THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 1 (2024)
https://doi.org/ 10.62019/abbdm.v4i1.108.

ASIAN BULLETIN OF BIG DATA MANAGEMENT
http://abbdm.com/
ISSN (Print): 2959-0795
ISSN (online): 2959-0809

Decoding Bovine Behavior: A Machine Learning Analysis of Disease and Event Detection

Ghalib Nadeem*, Zubair, Shahrukh Ali, Bismah Hasan, Basit Hasan, Sajid Yasin

Chronicle

Article history

Received: February 12, 2024
Received in the revised format: Feb 24, 2024
Accepted: Feb 27, 2024
Available online: Feb 29, 2024

Ghalib Nadeem* is currently affiliated with the department of Electronic Engineering, Iqra University, Karachi, Pakistan.
Email: ghalibnadeem@iqra.edu.pk

Zubair is currently affiliated with School of Control and Computer Engineering, NCEPU, Beijing, China.
Email:120214200040@ncepu.edu.cn

Shahrukh Ali is currently affiliated with the department of Information Security, NEDUET, Karachi, Pakistan.
Email: sarangshaikh55@gmail.com

Bismah Hasan and Sajid Yasin is currently affiliated with the department of Computer Science, Iqra University, Karachi, Pakistan.
Email: bismah@iqra.edu.pk
Email: sajidyasin@iqra.edu.pk

Basit Hasan is currently affiliated with the department of Software Engineering, SMIU, Karachi, Pakistan.
Email: basithq@smiu.edu.pk

Corresponding Author*

Abstract

The dairy industry globally serves millions, creates employment opportunities, and contributes to GDP and livelihoods, supporting numerous farmers and milk processors, and contributing to the GDP of many countries. Even with the high number of cows, a variety of illnesses such lameness, mastitis, metritis, foot and mouth disorders have a significant impact on the milk output. This demands the growth of smart methods to assist dairy producers in keeping an eye on the health, nutrition, and general well-being of their cows. To identify noteworthy bovine occurrences and illnesses, this research uses Machine Learning (ML) methods such as Extreme Gradient Boost (XGB), Naive Bayes (NB), and Perceptron. In order to uncover complex relationships and unearth undiscovered information about the characteristics and occurrences of bovine's disease like Acidosis, Calving, Estrus, Lameness, and Mastitis. We closely scrutinize four distinct datasets. Metrics like as Area under the Curve (AUC), F1 score, accuracy, precision, recall, and Receiver Operating Characteristic (ROC) curve are used to evaluate performance. When it comes to identifying estrus events, XGB has the best detection accuracy score 92.59%, whereas XGB can detect a variety of events and disease with the highest recall 100% and the highest precision 95% and AUC of 0.962 when it comes to identifying calving, mastitis, and lameness.

Keywords: Dairy Industry, Estrus Prediction, Machine Learning, Mastitis.

INTRODUCTION

A major obstacle to the world's food security is the prevalence of illnesses in dairy herds, which affects the dairy industry. Bovine diseases include lameness, mastitis, estrus, acidosis, and complications during calving present serious problems for farmers all over the world. These medical conditions put the general health of the herd in danger in addition to reducing production and raising veterinary expenses. Since dairy products satisfy the nutritional needs of millions of people worldwide, early identification and efficient disease management are critical to the sustainability and resilience of the dairy business. Furthermore, the economic consequences of disease outbreaks in dairy herds can be severe since they can lead to the loss of valuable breeding stock and lower the herd's market value. Due to the large number of animals and subtle symptoms, diseases in bovine cattle might be challenging to diagnose. Conventional techniques take an extensive amount of time, are arbitrary, and are liable to error by human beings. Recent advances in technology like machine learning have made detection of diseases more affordable and accessible than ever. To identify conditions like mastitis (N. Abdul Ghafoor et al., 2021), acidosis (Wagner et al., 2020), lameness (C.Post et al., 2020), calving (C.A. dos Santos et al., 2022), and estrus (Wang J et al., 2020), researchers have created ML-based algorithms. Individual cow health is continuously monitored by these ML-based systems using data from multiple sources, including sensors (Gertz et al., 2020), cameras (Wang et al., 2023), wearable (Alipio et al., 2022), and electronic health records (Mancuso et al., 2023).

In the reproductive cycle of dairy cows, calving is a crucial stage. The health and safety of the cow and the calf depend on the prompt diagnosis of calving. Failure to properly diagnose and manage calving can lead to a multitude of complications like dystocia and postpartum diseases. For the cow, it can result in uterine infections, retained placenta, and even death. Additionally, the calf may suffer from inadequate colostrum intake, which can weaken their immune system and make them more susceptible to diseases. Therefore, it is essential for farmers to closely monitor calving and seek veterinary assistance if any issues arise to ensure the well-being of both the cow and calf. To this aim, researchers have investigated a variety of strategies and tactics which includes animal behavior (Cantor et al., 2022), Acoustic monitoring has been utilized to detect distinct vocal patterns associated with calving (Alexander C. et al., 2020).

Mastitis, or inflammation of the breast gland, is often caused by a bacterial infection. It is a common, prohibitive disease that results in lower-quality and less milk produced by dairy cows. Bacteria like Escherichia coli or Staphylococcus aureus that enter the udder through the teat canal cause the infection. Unsanitary circumstances, contaminated bedding, or poor milking hygiene can all contribute to this. In order to stop the bacteria from spreading and preserve the herd's general well-being, immediate and effective control measures are essential. To tackle this circumstance, several machine learning techniques are used, including milk composition analysis (Ebrahimie E et al., 2021), and somatic cell counts (Bobbo T et al., 2021). The natural period of sexual receptivity known as estrus in female cows is the center of the reproductive cycle.

Accurate estrus identification is essential for optimal reproduction, productive breeding management, and high breeding yields. Reproduction is maximized and breeding success rates are increased when farmers are able to determine the ideal time to breed cows through accurate estrus detection. To achieve this target many researchers utilized multiple techniques such as bovine body temperature (Burnett et al., 2020), and using motion sensors and accelerometers for sensor-based tracking of cow activity pattern (Arcidiacono, C. et al. 2020). Acidosis is a metabolic disorder that dairy cows may experience due to a high-energy diet and an imbalance in the rumen's pH level. When the rumen's pH falls below normal, it can cause acidosis, which upsets the delicate balance of bacteria involved in digesting. Along with other health problems, it results in poor milk output and decreased feed consumption. Acidosis can cause lameness, digestive problems, and in extreme situations, even death if treatment is not received. Several researchers use techniques to early detection which includes observing the feed intake (Jaramillo-López et al., 2017), and rumen pH level (Gündüz, K.A. et al., 2022). Lameness in dairy cattle is a major risk to the animals' health and financial success. It is often caused by trauma to the legs and feet, which limits the cow's range of motion and produces agony. Lameness has two effects: it lowers productivity and increases the risk of different medical conditions. Additionally, as lameness prevents cows from moving to the feeding area, their nutrition may suffer and their milk production may be reduced. In order to accomplish early lameness detection, researchers have investigated a variety of approaches and strategies, utilizing machine learning's potential such as such as accelerometers, pressure mats (Taneja et al., 2020) to get the data to cope it earlier.

Through enhancing dairy productivity, insuring stability in the dairy sector, and enhancing animal welfare, artificial intelligence and machine learning are transforming medicine. Artificial Intelligence (AI) can provide real-time monitoring and automation, give non-invasive examinations, control complex data, and identify cattle diseases early through the use of ML models. Lameness, calving, estrus, mastitis, and acidosis are among the common disorders that dairy cows suffer from. Technology and veterinary medicine have a collaboration that goes beyond health to include sustainable farming methods. Modern AI advancements have paved the way for innovative methods of disease diagnosis, such as general medicine predictive algorithms (Nadeem, G. et al., 2023), and have demonstrated promise for a range of medical applications.

This research aims to develop a ML system for detecting physiological and pathological events in cows, specifically identifying and categorizing five major types of common cow diseases and events. GEA Farm Technology's CowView system in Bönen, Germany, will be used for training the system, which will be organized into four data sets. These sets contain information on 11 different types of events and 386 individual Holstein breed cows. These eleven different event types were categorized into three states: Six of the events included health, two involved reproductions, and three were stress-related events (Lardy, et al., 2022). Binary classification is selected, targeting events or diseases that will either occur or not. Various algorithms, including Extreme Gradient Boost, Naive Bayes, and Perceptron model, is applied to determine the probability of an event or disease occurring. Hyper-parameter tuning is performed to achieve the best performance for each machine learning model (Nadeem, G. et al., 2023). The XGB model performed better than other models in identifying and categorizing bovine illnesses and events on all accuracy measures.

The fields of agricultural informatics and bovine health management both benefit greatly from this study's gains. We provide a novel strategy that leverages machine learning and artificial intelligence to improve overall dairy productivity, reduce financial losses, and improve the accuracy of disease detection. The proposed automated disease detection system has the potential to revolutionize the dairy industry and benefit farmers, veterinarians, and other stakeholders by promoting early intervention, improving animal welfare, and optimizing dairy output. Our system employs cutting-edge algorithms capable of instantaneously analyzing huge amounts of data, thereby enabling the prompt administration of suitable medication and the early detection of diseases. This lessens the chance that infectious diseases may infect the cattle while also conserving valuable resources. Early disease diagnosis also enables caretaker to reduce losses by halting the disease's spread. This paper is organized as follows. After this introduction, we review the related work in the field of bovine welfare event detection systems. Then, we present the methodologies and ML model implementation that we used in this study. Next, we describe the data sets that we collected and processed for the model's training and testing phases. After that, we report the performance analysis and a comparison of AI model evaluations. Finally, we draw the conclusions and discuss the future work.

RELATED WORK

A number of researchers have investigated machine learning methods for the purpose of identifying or forecasting the emergence of different health problems. Nevertheless, just one or three diseases have been the focus of each investigation. Precision livestock farming (PLF) is a technology that uses sensors, data analytic, and machine learning to monitor and manage livestock production more efficiently (A.Sharma eet al., 2021). This approach can improve animal welfare and productivity by monitoring animal behavior, health, and environmental conditions. PLF can optimize feeding strategies and reduce waste by monitoring feed intake and growth rates. However, challenges such as data privacy and implementation costs need to be addressed to fully realize the potential of PLF in livestock production (A.Sharma eet al., 2021).

Researchers (Wang J et al., 2020) have improved dairy cow estrus detection and prediction by applying machine learning methods. They examined several algorithms, such as SVM, ANN, and BPNN, using data from a herd. When it came to estrus prediction and detection, the BPNN algorithm with a 0.5-hour time frame had the best accuracy. This study shows how machine learning algorithms might boost overall production and reproductive efficiency in dairy farming. The study emphasizes how crucial cutting-edge technologies like artificial neural networks (ANN) are for breeding choices, which improves reproductive efficiency (Wang J et al., 2020).

The study (X. Zhou et al., 2022) predicts common digestive diseases in dairy cows, such as metritis, mastitis, and lameness, using machine learning algorithms. XGB, Random Forest, KNN, and Rpart were among the eight algorithms the researchers used to create a predictive model. The Rpart method delivered the best results, with precision, accuracy, and AUC of 92.86%, 81.58%, and 0.908, respectively, reflecting that the models achieved great accuracy. In order to increase cow production and health, the study highlights the value of implementing cutting-edge technology like machine learning algorithms in dairy farming. It also emphasizes the need of early detection and treatment of these illnesses. The study emphasizes how crucial it is to use modern tools to protect dairy cows' health and welfare (X. Zhou et al., 2022). By utilizing various machine learning models to anticipate bovine events and diseases in the specified data-sets (Lardy, et al., 2022), the proposed work appears to be the first of its kind when compared to the aforementioned studies.

METHODOLOGY

In this study, we leverage data-sets that inherently possess target classes, making supervised machine learning the preferred approach for our predictive system, the entire flow of the proposed system is illustrated in Figure 1. The system involves acquiring data from four distinct data-sets, preprocessing it, and allocating them for training and testing.

The data is partitioned into training and testing phases, with a 80% training to 20% testing ratio. The data-sets are then used to train various ML models using Python 3.10.6 and libraries like SciKit-Learn, numpy, pandas, and tensor flow. Post-training, a rigorous validation process evaluates the performance of the models, culminating in a comparative analysis across various accuracy parameters. The ML process is meticulously structured, utilizing cutting-edge tools and methodologies to extract meaningful insights from complex data sets.

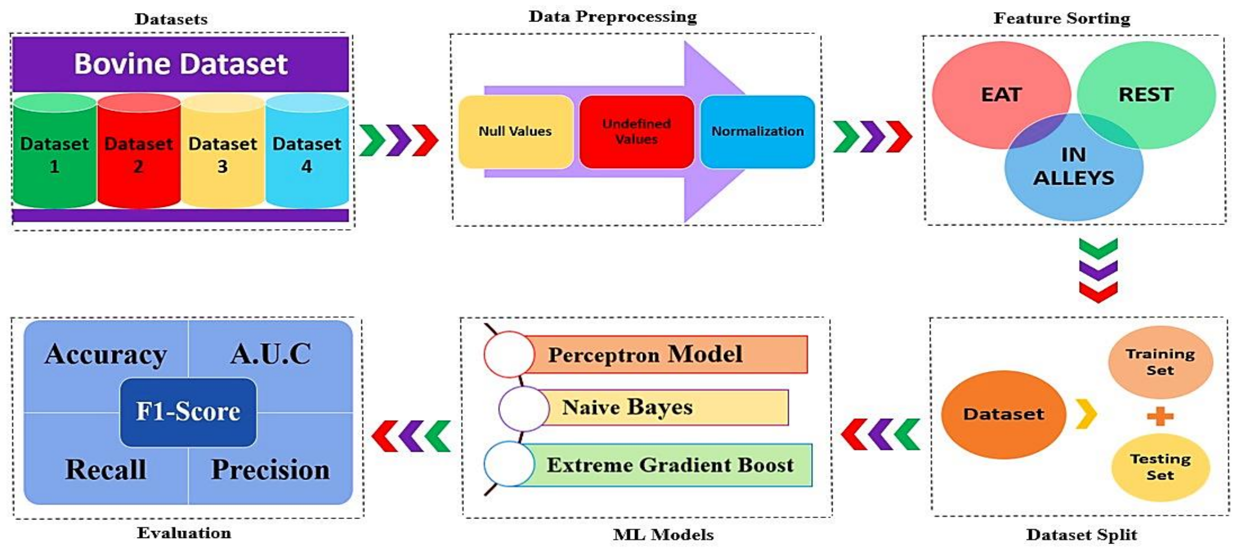

Figure 1.
Block diagram of entire proposed method.

DATA ACQUISITION

The data used in the study was gathered by GEA Farm Technology's CowView system in Bönen, Germany (Lardy, et al., 2022). Data is gathered using a tag attached to each cow's neck collar emits radio waves, which are picked up by antennae throughout the barn and used to triangulate the cow's position. Based on its position, one can deduce the cow's activity. Dataset is composed of four data-sets, the data-sets contains the information of 386 Holstein breed unique cows with a tag id, date, hourly scale aggregation of individual activity, time spent in 'eating', 'resting' and in walking 'in alleys' are in unit seconds for each hour scale and observed the occurrence of 11 types of events for each data-set, illustrated in table 1. Maintaining the Integrity of the Specifications.

Table 1.
Frequency of events occurrence in data-sets

Events	Occurrence			
	Dataset-1	Dataset-2	Dataset-3	Dataset-4
Calving	8	0	0	171
Ruminal Acidosis	0	271	0	0
Accidental Events	0	0	0	15
Mixing	72	0	0	0
Management Changes	0	168	0	2581
LPS injection	27	0	0	0
Estrus	41	7	26	257

Lameness	4	16	0	114
Other Disease	10	8	0	66
Mastitis	9	3	0	32
Other Disturbance	173	671	0	12223

DATA PRE-PROCESSING

One essential function is data pre-processing in determining the final performance of machine learning models (Habib, Beenish et al., 2022). It involves data cleaning as Missing values may seriously affect the accuracy as well as reliability of machine learning models, thus it's critical to manage them properly. It assure the models are trained on accurate and comprehensive data by removing null values and setting undefined data to zero. Similarly, the normalization as the Min-Max scaling approach is a popular technique that converts the data to a range between 0 and 1. It is also referred to as feature scaling or normalization. This approach divides the result by the range of values after deducting the minimum value of each characteristic. Scaling techniques can enhance machine learning algorithms' performance and accuracy by putting all features on a same scale. This prevents certain attributes from prevailing over others because of their wider range of values. This part also involves balancing the data which is the most critical step of data preprocessing because each event or disease in a given data set has an unbalanced number of normal and event occurrence states, as illustrated in table 1, which may lead the model to biased output on suppress class of dataset. Finally, it involves Feature Selection and the most prominent features that are correlated with each individual event and disease are chosen, labelled as EAT, REST, and IN-ALLEYS.

TRAINING AND TESTING SET

The training and testing phases are one of the most influential components of research strategy, lay the groundwork for creating and comprehensively evaluating specialized machine learning models meant for precision-driven cattle disease and event detection. To ensure a solid basis for modelling and evaluation, it's been precisely divided our data set into training and testing sets using an 80:20 ratio allocation. Notably, this method made it possible to integrate four different data sets, all of which added to a comprehensive perspective on real-world situations.

MACHINE LEARNING

Due to ML algorithms' significant complex data processing capabilities, they are frequently utilized in Bovine diagnosis, prediction, and other domains (Liu R et al., 2022). To find the optimum model for event/disease detection based on data, four machine learning models were used. The selection of classifier models is limited to those that have been widely recognized and utilized.

Extreme Gradient Boost

Another ensemble tree model is the Extreme Gradient Boosting XGB technique, which is a large-scale parallel technique that is trained incrementally and continuously adds new decision trees, mathematically it represents the algorithm by the second order Taylor expansion equation as in equation 1 (Morde, Vishal. Et al., 2019), where $y^{(t)}$ represents the forecast for the p-th feature example at the t-th increment. An objective loss function L represents the mathematical equation.

$$\boldsymbol{L(t)} = \sum_{\boldsymbol{p=1}}^{\boldsymbol{n}} \left(\boldsymbol{1}\left(\boldsymbol{y_p}.\boldsymbol{y_p}^{(\boldsymbol{p-1})}\right) + \boldsymbol{f_t}(\boldsymbol{x_p})) + \boldsymbol{\theta}(\boldsymbol{f_t})\right) \quad \textbf{(1)}$$

The SciKit-Learn package utilizing xgboost, the XGB classifier is built with a hyper-parameter of n-threads (number of parallel processes) = 1, objective linear logistic, max-depth = 8, n-estimators is of 450 and eta = 0.3, while the remaining parameters are set to their default values.

Naive Bayes

Using probability to forecast an object's probability in a classification job, a Naive Bayes classifier (NB) is a type of probabilistic machine learning model (Jackins, V. et al., 2021). Equation 2 illustrates the Bayes theorem, establishing the core of the classifier.

$$\boldsymbol{P(y|x)} = \frac{\mathbf{P}(\boldsymbol{x}|\boldsymbol{y})\mathbf{P}(\mathbf{y})}{\mathbf{P}(\boldsymbol{x})} \qquad (\mathbf{2})$$

In this case, variable x indicates attributes and variable y the target class of the event. Where P(y) is the prior probability, P(x|y) is the likelihood probability, P(y|x) is posterior probability and P(x) is the marginal posterior probability. In this case, the method would be as in equation 3.

$$\boldsymbol{P(y|x_1, \dots, x_n)} = \frac{\mathbf{P}(\boldsymbol{x_1}|\boldsymbol{y})\mathbf{P}(\boldsymbol{x_2}|\boldsymbol{y})\mathbf{P}(\boldsymbol{x_3}|\boldsymbol{y})\mathbf{P}(\mathbf{y})}{\mathbf{P}(\boldsymbol{x_1})\mathbf{P}(\boldsymbol{x_2})\mathbf{P}(\boldsymbol{x_3})} \qquad (\mathbf{3})$$

Equation 3 shows how to build the Naive Bayes classifier of nature Gaussian using the SciKit-Learn module GaussianNB, the alpha is set to 1 and the default values apply to all other parameters. The zero probability problem in the NB algorithm is solved by the additive smoothing parameter alpha.

Perceptron

A simple perceptron model by SciKit-Learn package perceptron is used as shown in equation 4, trained using the stochastic gradient descent optimization algorithm and activation function by default is step function with an eta of 0.0001 and initially by default weights are uniform and other remaining parameters are default values.

$$\boldsymbol{Z} = \mathbf{f}(\mathbf{b} + \mathbf{x}.\mathbf{w}); \mathbf{Z} = \mathbf{f}(\mathbf{b} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{x_i}.\boldsymbol{w_i})) \qquad (\mathbf{4})$$

Z is the current node's output, f is the activation function, xi is the input from the node in the preceding layer with weight wi, and b is the bias to adjust the spectrum of the summation value. Figure 2 shows the node model for perceptron model.

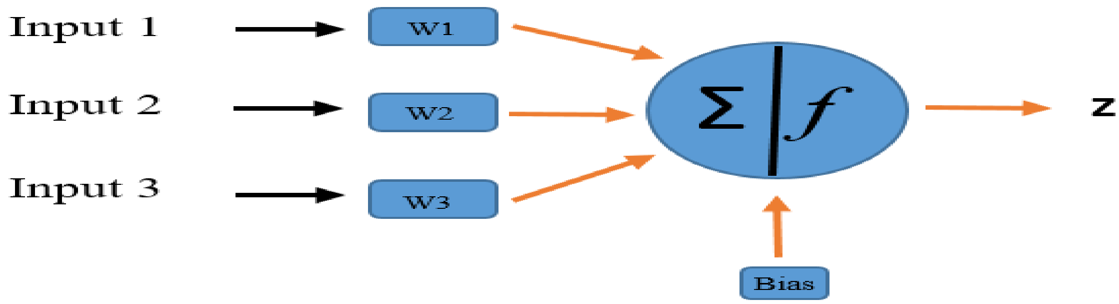

Figure 2.
Single perceptron model spectrum for 3 input attributes with weights and bias.

EVALUATION

How well a machine learning model performs with new and untrained data determines how effective the model is. The optimal configuration method for machine learning models was determined by analyzing the accuracy scores in conjunction with the F1-score, Precession, and Recall values (Habib, Beenish, et al., 2022). The area under the curve with respect to ROC curve are also shown against genuine positive and false negative data to further support the findings.

DATA-SET DESCRIPTION

There are several data sets available for detecting diseases in cows and monitoring their health; the data set utilized in this study was gathered from the Cow-View system (GEA Farm Technology, Bönen, Germany), Four data-sets make up the data. The following is the structure of the data-sets: A total of 386 Holstein breed unique cows having individuals data representing the amount of time spent "eating," "resting," and strolling "in alleys." Lastly, a Boolean is given for each of the 11 different categories of occurrences; a 1 means that the incident of that type was reported for this hour, and a 0 means that it wasn't. The INRAE Herbipôle experimental farm in Marcenat, France is the source of data sets 1 and 2. Data-sets 3 and 4 pertain to European commercial farms and include information on 28, 28, 30, and 300 cows that were observed for six months, two months, forty days, and a year, respectively with the approximately losses of 12.45% with the data for 107665 hours, 0.2% withholding data for 40246 hours, 10.4% for 26224 hours and 16.47% with the data for 2177207 hours (Lardy, et al., 2022). The frequency of events observed in each data-sets is in table 2. The events covered for the training and testing on ML models are: Acidosis, Calving, Estrus, Mastitis, and Lameness, other remaining factors are for the farm's observed during the data gathering.

Table 2.
Unbalanced distribution and frequency of even

Events	Dataset-1			Dataset-2			Dataset-3			Dataset-4		
	FREQ	0s	1s	FREQ	0s	1s	FREQ	0s	1s	FREQ	0s	1s
Estrus	41	106681	984	7	40079	168	26	25601	624	257	2171040	6168
Mastitis	9	1077449	216	3	40175	72	0	0	0	32	2176440	768
Lameness	4	107569	96	16	39863	384	0	0	0	144	2174472	2736
Calving	8	107473	192	0	0	0	0	0	0	171	2173104	4104
Acidosis	0	0	0	271	33743	6504	0	0	0	0	0	0

EXPERIMENTAL RESULTS

Accuracy ratings and other metric variables are used to determine the best setup approach for machine learning models. Equations 5, 6, 7, and 8 present the parameters that are utilized to evaluate our trained models' productivity, which are accuracy, recall, precision, F1 score, and AUC. We utilized ROC in relation to AUC for our ML approaches' performance assessment since it is a more robust performance assessor than a confusion matrix. The AUC, which may yield reliable accuracy results even in the presence of variations in classifier performance, is computed using the false positive and true positive values in the ROC curve. The performance of the classifiers may be evaluated using the mentioned metrics, which are represented in this study as (FN), (FP), (TN), and (TP).

$$\boldsymbol{Accuracy} = \frac{\mathbf{TP + TN}}{\boldsymbol{TP + FP + TN + FN}} \quad (5)$$

$$\boldsymbol{Recall} = \frac{\mathbf{TP}}{\boldsymbol{TP + FN}} \quad (6)$$

$$\boldsymbol{Precision} = \frac{\mathbf{TP}}{\boldsymbol{TP + FP}} \quad (7)$$

$$\boldsymbol{F1 - Score} = \frac{\mathbf{2 * Recall * Precision}}{\boldsymbol{Recall * Precision}} \quad (8)$$

When it comes to the values, TP used for "true positives," the frequency of individuals that accurately projected to be "positive," and TN here represent "true negatives," the frequency of accurately expected to be "negative." False positives (FP) are the frequency of individuals that projected to be positive but were in actual negative, and false negatives (FN) are the number of individuals who were predictably negative but were actually positive (Liu R, et al., 2022), (Abro,et al., 2021) (Abro,et al., 2023). In this study, five different events of Acidosis, Calving, Estrus, Lameness, and Mastitis are detected using three different machine learning models: XGB, Naïve Bayes, and Perceptron. The ROC curve and AUC metric were used to evaluate each model's performance for the individual event. The significance of the ROC curve in evaluating the performance of predictive models for different events. The ROC curves for the three models for the event of Acidosis are shown in figure 3, while the curves for the events of Calving, Estrus, Lameness, and Mastitis are shown in figures 4, 5, 6, and 7.

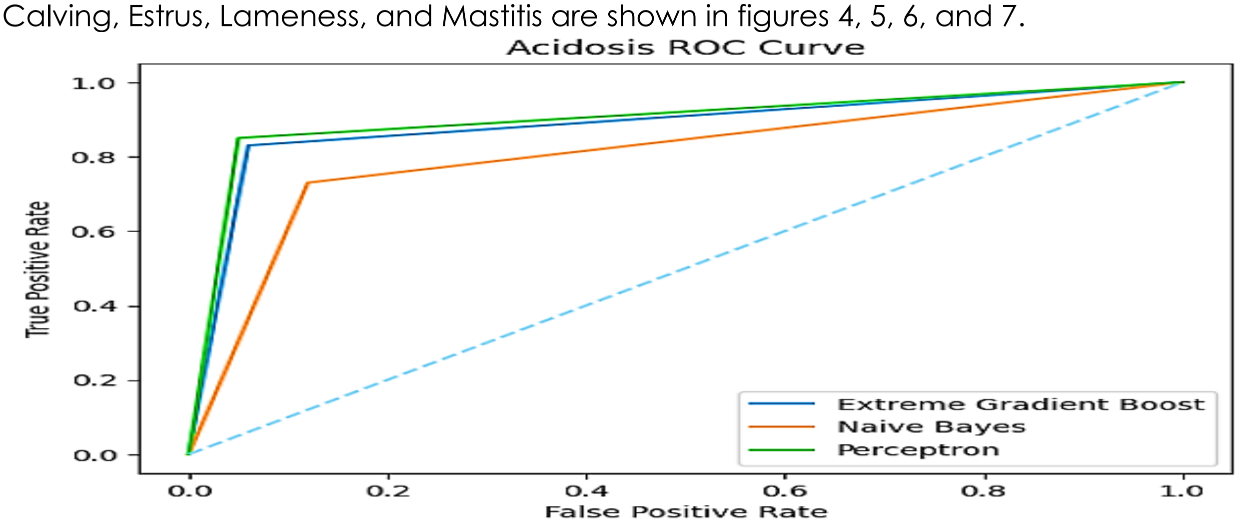

Figure 3.
ROC curve for the event of Acidosis.

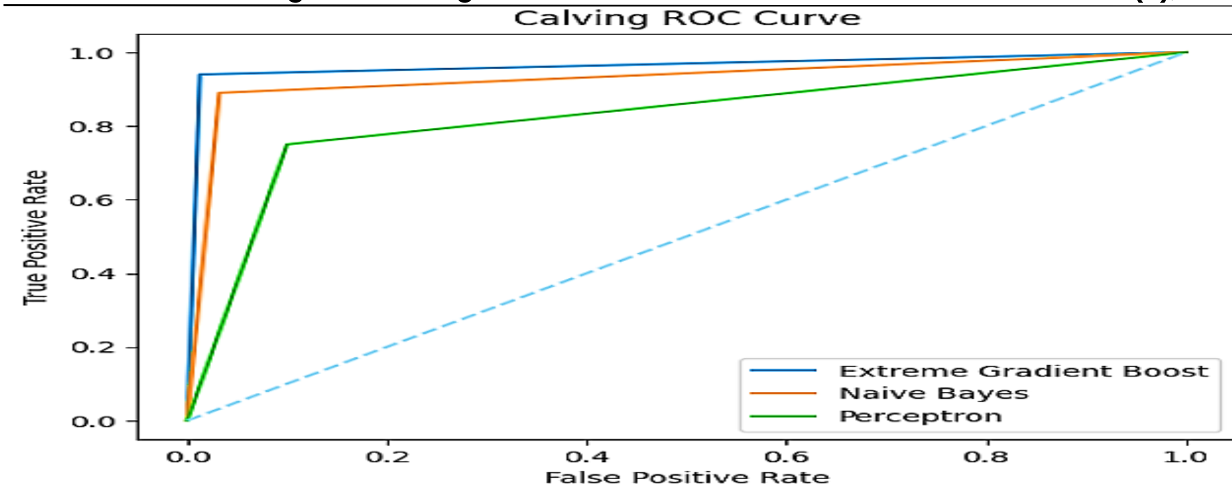

Figure 4.
ROC curve for the event of Calving

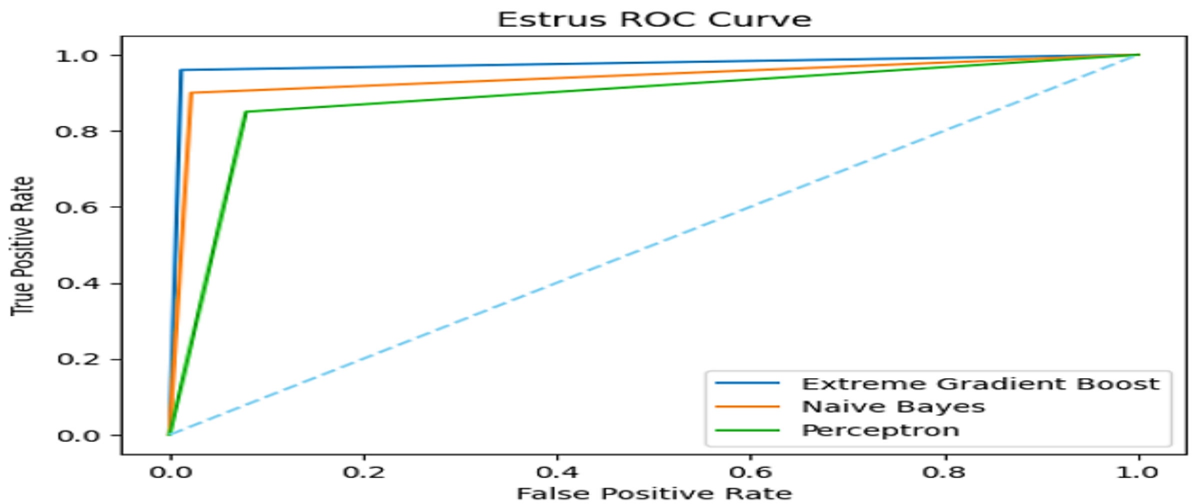

Figure 5.
ROC curve for the event of Estrus

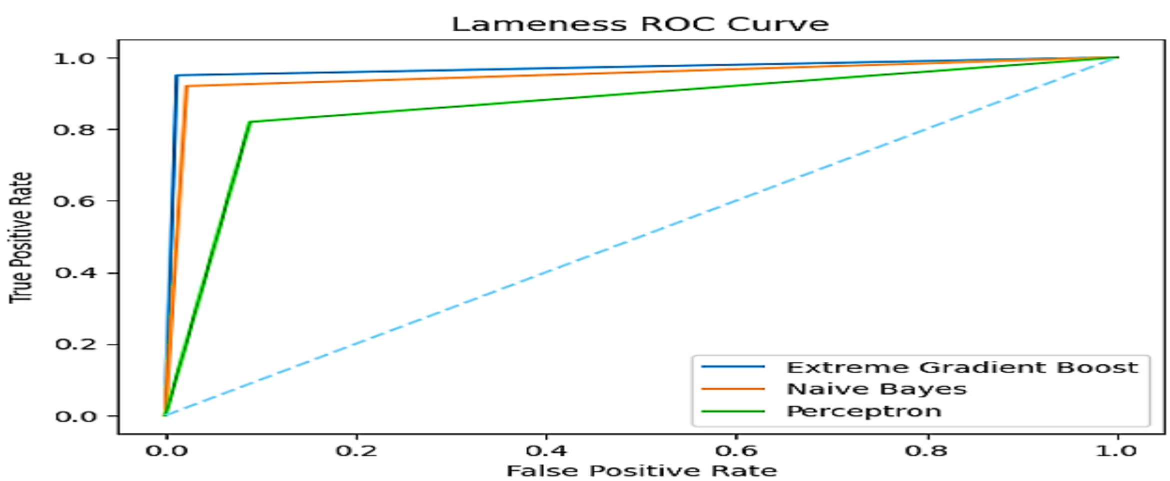

Figure 6.
ROC curve for the event of Lameness

Table 3.
Overall accuracy achievement

Events	Dataset	Accuracy		Precision		Recall		F1 score		AUC	
		Model	Value	Model	Value	Model	Value	Model	Value	Model	Value
Acidosis	2	NB	79.92	NB	84	SPM	83	NB	81	SPM	85
Mastitis	1	SPM	80.50	SPM	85	XGB	89	SPM	83	NB	89
	2	XGB	91.09	NB	92	XGB	100	XGB/NB	90	XGB/NB	96
	4	XGB	87.09	XGB	93	NB/SPM	87	NB	88	XGB	93
Calving	1	NB	82.56	NB	93	XGB/NB	84	NB	85	XGB/NB	92
	4	XGB	90.18	XGB	92	XGB	85	XGB	90	XGB	94
Estrus	1	XGB	92.59	NB	92	XGB	96	XGB	91	XGB	96
	2	NB	87.36	SPM	86	NB	92	XGB/SPM	80	XGB/NB	89
	3	SPM	85.39	NB	90	NB	84	NB	87	NB	94
	4	XGB	83.59	XGB	89	NB	82	XGB/NB	84	XGB	92
Lameness	1	NB	86.72	NB	90	XGB/NB	97	XGB/NB	86	XGB/NB	91
	2	NB	84.50	SPM	79	NB	94	SPM/NB	84	NB	89
	4	XGB	89.91	XGB	95	XGB/NB	88	XGB	89	XGB	96

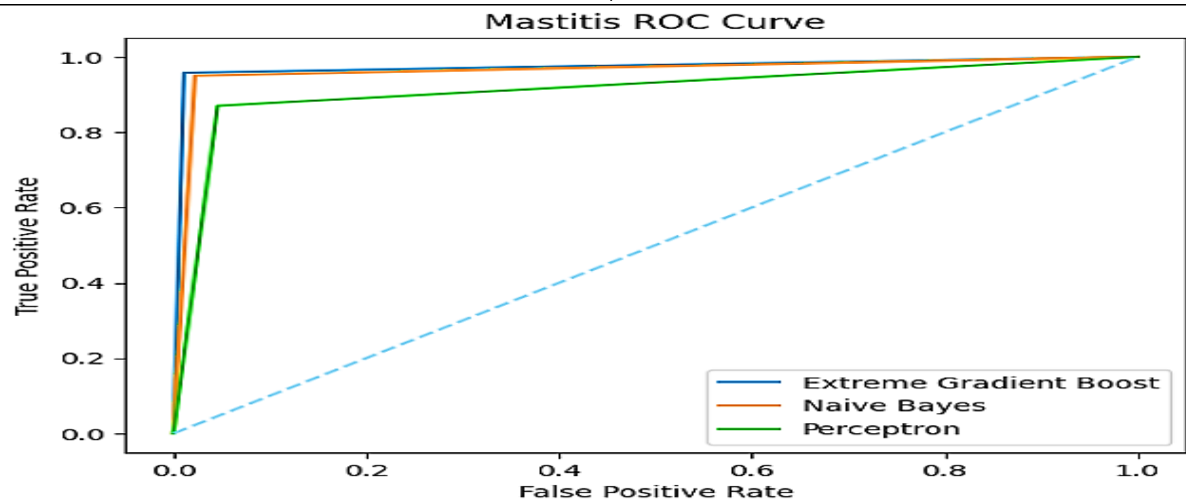

Figure 7.
ROC curve for the event of Mastitis

Model XGB demonstrated the highest performance on event Calving, Estrus, Mastitis and Lameness, with AUC values of 0.92, 0.89, and 0.95, respectively. Model Naïve Bayes performed closer to XGB on similar events, with an AUC value of 0.88. Perceptron model performed the best for the event of Acidosis, with an AUC value of 0.91. This indicates that model XGB was the best at distinguishing between the positive and negative cases in the dataset, while model Perceptron didn't perform that much well. The XGB model's overall performance analysis for various event detection is outlined in figure 8, similarly an overall performance of model which hits the highest metrics are shown in table 3.

The XGB model is outperforming the Naïve Bayes and Perceptron models due to several reasons where one of the most essential is Model Complexity, as the XGB is an ensemble method that builds multiple decision trees and combines them to make a final prediction. This allows it to capture complex patterns in the data, making it more flexible and accurate. Similarly, in features handling XGB does not make the premise that every attribute is independent, in contrast to Naive Bayes.

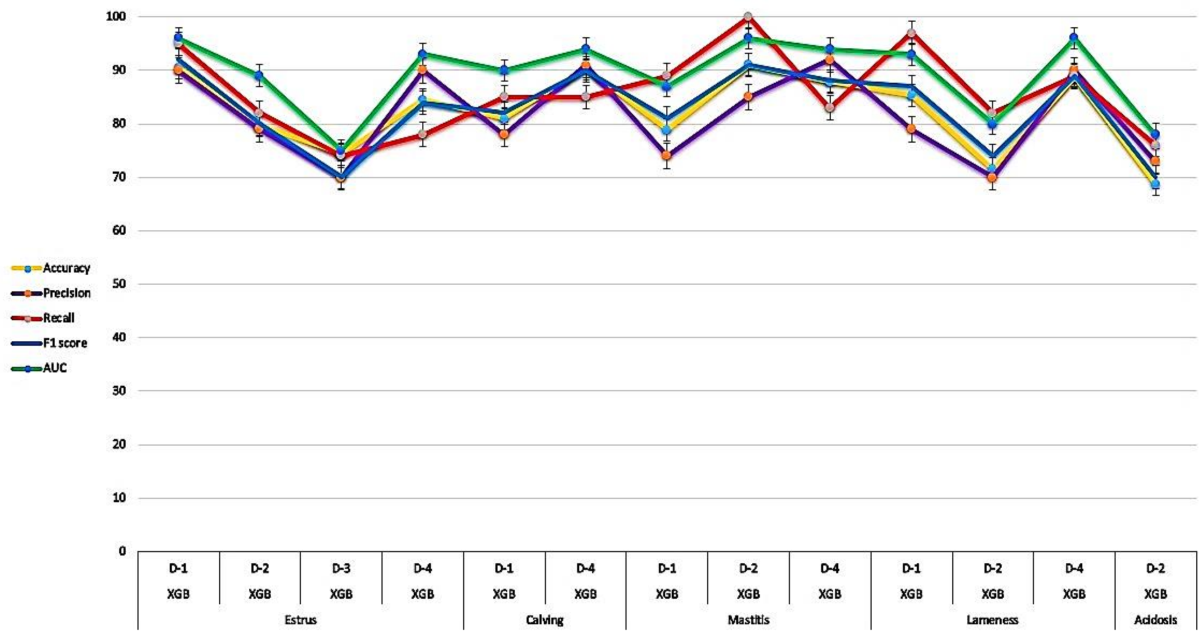

Figure 8.
Comprehensive XGB model's accuracy metrics analysis.

As a result, it can manage feature interactions more effectively. However, as the Perceptron is a linear model, it could not function effectively if the data cannot be separated linearly. Despite being computationally fast, Naive Bayes relies heavily on the independence of features, which may not hold for many datasets from the real world. Therefore, the superior performance of XGB in bovine disease and event detection is the attributed to its ability to model complex relationships, handle feature interactions, prevent overfitting, and deal with missing values effectively. However, it's important to note that the performance of these models can vary depending on the specific characteristics of the dataset and the problem at hand.

CONCLUSION

This research presents a thorough review of machine learning for early bovine disease fand event detection, highlighting their critical role in enhancing dairy production worldwide. The results we have obtained highlight the trans-formative capacity of machine learning-driven insights, which may provide farmers with prompt interventions, increased efficiency, and decreased financial losses within the dairy sector. Three prominent ML models were used: Naive Bayes, Perceptron Model, and Extreme Gradient Boost. Across a variety of data sets, the XGB showed strong performance in event classification, exhibiting high metrics such as F1 score, accuracy, recall, precision, and AUC. Its prognosis for Mastitis was accurate at 91.09%, while its accuracy in detecting Estrus was 92.59%. The model also demonstrated potential in early event detection by showing high accuracy in Calving detection, Lameness detection, and Acidosis prediction. These findings demonstrate how well the XGB detects events in a variety of settings and data sources. This research demonstrates the promise of data-centric innovation, paving the way for a future where technology integrates with agricultural tradition to create a resilient, efficient, and compassionate dairy farming landscape.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the supervisors and their different affiliations.

Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.

Availability of data and material: In the approach, the data sources for the variables are stated.

Authors' contributions: Each author participated equally to the creation of this work.

Conflicts of Interests: The authors declare no conflict of interest.

Consent to Participate: Yes

Consent for publication and Ethical approval: The first two datasets were acquired under the conditions of experimentation. The instructions for utilizing the animals were in line with the EU Directive 2010/63's framework for the protection of animals used in research. The French Ministry of Agriculture has approved the INRAE "Herbipôle" experimental facility (https://doi.org/10.15454/1.5572318050509348E12, UE 1414, Marcenat, France) to conduct studies on live animals (EEA accreditation #C15-114-01). The regional ethics committee reviewed and approved the protocol used in dataset 1 (approval: APAFIS2015043014541577). Such permission was not necessary for the methodology employed in dataset 2. Measurements, feeding, housing, animal handling, and other aspects of the on-farm routine included, thus datasets 3 and 4 were acquired without the need for any further animal interventions.

REFERENCES

A. Sharma, A. Jain, P. Gupta and V. Chowdary, "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," in IEEE Access, vol. 9, pp. 4843-4873, 2021, doi: 10.1109/ACCESS.2020.3048415.

Abro, A. A., Khan, A. A., Talpur, M. S. H., Kayijuka, I., & Yaşar, E. (2021). Machine learning classifiers: a brief primer. University of Sindh Journal of Information and Communication Technology, 5(2), 63-68.

Abro, A. A., Siddique, W. A., Talpur, M. S. H., Jumani, A. K., & Yaşar, E. (2023). A combined approach of base and meta learners for hybrid system. Turkish Journal of Engineering, 7(1), 25-32.

Alexandra C. Green, et al. "Vocal Production in Postpartum Dairy Cows: Temporal Organization and Association with Maternal and Stress Behaviors." Journal of Dairy Science, 31 Oct. 2020, www.sciencedirect.com /science/article/pii/ S0022030220308778.

Alipio, M., \& Villena, M. L. (2022). Intelligent wearable devices and biosensors for monitoring cattle health conditions: A review and classification. Smart Health, 100369.

Arcidiacono, C., et al. "Moving Mean-Based Algorithm for Dairy Cow's Oestrus Detection from Uniaxial-Accelerometer Data Acquired in a Free-Stall Barn." Computers and Electronics in Agriculture, vol. 175, 8 Aug. 2020, p. 105498, https://doi.org/10.1016/j.compag.2020.105498.

Bobbo T, Biffani S, Taccioli C, Penasa M, Cassandro M. Comparison of machine learning methods to predict udder health status based on somatic cell counts in dairy cows. Sci Rep. 2021 Jul 1;11(1):13642. doi: 10.1038/s41598-021-93056-4. PMID: 34211046; PMCID: PMC8249463.

Burnett, Tracy A., et al. "Rumen-Reticular Temperature during Estrus and Ovulation Using Automated Activity Monitors in Dairy Cows." Frontiers, Frontiers, 23 Nov. 2020, https://www.frontiersin.org/articles/10.3389/fvets.2020.597512/full.

C. A. dos Santos, N. M. D. Landim, H. X. de Araújo, and T. do P. Paim, “Automated Systems for Estrous and Calving Detection in Dairy Cattle,” AgriEngineering, vol. 4, no. 2, pp. 475–482, May 2022, doi: 10.3390/agriengineering4020031.

C. Post, C. Rietz, W. Büscher, and U. Müller, “Using Sensor Data to Detect Lameness and Mastitis Treatment Events in Dairy Cows: A Comparison of Classification Models,” Sensors, vol. 20, no. 14, p. 3863, Jul. 2020, doi: 10.3390/s20143863.

Cantor, Melissa C., et al. “Using Machine Learning and Behavioral Patterns Observed by Automated Feeders and Accelerometers for the Early Indication of Clinical Bovine Respiratory Disease Status in Preweaned Dairy Calves.” Frontiers, Frontiers, 30 May 2022, https://www.frontiersin.org/articles/10.3389/fanim.2022.852359/full.

Ebrahimie E, Mohammadi-Dehcheshmeh M, Laven R, Petrovski KR. Rule Discovery in Milk Content towards Mastitis Diagnosis: Dealing with Farm Heterogeneity over Multiple Years through Classification Based on Associations. Animals (Basel). 2021 Jun 1;11(6):1638. doi: 10.3390/ani11061638. PMID: 34205858; PMCID: PMC8227403.

Gertz, M., Große-Butenuth, K., Junge, W., Maassen-Francke, B., Renner, C., Sparenberg, H., \& Krieter, J. (2020). Using the XGBoost algorithm to classify neck and leg activity sensor data using on-farm health recordings for locomotor-associated diseases. Computers and electronics in agriculture, 173, 105404.

Gündüz, K.A., and F. Başçiftçi. “IOT-Based Ph Monitoring for Detection of Rumen Acidosis.” Arquivo Brasileiro De Medicina Veterinária e Zootecnia, Universidade Federal De Minas Gerais, Escola De Veterinária, 10 June 2022, https://www.scielo.br/j/abmvz/a/Tynk37MCT6mwVBGMtttRkGt/.

Habib, Beenish, and Farida Khursheed. "Performance evaluation of machine learning models for distributed denial of service attack detection using improved feature selection and hyper-parameter optimization techniques." Concurrency and Computation: Practiceand Experience 34, no. 26 (2022): e7299.

Jackins, V., Vimal, S., Kaliappan, M. et al. AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. J Supercomput 77, 5198–5219 (2021). https://doi.org/10.1007/s11227-020-03481-x

Jaramillo-López, Esaúl, Mateo F. Itza-Ortiz, Gwendolyne Peraza-Mercado, and José M. Carrera-Chávez. "Ruminal acidosis: strategies for its control." Austral journal of veterinary sciences 49, no. 3 (2017): 139-148.

Lardy, Romain, M-M. Mialon, N. Wagner, Y. Gaudron, Bruno Meunier, K. Helle Sloth, Dorothee Ledoux et al. "Understanding anomalies in animal behaviour: data on cow activity in relation to health and welfare." Animal-Open Space 1, no. 1 (2022): 100004.

Liu R, Xu Z, Teng J, Pan X, Lin Q, Cai X, Diao S, Feng X, Yuan X, Li J, Zhang Z. Evaluation of six machine learning classification algorithms in pig breed identification using SNPs array data. Anim Genet. 2023 Apr;54(2):113-122. doi: 10.1111/age.13279. Epub 2022 Dec 2. PMID: 36461674.

Mancuso, D., Castagnolo, G., \& Porto, S. M. (2023). Cow Behavioural Activities in Extensive Farms: Challenges of Adopting Automatic Monitoring Systems. Sensors, 23(8), 3828.

Morde, Vishal. “XGBoost Algorithm: Long May She Reign!” Medium, Towards Data Science, 8 Apr. 2019, https://towardsdatascience.com/https-medium-com-vishalmordexgboost-algorithm-long-she-may-rein-edd9f99be63d.

N. Abdul Ghafoor and B. Sitkowska, “MasPA: A Machine Learning Application to Predict Risk of Mastitis in Cattle from AMS Sensor Data,” AgriEngineering, vol. 3, no. 3, pp. 575–584, Aug. 2021, doi: 10.3390/agriengineering3030037.

Nadeem, G., Rehman, Y., Khaliq, A., Khalid, H., \& Anis, M. I. (2023, March). Artificial Intelligence based prediction system for General Medicine. In 2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-6). IEEE.

Taneja, Mohit, John Byabazaire, Nikita Jalodia, Alan Davy, Cristian Olariu, and Paul Malone. "Machine learning based fog computing assisted data-driven approach for early lameness detection in dairy cattle." Computers and Electronics in Agriculture 171 (2020): 105286.

Wagner, Nicolas, Violaine Antoine, Marie-Madeleine Mialon, Romain Lardy, Mathieu Silberberg, Jonas Koko, and Isabelle Veissier. "Machine learning to detect behavioural anomalies in dairy cows under subacute ruminal acidosis." Computers and electronics in agriculture 170 (2020): 105233.

Wang J, Bell M, Liu X, Liu G. Machine-Learning Techniques Can Enhance Dairy Cow Estrus Detection Using Location and Acceleration Data. Animals (Basel). 2020 Jul 8;10(7):1160. doi:10.3390/ani10071160. PMID: 32650526; PMCID: PMC7401617.

Wang, Y., Li, Q., Chu, M., Kang, X., \& Liu, G. (2023). Application of infrared thermography and machine learning techniques in cattle health assessments: A review. Biosystems Engineering, 230, 361-387.

X. Zhou et al., “The Early Prediction of Common Disorders in Dairy Cows Monitored by Automatic Systems with Machine Learning Algorithms,” Animals, vol. 12, no. 10, p. 1251, May 2022, doi: 10.3390/ani12101251.