 (352)
Average Exit Weight, lbs (kg)	1,297 (588)	1,328 (602)	1,290 (585)	1,326 (601)
Average DOF³	172	163	187	162
Average BRD⁴ Morbidity Risk, %	29.2	9.9	27.6	12.3
Average BRD⁴ Mortality Risk, %	4.6	0.4	2.0	1.4

3166 ¹ Meta = metaphylaxis

3167 ²No Meta = no metaphylaxis

3168 ³DOF = days on feed

3169 ⁴BRD = bovine respiratory disease

3170 **Table 5.2.** Table showing all performance metrics for each model series and model run. Models
3171 included boosted decision tree, logistic regression, neural network, and random forest.

Model Series	Model	Average AUC-ROC ¹	St. Dev. ² AUC-ROC ¹	Sn(%) ³	Sp(%) ⁴	PPV ⁵	NPV ⁶
Feedyard Actual				24.6	87.0	0.39	0.78
Baseline	BDT ⁷	0.93	0.01	91.0	73.9	0.90	0.77
	LR ⁸	0.81	0.02	87.9	36.0	0.78	0.55
	NN ⁹	0.90	0.01	91.3	64.3	0.87	0.74
	RF ¹⁰	0.93	0.01	91.0	73.1	0.89	0.76
Baseline + Econ	BDT	0.94	0.01	91.2	75.9	0.90	0.77
	LR	0.92	0.01	90.4	71.0	0.89	0.74

	NN	0.93	0.01	90.5	74.2	0.90	0.76
	RF	0.93	0.01	89.7	75.8	0.90	0.75
Baseline Subset	BDT	0.89	0.03	87.5	67.2	0.86	0.72
	LR	0.80	0.05	86.0	38.2	0.76	0.54
	NN	0.84	0.05	85.8	62.2	0.84	0.68
	RF	0.89	0.03	89.3	65.2	0.85	0.74
Baseline + Origin	BDT	0.89	0.03	87.8	66.3	0.85	0.71
	LR	0.79	0.03	86.0	39.1	0.76	0.55
	NN	0.79	0.04	80.0	55.2	0.81	0.58
	RF	0.87	0.04	88.7	56.3	0.82	0.69
Baseline + Origin + Econ	BDT	0.91	0.03	88.1	68.6	0.87	0.73
	LR	0.87	0.04	89.0	61.2	0.83	0.73
	NN	0.88	0.04	86.7	69.4	0.86	0.71
	RF	0.90	0.03	90.1	66.7	0.86	0.75

3172 ¹AUC-ROC = area under the receiver operating characteristics curve

3173 ²St. Dev. = standard deviation

3174 ³SN = sensitivity

3175 ⁴SP = specificity

3176 ⁵PPV = positive predictive value

3177 ⁶NPV = negative predictive value

3178 ⁷BDT = boosted decision tree

3179 ⁸LR = logistic regression

3180 ⁹NN = neural network

3181 ¹⁰RF = random forest

3182

Chapter 6 - Evaluation of 3-, 6-, and 9-day post-treatment intervals following pradofloxacin for treatment of bovine respiratory disease in stocker cattle

L. Heinen¹, B. V. Lubbers¹, R.K. Tessman², R.G. Amachawadi¹, B.J. White¹

¹Beef Cattle Institute, Kansas State University, Manhattan, KS 66506, USA

²Elanco Animal Health, Greenfield, IN 45140, USA

This research is published in the Bovine Practitioner journal and is reproduced with permission from publisher as formatted in referenced publication. Lilli Heinen, Brian V. Lubbers, Ronald K. Tessman, Raghavendra G. Amachawadi, Brad J. White. Evaluation of 3-, 6-, and 9-day post-treatment intervals following pradofloxacin for treatment of bovine respiratory disease in stocker cattle, Bovine Practitioner Volume 59, Issue 2, August 2025,

Abstract

Pradofloxacin is a fluroquinolone antimicrobial approved for treatment of bovine respiratory disease (BRD) in cattle in the United States. The primary objective of this study was to compare different post-treatment intervals (PTI) following therapy with pradofloxacin on first treatment success and case fatality in stocker cattle naturally affected with BRD. The days to death following enrollment was evaluated as a secondary outcome. Cattle at five commercial stocker operations were identified with BRD by site personnel and were randomized in a 1:1:1 ratio to 3-, 6-, or 9-day PTI within each cohort. Following treatment with pradofloxacin, cattle were followed for 45 days to monitor health outcomes. First treatment success was defined as no additional treatment for BRD and no mortality. Case fatality was defined as death due to any cause. Generalized linear mixed models were used for statistical analysis. Weight at treatment and rectal temperature were not different between PTI groups. First treatment success ($p = 0.51$) and case fatality ($p = 0.52$) were not different among PTI treatment groups. Days to death were not different between PTI treatment groups ($p = 0.20$). These results will assist with determining optimal use of pradofloxacin in stocker cattle.

Keywords: antimicrobial, BRD, health

Introduction

Pradofloxacin (Pradalex™)^a is a third-generation fluoroquinolone approved in the United States for the treatment of bovine respiratory disease (BRD) associated with *Mannheimia haemolytica*, *Pasteurella multocida*, *Histophilus somni* and *Mycoplasma bovis* in beef and dairy cattle between 2-12 months of age intended for slaughter or in breeding cattle less than 1 year of age. Pradofloxacin is a bactericidal antimicrobial agent with activity against Gram-positive and Gram-negative bacteria as well as *Chlamydia* and *Mycoplasma* species. Additionally, its spectrum also includes anaerobic bacterial species.¹ Establishment of pharmacokinetic and antimicrobial resistance performance is important for evaluation of an antimicrobial for treatment of disease.

Pharmacokinetic studies in cattle infected with experimentally induced BRD have demonstrated that therapeutic concentrations in the pulmonary epithelial lining fluid (PELF) are reached in 7 hours following a single injection of pradofloxacin at 10 mg/kg.² To date, no pharmacokinetic studies have been published in cattle treated with pradofloxacin for naturally occurring BRD. Pharmacokinetic data allows for understanding of therapeutic concentrations in the body and determination of appropriate dosage; however, determination of the optimal interval between treatment administration and evaluation of the treatment's success, called the post-treatment interval (PTI), should be based on clinical data.

In vitro analysis of bovine isolates of *M. haemolytica* and *P. multocida* reports mutant prevention concentrations (MPCs) of less than 0.063 microgram per milliliter. This is four-fold below the reported susceptibility breakpoint for pradofloxacin.¹ The same study reports that all 41 tested strains of *M. haemolytica* and *P. multocida* demonstrated minimum inhibitory concentration (MIC) values below the breakpoint for pradofloxacin.¹ These findings show that pradofloxacin has favorable in vitro activity against two main BRD pathogens and a low likelihood for antimicrobial resistance selection. As antimicrobial resistant isolates from BRD cases are reported more frequently, minimizing antimicrobial exposure may slow the progression of resistance.² Establishment of an optimal PTI for pradofloxacin in cattle treated for BRD would ensure adequate time for resolution of clinical signs and reduce unnecessary antimicrobial use.

While understanding pharmacological and antimicrobial resistance properties of an antimicrobial are important, clinical performance and determination of treatment success or failure is equally important. To date, a single study has compared the efficacy of pradofloxacin to tulathromycin and florfenicol in steers experimentally induced with BRD.³ This study demonstrated no difference in time to resolution of clinical signs in the treatment groups. When evaluating clinical success or failure of a treatment, emphasis is placed on the PTI. Single injection antimicrobial therapy available for BRD treatment has placed even more importance on PTIs. Many studies have investigated the impact of different PTIs on the treatment outcome using other single injection antimicrobials in cattle diagnosed with BRD; however, no studies have been published investigating PTIs following pradofloxacin treatment for BRD in cattle.⁴⁻⁸

The primary study objective was to evaluate the effect of 3-, 6- and 9-day post-treatment intervals (PTI) on first treatment success and case fatality in naturally occurring bovine respiratory disease (BRD) in stocker calves. Secondary outcomes included days-to-death (DTD) in the same population. Our hypothesis was that the 6- and 9-day PTI treatment groups will demonstrate improved first treatment success rates and case fatality rates as compared to the 3-day group. Additionally, we hypothesized there would be no differences in DTD between PTI treatment group case fatalities.

Materials and Methods

This study was performed at 5 commercial stocker operations in Kansas and Indiana. The study began in October 2024 and concluded in February 2025. All procedures were approved by the Kansas State University Institutional Animal Care and Use Committee prior to initiation of the study (Protocol Number 5055).

Sample size calculation

First treatment success was used for sample size estimates. Sample size was calculated to detect a difference of 15% between the 3-day group (estimated 70% first treatment success) and the 9-day group (estimated 85% first treatment success). A 95% confidence interval (alpha set at 0.05) was selected with beta set to 0.80. Estimated sample size was calculated at 118 per treatment group. Study numbers were inflated by 5% and rounded up to account for missing data

and loss to follow-up resulting in 125 animals per treatment group to achieve desired statistical power.

Study population

Cohorts of cattle arriving at each operation were required to meet certain criteria. The cohorts could only contain animals greater than 2 months of age and only native beef breeds. Dairy breeds and beef-on-dairy crossbred cattle were not eligible for enrollment. Mixed sex cohorts and those containing intact bulls were eligible for study inclusion. At cohort arrival, all operations were required to adhere to a standard processing protocol including metaphylaxis with tulathromycin (Increxxa™^b) at the label dose of 2.5 mg/kg, vaccination for respiratory pathogens (bovine viral diarrhea type 1 and 2, infectious bovine rhinotracheitis, bovine parainfluenza 3, bovine respiratory syncytial virus, and *M. haemolytica*) (Titanium®^{5c} and Nuplura®^d) and parasite treatment with injectable moxidectin (Cydectin®^e) at 0.2 mg/kg. Additional procedures (such as castration, dehorning, BVDV PI testing, and/or administration of growth implants) were allowable based on the determination of the veterinarian of record for each study site. A mandatory 5-day post-metaphylaxis interval (PMI) was required for all study animals; animals requiring antimicrobial therapy during this period were treated at the discretion of on-site personnel, but were not eligible for enrollment in the current study.

Inclusion criteria

Following the 5-day PMI, cattle were observed daily for clinical signs of BRD using a standardized BRD case definition for enrollment. An animal was considered a BRD case if it displayed at least two of the following clinical signs: anorexia, depression, increased respiratory rate or effort, nasal discharge, cough, or a rectal temperature greater or equal to 104 degrees Fahrenheit (40 degrees Celsius). Individual cattle meeting the BRD case definition were administered a dose of pradofloxacin at 10 mg/kg and randomly assigned to a post-treatment interval (PTI).

3298 **Feed and housing**

3299 Site A is located in central Kansas and typically purchases high risk calves out of the
3300 southeast United States. Cattle for the trial were purchased from auction markets in Kentucky,
3301 Arkansas and Missouri. All enrolled cattle were steers, although some of the cohorts had 15-20%
3302 bulls at arrival. Average arrival weights ranged between 575 to 605 lbs (261 to 274 kgs). For the
3303 first 21 days on feed, cattle were fed a ration containing 14% ground hay, 14% corn, 35% wet
3304 distillers grains, 33.5% corn silage and 3.5% supplement mix. The ration changes after 21 days
3305 to incorporate 15% sorghum silage with a subsequent decrease in ground hay and corn silage.

3306 Site B is located in central Kansas and is used primarily as a grow facility for a larger
3307 feedlot operation. Pens consist of traditional dirt floor pens and grass traps. Maximum capacity at
3308 the facility is approximately 11,000 head and cattle are kept on pasture from May to August.
3309 Cattle for the trial came from 2 cohorts of native breed heifers purchased from a sale barn in
3310 Oklahoma City at average weights of 450 lbs (204 kgs) and 550 lbs (249 kgs). At arrival, cattle
3311 started on a ration consisting of 22% rolled corn, 6.5% sorghum silage, 38% ground hay, 30%
3312 modified distillers grains, and 4% liquid protein. At this site, cattle remain on the starter ration
3313 for 3 weeks before beginning a step-up program to the final grow ration which contains a higher
3314 corn and modified distillers percentage. Cattle typically remain at this site for 90 days before
3315 moving to the finishing feedlot.

3316 Site C starts 1,500 calves per year on grass in the Flint Hills region of Kansas. Calves
3317 receive brome and prairie hay ad libitum and the starter feed contains soybean hulls, dry distillers
3318 grains, oats and corn. Calves are fed until they are eating 2% of their body weight in hay and 1%
3319 of their body weight in starter feed. Typically, this operation feeds cattle from mid-October to
3320 mid-March. Trial calves at this site were all heifer calves sourced from a southwest Missouri
3321 auction market and shipped to central Kansas. Cattle weighed an average of 504 lb. (229 kgs) at
3322 arrival and all tested negative for persistent bovine viral diarrhea virus infection (BVDV PI)
3323 using an ear notch test.

3324 Site D purchases cattle from livestock auction markets year-round to start on grass in
3325 central Kansas. Cattle had access to 30-acre pastures and were fed prairie hay ad libitum along
3326 with dry distillers grains and cracked corn in a 1:1 ratio. The cattle purchased for this study were
3327 purchased from northern Alabama and southern Tennessee auction markets over 3 days, most as

singles or pairs, and transported for 12 hours. Approximately 80% of the cattle were bulls with the remaining 20% as steers. At arrival, cattle weighed 630 lbs (286 kgs) on average.

Site E is a row-crop and stocker operation in Kentucky. Cattle are purchased from several regional auction markets and transported to the facility in groups throughout the year. Cattle have access to grass pastures attached to dirt floor pens.

Site F feeds 6,000 calves per year, keeping about 1,200 head on site at one time, year-round. Pens are approximately 15 acres each and are a combination of dirt floor and pasture. Cattle are fed a ration consisting of hay, corn silage, wet distillers grain, hominy and mineral mix. Cattle for the study were sourced from multiple sale barns in Kentucky, Tennessee and Georgia in October. They were placed in groups of 70-80 head upon arrival to the operation in Indiana with an average arrival weight of 480 lbs (218 kgs). All cattle from this site were steers.

No feed-grade antimicrobials were administered in the rations at any sites. Other feed additives such as melengesterol acetate or amprolium were allowed, if necessary. Water access was provided ad libitum through automatic waterers at each operation. Enrollment was not restricted by site. Animals were enrolled on a rolling basis until a total of 400 animals was enrolled.

Enrollment

Animals were randomly assigned in a 1:1:1 ratio within each cohort to one of three PTI groups: 3-day, 6-day, and 9-day. Study participants were given an enrollment form to use to assign animals to one PTI group in groups of three. Complete blinding of study personnel to treatment group was not possible given that the same individuals were enrolling animals in the study and administering treatments.

A random number generator (Excel) was used to create a random number for each blank on the enrollment form. For each group of three rows on the enrollment form, the smallest number was assigned to the 3-day PTI, the largest number was assigned to the 9-day PTI, and the middle number was assigned to the 6-day PTI. The first calf meeting the BRD case definition was assigned to the first blank. The next calf from the same cohort was assigned to the second blank. This randomization strategy was performed in an effort to distribute PTI groups evenly between cohorts. Ear tags were used to denote the date the calf would be eligible for treatment. No information about the date the calf was originally pulled was included on the ear tag. Weight at the time of treatment and rectal temperature were recorded.

3359

3360 **Observation period**

3361 Following drug administration and assignment to PTI group, cattle were returned to their
3362 home pen and followed for 45 days to monitor for subsequent health outcomes. Following the
3363 assigned PTI, cattle were eligible for retreatment if they met the initial case definition for BRD.
3364 Animals that required additional treatment(s) were treated according to individual operation
3365 protocol but were not treated with pradofloxacin again. Health outcomes of interest were: first
3366 treatment success (FTS), case fatality (CF), and days to death (DTD). First treatment success
3367 (FTS) and case fatality (CF) were determined at the individual animal level and converted to
3368 binary outcomes. Treatment success was defined as not requiring additional treatment for BRD
3369 and not experiencing mortality within 45 days of initial treatment. Case fatality was defined as
3370 mortality due to any cause within 45 days of initial treatment. The variable characterizing the
3371 number of days between initial treatment and death for those animals experiencing a case fatality
3372 was left as a continuous numerical variable and was calculated as days to death (DTD). Cattle for
3373 which 45-day observation data was unavailable and cattle which were treated before their
3374 assigned PTI was over were removed from final analysis. Post-mortem examinations were not
3375 performed consistently across sites to determine cause of death. As such, an animal was
3376 considered a case fatality if it died following enrollment without regards for true cause of death.

3377

3378 **Statistical analysis**

3379 Data management steps and subsequent statistical analysis were carried out in R
3380 software.⁹ Sex was not included in the models as it was confounded by site and cohort. Weight
3381 and rectal temperature were categorized for analysis. Weight was categorized into four groups:
3382 450 lbs. (204 kgs) or less, 451-550 lbs. (205-249 kgs.), 551-650 lbs. (250-295 kgs.), and greater
3383 than 650 lbs. (295 kgs). Temperature was categorized: normal (below 104 degrees Fahrenheit, 40
3384 degrees Celsius), and febrile (greater than or equal to 104 degrees Fahrenheit, 40 degrees
3385 Celsius) based on pre-established case definition.

3386 Generalized linear mixed models were used for statistical analysis of FTS and CF with
3387 binomial distributions. To model DTD, a generalized linear mixed model with a Poisson
3388 distribution and log link function was utilized. In all models, a random effect of cohort nested

within site was included to account for lack of independence. Fixed effects included rectal temperature category, weight at enrollment category, and PTI group. Rectal temperature and weight categories remained in the model as experimental design factors. Pairwise comparisons for PTI group were performed using Tukey's adjustment.

Results

A total of 446 animals were enrolled in the study. One animal was removed as it was enrolled twice, and 14 animals were removed due to re-treatment before their assigned PTI was completed. Site E was removed less than a week from the start of the study due to a higher-than-expected morbidity and labor constraints. All animals enrolled at Site E were removed from the study ($n = 25$). Table 1 displays the total number of animals enrolled at each site excluding the 40 animals that were removed.

Descriptive characteristics for the 406 enrolled cattle are presented by treatment group in Table 2. A chi-square test of independence was applied to determine differences in sex between PTI treatments groups. There were no statistical differences between the counts of heifers, steers, or bulls between the PTI treatment groups ($p\text{-value} > 0.05$). Site B experienced an outbreak of nervous coccidiosis approximately half-way through the trial. The affected pens of cattle received a 5-day course of amprolium pellets at 454 mg per 100 lb. of body weight. Three calves enrolled in the trial demonstrated neurologic signs and were independently treated with oral sulfa boluses and thiamine (vitamin B1) injection ($n = 3$). The treatment for nervous coccidiosis did not meet criteria for BRD re-treatment; however, two calves subsequently died following treatment, and these were included as treatment failures and case fatalities. The surviving calf was considered a treatment success as it did not require additional treatment for BRD and did not experience mortality.

First treatment success

First treatment success risk across all study sites was 61.8%. There were no statistically significant differences in first treatment success between PTI treatment groups ($p = 0.51$). Other fixed effects including rectal temperature category and weight at treatment category were not

statistically associated with first treatment success. Numerically, the 9-day PTI group had the poorest treatment success risk and the 6-day group had the best risk.

Case fatality and days to death

There were also no significant differences in case fatality between PTI treatment groups ($p = 0.52$). Fixed effects including rectal temperature category and treatment weight category were not statistically associated with case fatality. Days to death was not statistically different ($p = 0.20$) between the three treatment groups. Table 3 reports the model-estimated means and standard errors of the mean (SEM) by treatment groups for the three outcome measures (FTS, CF, and DTD).

Discussion

To the authors' knowledge, this is the first study to investigate PTIs for pradofloxacin in naturally occurring BRD. Furthermore, it is one of the first investigations evaluating PTIs in stocker calves for any antimicrobial. Results of this study indicate that there are no differences in FTS, CF and DTD between the investigated PTIs. These results help provide recommendations for antimicrobial use in stocker calves treated for naturally occurring BRD.

Previous studies in feedlot cattle with different antimicrobials have demonstrated the value of PTIs for improving treatment outcomes.⁴⁻⁸ A multi-site investigation in feedlot cattle comparing 3-, 5- and 7-day PTIs following administration of ceftiofur crystalline free acid (CCFA) for BRD found statistically significant ($P < 0.05$) differences in 28-day treatment success rates between the 3- (65.6%) and 7-day (77.5%) groups. There were no differences in treatment success between other PTI groups and no differences in mortality rates between PTI groups.⁵ Another study evaluating CCFA in feedlot cattle was repeated over a 56-day observation period.⁶ This study found improved treatment success risk among the 7-day group as compared to the 3-day group. There were no other differences in treatment success and no differences in mortality between treatment groups.⁶ The CCFA studies demonstrated similar findings despite evaluating treatment outcomes over different lengths of time (28 days vs. 56 days).

A study evaluating PTIs in feedlot cattle treated for BRD with tilmicosin found a significant ($P = 0.05$) difference in first treatment success rates between the 3- (67.9%) and the 9-day (86.9%) PTI groups. The study found no significant ($P = 0.39$) differences in case fatality between PTI groups.⁷ These 3 studies demonstrate that first treatment success rates were improved by implementing PTIs greater than 3 days following initial treatment. Extension of the PTI to 7 or even 9 days in the aforementioned studies likely allowed the cattle in these studies to return back to a “healthy” appearance, resulting in fewer second or third treatments.

Post-treatment intervals of 7, 10 and 14 days were evaluated following treatment of BRD in feedlot cattle with tulathromycin.⁴ Statistical analysis found no significant ($P > 0.05$) differences for first treatment success and case fatality rates between PTI treatment groups. This research demonstrated that longer PTIs (greater than 7 days) had no impact on treatment success or mortality rates. A study by Theurer et al. evaluated prolonged PTI intervals following therapy with gamithromycin.⁸ The study compared 3-, 6-, 9- and 12-day PTIs in feedlot cattle. Differences in first treatment success ($P = 0.012$) and case fatality ($P = 0.032$) between the PTI treatment groups were notable. Specifically, differences in first treatment success were seen between the 3-day (56.8%) and 9-day (72.3%) groups. Differences in case fatality were observed between the 9-day (7.2%) and the 12-day (15.0%) groups. While previous BRD PTI studies demonstrated that longer PTIs improved clinical outcomes, the gamithromycin PTI study demonstrated that there is likely an upper limit to PTIs at which point clinical outcomes are negatively impacted.⁸

The current study found no statistically significant differences in case fatality rates between treatment groups. This finding is consistent with other reported PTI studies with the exception of the study by Theurer et al.⁸ However, the longest PTI investigated in the current was 9 days whereas the Theurer et al. study found a significantly higher case fatality in their 12-day PTI group. The current study also found no significant differences in first treatment success, a finding that is at odds with other studies.⁵⁻⁸ Original sample size calculations anticipated a greater difference in first treatment success risk between the 3- and 9-day group, so it is likely the current study is underpowered. Additionally, a permissive BRD case definition was used, meaning that the criteria for initial enrollment and retreatment were easily met. Doing so allowed for high external validity, as this is representative of current standard practice in the industry; however, it may have resulted in the treatment of animals that would have otherwise recovered

without intervention. This could have led to higher treatment success rates and lower case fatality rates. Previous PTI studies utilized BRD case criteria for enrollment that differed from what was used in the current study which may have contributed to differences in study findings.^{7,8}

At the time of this writing, only one study has been published evaluating the efficacy of pradofloxacin compared to other antimicrobials for the treatment of BRD. As a secondary outcome, Foster et al. evaluated clinical outcomes in dairy steers challenged with *M. haemolytica* and subsequently treated with pradofloxacin, tulathromycin or florfenicol.³ All steers demonstrated rapid resolution of clinical signs following treatment administration and there were no statistically significant ($P > 0.05$) differences between groups with regard to depression score, respiratory score or rectal temperature from day of treatment to study endpoint (all steers were humanely euthanized on day 12 post-treatment).

DTD has not been evaluated in case fatalities in previous PTI studies. Motivation for investigation of DTD as an outcome was brought about by producer concerns that adhering to longer PTIs results in cattle dying later due to delayed retreatment. Median values for the current study's case fatality were all below 20 days with a maximum value of 43. Smith et al. reported on DTD following initial BRD treatment in feedlot cattle across several regions in the United States.¹⁰ Their first and third quartile values are 4 and 43, closely matching our data. The current study demonstrated no significant differences in DTD between PTI treatment groups, indicating that the PTI investigated in this study do not impact timing of death. Furthermore, the distribution of DTD is similar to that reported in the literature for feedlot cattle.

Many meta-analyses have been published comparing efficacy of several antimicrobials for the treatment and control of BRD.¹¹⁻¹⁵ The meta-analyses were published before the approval for pradofloxacin for treatment of BRD but most report on the comparative efficacy of other fluoroquinolone antimicrobials. The meta-analyses summarize that both enrofloxacin and danofloxacin result in lower re-treatment risk compared to other antimicrobials except tulathromycin.¹²⁻¹⁴ It is important to mention, however, that the meta-analyses reported here do not account for a PTI which is needed to direct comparisons between antimicrobials when evaluating efficacy.

One of the major strengths of this study is the external validity of the study population. This study was conducted in commercial operations with cattle procurement, feeding and

management typical of current stocker cattle industry practices. Although processing and health evaluations were standardized by study protocol, they were also representative of stocker production practices. One of the limitations of this study is the incomplete blinding of cattle health evaluators to treatment groups. It was not possible to keep all individuals at these operations blinded to the treatment groups due to labor constraints; however, each site had multiple animals enrolling each day, making it more difficult to identify which animals belonged to each PTI treatment group due to the numerous eligible retreatment dates. Another limitation of this study is the permissive BRD case definition. This could have resulted in more spontaneous recoveries, leading to higher treatment success rates and fewer case fatalities. Additionally, permissive retreatment criteria could have led to a greater number of retreatments. Together with the spontaneous recoveries, the current BRD case definition could have led to few differences between PTI treatment groups. Despite this limitation, the enrollment criteria used in the current study was an accurate reflection of BRD case definitions on these operations, further contributing to external validity. As not all of the sites in this study typically use a rectal temperature cut-off for treatment, a rectal temperature cut point of 104 °F (40 °C) or greater was only included as one of the enrollment criteria.

Evaluation of clinical outcomes following treatment is important for determination of the optimal PTI. This trial found no significant differences in first treatment success and case fatality between PTI treatment groups. Other studies have concluded that longer PTIs demonstrate favorable FTS, although the findings are not consistent in PTIs 12 days or longer. Lack of statistical differences in case fatality between treatment groups is supported by the literature. Low statistical power and a permissive BRD case definition likely contributed to the findings in the current study. DTD was not statistically different by treatment group. Results of this study should not be extrapolated to other populations of cattle (such as those in the finishing phase of production) without confirmation of these findings in additional studies. The results of this study will help practitioners determine the optimal application of pradofloxacin related to health outcomes in beef stocker cattle.

End Notes

^aPradalex, Elanco Animal Health, Greenfield, IN

^bIncrexxa, Elanco Animal Health, Greenfield, IN

^cTitanium 5, Elanco Animal Health, Greenfield, IN

^dNuplura, Elanco Animal Health, Greenfield, IN

^eCydectin, Elanco Animal Health, Greenfield, IN

References

1. Blondeau JM, Fitch SD. Comparative in vitro killing by Pradofloxacin in comparison to Ceftiofur, Enrofloxacin, Florfenicol, Marbofloxacin, Tildipirosin, Tilmicosin and Tulathromycin against bovine respiratory bacterial pathogens. *Microorganisms*. 2024; 12:996.
2. Lubbers B, Hanzlicek G. Antimicrobial multidrug resistance and coresistance patterns of *Mannheimia haemolytica* isolated from bovine respiratory disease cases – a three-year (2009-2011) retrospective analysis. *J Vet Diagn Invest*. 2013; 25:413-417.
3. Foster DM, Halleran JL, Jacob ME, et al. Pharmacokinetics of pradofloxacin, florfenicol, and tulathromycin and response to treatment of steers experimentally infected with *Mannheimia hemolytica*. *J Vet Intern Med*. 2024;39: e17270.
4. Pfizer. Efficacy of Draxxin® followed by 7-, 10-, or 14-day post-treatment intervals against naturally occurring bovine respiratory disease. Prepared from study report 1133\$-60-05-489; 2006.
5. Pfizer. Outcomes of 3-, 5-, or 7-day post-treatment intervals after a single administration of Excede®. Technical bulletin no. EXD04023; 2004.
6. Pfizer. Efficacy of Excede® followed by 3- or 7-day post-treatment intervals vs Baytril® followed by a 3-day poster treatment interval in treatment of bovine respiratory disease. Technical bulletin no. EXD04024; 2004.
7. Apley M. Treatment of calves with bovine respiratory disease: duration of therapy and posttreatment intervals. *Vet Clin North Am Food Anim Pract*. 2015;31:441-453.
8. Theurer ME, Fox JT, Portillo TA, et al. Comparison of gamithromycin post-treatment intervals for beef cattle naturally affected with bovine respiratory disease. *Bov Pract*. 2020;54:105-111.
9. RStudio Team. RStudio: integrated development for R. Boston, MA: RStudio; 2024. Available at: <http://www.rstudio.com/>.
10. Smith K, White B, Amrine D, et al. Evaluation of first treatment timing, fatal disease onset, and days from first treatment to death associated with bovine respiratory disease in feedlot cattle. *Vet Sci*. 2023; 10:204.

- 3571 11. Abell K, Theurer M, Larson R, et al. A mixed treatment comparison meta-analysis of
3572 metaphylaxis treatments for bovine respiratory disease in beef cattle. *J Anim Sci.* 2017;
3573 95:626-635.
- 3574 12. Nautrup B, Vlaenderen I, Decker M, et al. Antimicrobial drug use for control and
3575 treatment of bovine respiratory disease in U.S. feedlot cattle: a meta-analysis of
3576 comparative studies versus tulathromycin. *Bov Pract.* 2017; 51:1-13.
- 3577 13. O'Connor A, Coetzee J, da Silva N, et al. A mixed treatment comparison meta-analysis
3578 of antibiotic treatments for bovine respiratory disease. *Prev Vet Med.* 2013; 110:77-87.
- 3579 14. O'Connor A, Yuan C, Cullen J, et al. A mixed treatment meta-analysis of antibiotic
3580 treatment options for bovine respiratory disease – an update. *Prev Vet Med.* 2016;15:130-
3581 139.
- 3582 15. O'Connor A, Hu D, Totton S, et al. A systemic review and network meta-analysis of
3583 injectable antibiotic options for the control of bovine respiratory disease in the first 45
3584 days post arrival at the feedlot. *Anim Health Res Rev* 2021; 20:163-181.
- 3585
- 3586

Table 6.1. Characteristics of cattle at study enrollment by study location

<i>Site</i>	<i>Total number enrolled</i>	<i>Average weight, lbs. (kgs)</i>	<i>Sex</i>
A	147	575 (261)-605 (274)	Steer
B	155	450 (204)-550 (249)	Heifer
C	29	504 (229)	Heifer
D	44	630 (286)	Bull, Steer
F	31	480 (218)	Steer

Table 6.2. Characteristics of cattle at study enrollment based on treatment group and across the study population.

	<i>3-day PTI</i>	<i>6-day PTI</i>	<i>9-day PTI</i>	<i>Overall</i>
% Bull*	7.4 (10/136)	8.3 (11/133)	9.5 (13/137)	8.4 (34/406)
% Steer*	39.7 (54/136)	39.8 (53/133)	35.0 (48/137)	38.2 (155/406)
% Heifer*	52.2 (71/136)	51.9 (69/133)	54.7 (75/137)	53.0 (215/406)
Weight at treatment, lbs. (kgs)	547 (248)	548 (249)	546 (249)	547 (248)
Temperature at treatment, degrees Fahrenheit (degree Celsius)	104.9 (40.5)	104.9 (40.5)	104.8 (40.4)	104.9 (40.5)

Table 6.3. Outcome means ($\pm$ SEM) by treatment group from generalized linear mixed models with random intercept to account for lack of independence in cohorts within sites.

<i>Outcome</i>	<i>3-day PTI</i>	<i>6-day PTI</i>	<i>9-day PTI</i>	<i>Overall P-value</i>
First treatment success risk, %	71.24 $\pm$ 7.47	74.72 $\pm$ 6.74	68.82 $\pm$ 7.71	0.51
Case fatality risk, %	5.35 $\pm$ 2.59	6.84 $\pm$ 3.01	8.93 $\pm$ 3.80	0.52
Days to death	18.24 $\pm$ 2.47	16.28 $\pm$ 2.22	18.18 $\pm$ 2.15	0.20

Appendix A - Supplement to Chapter 3

3597 **Appendix Table A.1. Complete list of variables created or obtained to describe the data.**

Variable Category	Variable	Description
Cohort level	Arrival weight of the cohort	Average weight of the cohort on arrival
	Year of arrival	Year the cohort arrived
	Quarter of arrival	Quarter of the year that the cohort arrived (1 = Jan. – Mar., 2 = Apr. – Jun., 3 = July – Sept., 4 = Oct. – Dec.)
	Month of arrival	Month of the year that the cohort arrived
	Day of arrival	Day of the week that the cohort arrived (1 = Mon., 2 = Tues., 3 = Wed., 4 = Thurs., 5 = Fri., 6 = Sat., 7 = Sun.)
	Arrived on a _____	Binary variables (1 = yes, 0 = no) for each day of the week a cohort could have arrived
	Arrived on a holiday	Binary variable (1 = yes, 0 = no) for whether or not a cohort arrived on a holiday; includes all federal holidays
	Sex	The sex of the cohort; could be steer, heifer, or mixed
	Date of max fill	Date that the pen reached capacity (would be the same as date of arrival if the pen was filled on that day)
	DOF of max fill	Days on feed that the lot took to reach capacity (date of arrival – date of max fill)
	Total head in	Number of head in the cohort on day of arrival

	Number of pulls on the same day	Count of additional pulls for any reason in that cohort on a specific event date
	Number of pulls last week	Count of pulls for any reason in that cohort in the last 7 days before the event date
	Distance between origin and feedyard	Calculated distance in kilometers between a cohort's origin and the feedyard at which they are being fed
	Total deads in cohort	Cumulative count of mortalities for any reason in a cohort up to the event date
	Total rails in cohort	Cumulative count of rails for any reasons in a cohort up to the event date
	Origin latitude	Latitude in degrees of the cohort's origin
	Origin longitude	Longitude in degrees of the cohort's origin
Individual animal level	Year of BRD1	Year the animal received its first treatment for BRD
	Quarter of BRD1	Quarter of the year that the animal received its first treatment for BRD (1 = Jan. – Mar., 2 = Apr. – Jun, 3 = July. – Sept., 4 = Oct. – Dec.)
	Month of BRD1	Month of the year that the animal received its first treatment for BRD
	Day of BRD1	Day of the week that the animal received its first treatment for BRD (1 = Mon., 2 = Tues., 3 = Wed., 4 = Thurs., 5 = Fri., 6 = Sat., 7 = Sun.)
	Treated for BRD1 on a _____	Binary variables (1 = yes, 0 = no) for each day of the

	week an animal could have received its first treatment for BRD
Treated for BRD1 on a holiday	Binary variable (1 = yes, 0 = no) for whether or not an animal received its first treatment for BRD on a holiday; includes all federal holidays
Event DOF	Days spent on feed until first treatment for BRD (date of arrival – event date)
Temperature	Rectal temperature obtained at first treatment for BRD
Event weight	Weight at first treatment for BRD
Treated for GI previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for digestive issues prior to the first treatment for BRD
Treated for MS previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for musculoskeletal issues prior to the first treatment for BRD
Treated for repro previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for reproductive issues prior to the first treatment for BRD
Treated for other previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for any other issue prior to the first treatment for BRD
ADG at BRD1	Calculated average daily gain at event date ((event weight – arrival weight)/event DOF)

Feedyard level	Number of pulls on the same day at feedyard	Count of additional treatments at the yard on the same event date
	Number of pulls last week at feedyard	Cumulative count of all treatments at the yard in the last 7 days prior to the event date
	Total cattle on feed at yard	Total count of cattle on feed at the feedyard on the event date
	Moving averages of pulls	Moving averages of the number of treatments at the feedyard in 3-, 5-, 7-, 14-, 21-, and 28-day increments prior to day of first treatment for BRD
	Divergence from moving average	The difference in the event date's number of treatments from each of the increments of moving averages of treatments at the feedyard
	Number of cattle that arrived on same day	Count of total head in the yard for a given arrival date

3598 BRD1 indicates the first treatment for BRD.

3599

Appendix B - Supplement to Chapter 4

3600

Appendix Table B.1. Complete list of variables included in both the BRD1 and BRD2 datasets.

3601

Variable Category	Variable	Description
Cohort level	Arrival weight of the cohort	Average weight of the cohort on arrival
	Year of arrival	Year the cohort arrived
	Quarter of arrival	Quarter of the year that the cohort arrived (1 = Jan. – Mar., 2 = Apr. – Jun., 3 = July – Sept., 4 = Oct. – Dec.)
	Month of arrival	Month of the year that the cohort arrived
	Day of arrival	Day of the week that the cohort arrived (1 = Mon., 2 = Tues., 3 = Wed., 4 = Thurs., 5 = Fri., 6 = Sat., 7 = Sun.)
	Arrived on a _____	Binary variables (1 = yes, 0 = no) for each day of the week a cohort could have arrived
	Arrived on a holiday	Binary variable (1 = yes, 0 = no) for whether or not a cohort arrived on a holiday; includes all federal holidays
	Sex	The sex of the cohort; could be steer, heifer, or mixed
	Month of max fill	Month of the year that the pen reached capacity (would be the same as date of arrival if the pen was filled on that day)
	Day of max fill	Day of the week that the pen reached capacity (1 = Mon., 2 = Tues., 3 = Wed., 4 = Thurs., 5 = Fri., 6 = Sat., 7 = Sun.)

	DOF of max fill	Days on feed that the lot took to reach capacity (date of arrival – date of max fill)
	Total head in	Number of head in the cohort on day of arrival
	Number of treatments on the same day	Count of additional treatments for any reason in that cohort on a specific event date
	Number of treatments last week	Count of treatments for any reason in that cohort in the last 7 days before the event date
	Number of treatments to date	Count of treatments for any reason in that cohort since it arrived on the feedyard
	Distance between origin and feedyard	Calculated distance in kilometers between a cohort's origin and the feedyard at which they are being fed
	Total deads in cohort	Cumulative count of mortalities for any reason in a cohort up to the event date
	Total rails in cohort	Cumulative count of animals culled (railed) for any reason in a cohort up to the event date
	Origin latitude	Latitude in degrees of the cohort's origin
	Origin longitude	Longitude in degrees of the cohort's origin
Individual animal level	Year of BRD1	Year the animal received its first treatment for BRD
	Quarter of BRD1	Quarter of the year that the animal received its first treatment for BRD (1 = Jan. – Mar., 2 = Apr. – Jun, 3 = July. – Sept., 4 = Oct. – Dec.)

Month of BRD1	Month of the year that the animal received its first treatment for BRD
Day of BRD1	Day of the week that the animal received its first treatment for BRD (1 = Mon., 2 = Tues., 3 = Wed., 4 = Thurs., 5 = Fri., 6 = Sat., 7 = Sun.)
Treated for BRD1 on a _____	Binary variables (1 = yes, 0 = no) for each day of the week an animal could have received its first treatment for BRD
Treated for BRD1 on a holiday	Binary variable (1 = yes, 0 = no) for whether or not an animal received its first treatment for BRD on a holiday; includes all federal holidays
Event DOF	Days spent on feed until first treatment for BRD (date of arrival – event date)
Temperature	Rectal temperature obtained at first treatment for BRD
Event weight	Weight at first treatment for BRD
Treated for GI previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for digestive issues prior to the first treatment for BRD
Treated for MS previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for musculoskeletal issues prior to the first treatment for BRD
Treated for repro previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for

		reproductive issues prior to the first treatment for BRD
	Treated for other previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for any other issue prior to the first treatment for BRD
	ADG at BRD1	Calculated average daily gain at event date ((event weight – arrival weight)/event DOF)
Feedyard level	Number of treatments on the same day at feedyard	Count of additional treatments at the yard on the same event date
	Number of treatments last week at feedyard	Cumulative count of all treatments at the yard in the last 7 days prior to the event date
	Moving averages of treatments	Moving averages of the number of treatments at the feedyard in 3-, 5-, 7-, 14-, 21-, and 28-day increments prior to day of first treatment for BRD
	Divergence from moving average	The difference in the event date's number of treatments from each of the increments of moving averages of treatments at the feedyard

3602 BRD1 indicates the first treatment for BRD. BRD2 indicates the second treatment for BRD.

3603 **Appendix Table B.2. Complete list of additional variables included in only the BRD2**
3604 **dataset.**

Variable Category	Variable	Description
Cohort level	Number of pulls on the same day	Count of additional pulls for any reason in that cohort on the day of BRD2
	Number of pulls last week	Count of pulls for any reason in that cohort in the last 7 days before BRD2

	Updated number of treatments to date	Updated count of treatments for any reason in that cohort since it arrived on the feedyard
	Updated total deads in cohort	Updated cumulative count of mortalities for any reason in a cohort up to BRD2
	Updated total rails in cohort	Updated cumulative count of rails for any reasons in a cohort up to BRD2
Individual animal level	Year of BRD2	Year the animal received its second treatment for BRD
	Quarter of BRD2	Quarter of the year that the animal received its second treatment for BRD (1 = Jan. – Mar., 2 = Apr. – Jun, 3 = July. – Sept., 4 = Oct. – Dec.)
	Month of BRD2	Month of the year that the animal received its second treatment for BRD
	Day of BRD2	Day of the week that the animal received its second treatment for BRD (1 = Mon., 2 = Tues., 3 = Wed., 4 = Thurs., 5 = Fri., 6 = Sat., 7 = Sun.)
	Treated for BRD2 on a _____	Binary variables (1 = yes, 0 = no) for each day of the week an animal could have received its second treatment for BRD
	Treated for BRD2 on a holiday	Binary variable (1 = yes, 0 = no) for whether or not an animal received its second treatment for BRD on a holiday; includes all federal holidays
	Second event DOF	Days spent on feed until second treatment for BRD (date of arrival – event date)

Second event temperature	Rectal temperature obtained at second treatment for BRD
Second event weight	Weight at second treatment for BRD
Updated treated for GI previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for digestive issues prior to the second treatment for BRD
Updated treated for MS previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for musculoskeletal issues prior to the second treatment for BRD
Updated treated for repro previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for reproductive issues prior to the second treatment for BRD
Updated treated for other previously	Binary variable (1 = yes, 0 = no) for whether or not an animal was treated for any other issue prior to the second treatment for BRD
ADG at BRD2	Calculated average daily gain at event date ((event weight – arrival weight)/event DOF)
Drug class	The antimicrobial class which was used as first treatment for BRD (beta lactam, fluoroquinolone, macrolide, tetracycline, phenicol)
Drug name	The specific antimicrobial used at first treatment for BRD (ceftiofur, danofloxacin, enrofloxacin, florfenicol, gamithromycin,

		oxytetracycline, tildipirosin, tilmicosin, tulathromycin)
	Days between treatments	The number of days between first and second treatment for BRD (second event date – first event date)
Feedyard level	Updated number of treatments on the same day at feedyard	Count of additional treatments at the yard on the same event date
	Updated number of treatments last week at feedyard	Cumulative count of all treatments at the yard in the last 7 days prior to the event date
	Updated moving averages of treatments	Moving averages of the number of treatments at the feedyard in 3-, 5-, 7-, 14-, 21-, and 28-day increments prior to day of second treatment for BRD
	Updated divergence from moving average	The difference in the event date's number of treatments from each of the increments of moving averages of treatments at the feedyard

3605 BRD2 indicates the second treatment for BRD.

3606 **Appendix Table B.3. Summary of the features selected for inclusion in the BRD1 model**
3607 **with associated chi square correlation coefficients.**

Feature	Chi Square Correlation Coefficient
Event DOF	4573.702274
Event Weight	1588.435581
Temp	1265.809628
Lot Arrival Weight	1032.181785
Death Count	909.6541482
Railer Count	906.9989237
Difference from Yard Pull Moving 7 Day Average	616.5512241
Difference from Yard Pull Moving 5 Day Average	587.9573396
Difference from Yard Pull Moving 3 Day Average	530.6411536

Difference from Yard Pull Moving 21 Day Average	488.7084571
Difference from Yard Pull Moving 14 Day Average	464.809555
Pulls Last Week	423.6559944
Difference from Yard Pull Moving 28 Day Average	336.2323994
Lot Head In	280.5218564
Origin Longitude	250.4637661
ADG at Pull	171.0379668
Origin Latitude	167.5820331
Pulls Today	166.8064602
Sex	159.4191228
Arrival Month	123.291462
Lot Max Month	123.2709112
Pull Month	121.3196543
Distance	109.5110013
Arrival Quarter	95.74068774
Yard Pulls Moving 7 Day Average	83.40157531
Yard Pulls Moving 5 Day Average	83.22193657
Yard Pulls Moving 3 Day Average	75.91059225
Pull Quarter	75.70398169
Yard Pulls Moving 28 Day Average	74.67447512
Yard Pulls Today	74.0621834

3608

3609 **Appendix Table B.4. Summary of the features selected for inclusion in the BRD2 model**
3610 **with associated chi square correlation coefficients.**

Feature	Chi Square Correlation Coefficient
Lot Arrival Weight	5883.277922
Second Pull Event DOF	834.7695641
Drug Name	747.6033154
Second Pull Event Weight	727.9967012
Event Weight	714.3840295
Event DOF	688.6798818

Drug Class	674.6072373
ADG at Second Pull	313.7334169
ADG at Pull	289.1462645
Days Between Pulls	266.4971166
Second Pull Difference from Yard Pull Moving 28 Day Average	254.5541031
Second Pull Difference from Yard Pull Moving 21 Day Average	244.2552955
Second Pull Difference from Yard Pull Moving 14 Day Average	242.3834577
Second Pull Month	193.6363286
Difference from Yard Pull Moving 21 Day Average	184.2238347
Lot Head In	175.7987889
Pulls Today	171.891481
Pulls Last Week	166.7031833
Second Pull Yard Pull Moving 7 Day Average	163.2569589
Second Pull Difference from Yard Pull Moving 7 Day Average	162.219898
Difference from Yard Pull Moving 28 Day Average	160.2515969
Second Pull Yard Pull Moving 3 Day Average	158.4757184
Second Pull Yard Pulls Today	158.0993427
Second Pull Yard Pull Moving 5 Day Average	156.3171592
Second Pull Yard Pulls Last Week	152.0904538
Second Pull Yard Pull Moving 28 Day Average	151.8417672
Second Pull Yard Pull Moving 14 Day Average	148.9196045
Second Pull Yard Pull Moving 21 Day Average	145.2769934
Difference from Yard Pull Moving 14 Day Average	144.0239092
Death Count	143.0168193
Second Pull Pulls Today	142.6409059
Second Pull Difference from Yard Pull Moving 5 Day Average	139.6821739
Origin Longitude	134.7853158
Difference from Yard Pull Moving 7 Day Average	133.9455601

Pull Month	133.4447939
Yard Pulls Moving 21 Day Average	132.9032049
Second Pull Difference from Yard Pull Moving 3 Day Average	130.3977869
Yard Pulls Moving 14 Days Average	129.7740312
Yard Pulls Last Week	129.325457
Railer Count	127.9973476
Second Pull Railer Count	127.7113414
Second Pull Pulls Last Week	126.4887276
Sex	126.4271189
Arrival Month	125.7135927
Temp	125.4594175
Yard Pull Moving 28 Day Average	125.0878284
Lot Max Month	122.5586843
Yard Pulls Today	122.4779628
Yard Pulls Moving 3 Day Average	121.4890673
Yard Pulls Moving 7 Day Average	121.1558651

3611

3612

Appendix C - Supplement to Chapter 5

3613 **Appendix Table C.1. Table describing the ranking of each variable for each predictive**
 3614 **algorithm within each dataset**

Dataset	Model	Feature Importance Ranking
Baseline	BDT ¹	Purchase Month Purchase Year Average Arrival Weight Sex Head Received
	LR ²	Purchase Month Purchase Year Average Arrival Weight Sex Head Received
	NN ³	Purchase Month Purchase Year Average Arrival Weight Sex Head Received
	RF ⁴	Purchase Month Purchase Year Average Arrival Weight Sex Head Received
Baseline Subset	BDT	Purchase Month Purchase Year Average Arrival Weight Sex Head Received
	LR	Purchase Month Average Arrival Weight Purchase Year Sex Head Received
	NN	Purchase Month

		Purchase Year Average Arrival Weight Head Received Sex
	RF	Purchase Month Purchase Year Average Arrival Weight Head Received Sex
Baseline + Econ	BDT	Purchase Month Average Arrival Weight Average Corn Futures Feeder Futures Difference from Average Brent Oil Futures Sex Head Received Average Brent Oil Futures Average Feeder Futures Difference from Average Fat Futures Fat Futures Purchase Year Average Interest Rate Difference from Average Feeder Futures Difference from Average Corn Futures Average Fat Futures Interest Rate Difference from Average Soybean Futures Difference from Average Interest Rate Average Soybean Futures Average Soybean Meal Futures Brent Oil Futures Corn Futures Soybean Meal Futures Difference from Average Soybean Meal Futures Soybean Futures
	LR	Feeder Futures Purchase Year

		Fat Futures Difference from Average Feeder Futures Average Interest Rate Difference from Average Fat Futures Average Soybean Meal Futures Average Feeder Futures Average Corn Futures Soybean Meal Futures Average Arrival Weight Average Soybean Futures Soybean Futures Average Fat Futures Difference from Average Soybean Futures Difference from Average Corn Futures Average Brent Oil Futures Difference from Average Brent Oil Futures Interest Rate Purchase Month Corn Futures Brent Oil Futures Sex Head Received Difference from Average Soybean Meal Futures Difference from Average Interest Rate
	NN	Average Feeder Futures Feeder Futures Fat Futures Purchase Month Average Arrival Weight Difference from Average Feeder Futures Average Interest Rate Average Corn Futures Difference from Average Fat Futures Average Fat Futures Average Soybean Futures Interest Rate Difference Average Brent Oil Futures Sex

		Brent Oil Futures Average Soybean Meal Futures Difference from Average Soybean Meal Futures Average Brent Oil Futures Soybean Meal Futures Purchase Year Soybean Futures Corn Futures Difference from Average Corn Futures Difference from Average Soybean Futures Head Received Difference from Average Interest Rate
	RF	Average Arrival Weight Average Corn Futures Sex Purchase Month Head Received Difference from Average Soybean Meal Futures Average Fat Futures Corn Futures Difference from Average Brent Oil Futures Average Soybean Futures Fat Futures Interest Rate Average Brent Oil Futures Soybean Meal Futures Average Feeder Futures Average Soybean Meal Futures Difference from Average Fat Futures Difference from Average Feeder Futures Difference from Average Interest Rate Soybean Futures Difference from Average Soybean Futures Brent Oil Futures Average Interest Rate Difference from Average Corn Futures Purchase Year Feeder Futures

Baseline + Origin	BDT	Purchase Month Purchase Year Average Arrival Weight Latitude of Origin Sex City of Origin Longitude of Origin Head Received State of Origin
	LR	City of Origin Purchase Month State of Origin Purchase Year Sex Head Received Average Arrival Weight Latitude of Origin Longitude of Origin
	NN	Purchase Month Purchase Year City of Origin Sex Head Received Longitude of Origin Latitude of Origin Average Arrival Weight State of Origin
	RF	Purchase Month Purchase Year Average Arrival Weight State of Origin Longitude of Origin City of Origin Head Received Latitude of Origin Sex

Baseline + Origin + Econ	BDT	Purchase Month Difference from Average Brent Oil Futures Feeder Futures Average Arrival Weight Difference from Average Corn Futures Sex Average Fat Futures Corn Futures Fat Futures Soybean Meal Futures Difference from Average Fat Futures City of Origin Average Feeder Futures Difference from Average Soybean Meal Futures Difference from Average Interest Rate Head Received Purchase Year Longitude of Origin Average Soybean Meal Futures Brent Oil Futures Average Interest Rate Latitude of Origin Average Brent Oil Futures Average Corn Futures Soybean Futures Difference from Average Feeder Futures Difference from Average Soybean Futures State of Origin Average Soybean Futures Interest Rate
	LR	Feeder Futures Fat Futures Average Feeder Futures Difference from Average Feeder Futures Difference from Average Fat Futures City of Origin Purchase Year State of Origin

		Difference from Average Soybean Futures Soybean Futures Average Soybean Futures Average Interest Rate Difference from Average Interest Rate Difference from Average Soybean Meal Futures Sex Average Arrival Weight Head Received Average Corn Futures Longitude of Origin Corn Futures Difference from Average Corn Futures Purchase Month Average Brent Oil Futures Average Soybean Meal Futures Soybean Meal Futures Latitude of Origin Average Fat Futures Brent Oil Futures Interest Rate Difference from Average Brent Oil Futures
	NN	Feeder Futures Average Feeder Futures Difference from Average Feeder Futures Sex Interest Rate Average Soybean Futures Brent Oil Futures Soybean Futures Difference from Average Interest Rate Purchase Month Difference from Average Soybean Futures Purchase Year Longitude of Origin Average Corn Futures Difference from Average Brent Oil Futures Difference from Average Corn Futures

		Average Interest Rate Head Received Average Soybean Meal Futures Corn Futures City of Origin Latitude of Origin Average Fat Futures Difference from Average Soybean Meal Futures Average Arrival Weight Average Brent Oil Futures Soybean Meal Futures State of Origin Difference from Average Fat Futures Fat Futures
	RF	Average Arrival Weight State of Origin Purchase Month Soybean Futures Head Received Purchase Year Average Brent Oil Futures Average Soybean Meal Futures Brent Oil Futures Feeder Futures Latitude of Origin Average Feeder Futures Average Soybean Futures Interest Rate Difference from Average Corn Futures Average Corn Futures Difference from Average Feeder Futures Sex Longitude of Origin Corn Futures Fat Futures City of Origin Average Fat Futures Average Interest Rate

		Difference from Average Soybean Futures Soybean Meal Futures Difference from Average Brent Oil Futures Difference from Average Fat Futures Difference from Average Interest Rate Difference from Average Soybean Meal Futures
--	--	--

3615 ¹BDT = boosted decision tree

3616 ²LR = logistic regression

3617 ³NN = neural network

3618 ⁴RF = random forest