Harnessing uncertainty: a deep mechanistic approach for cautious diagnostic and forecast of Bovine Respiratory Disease

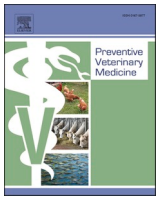
Theophile Ghislain Loic Eyango Tabi, Maud Rouault, Victoria Potdevin, Xavier L'hostis, Sebastien Assie, Sébastien Picault, Nicolas Parisey

Preventive Veterinary Medicine 233 (2024) 106354

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



Harnessing uncertainty: A deep mechanistic approach for cautious diagnostic and forecast of Bovine Respiratory Disease

Théophile Ghislain Loïc Eyango Tabi [a,b,c,*], Maud Rouault [a], Victoria Potdevin [b], Xavier L'hostis [b], Sébastien Assié [a], Sébastien Picault [a,1], Nicolas Parisey [c,1]

[a] Oniris, INRAE, BIOEPAR, 44300, Nantes, France
[b] ADVENTIEL, 7 boulevard nominoë, 35740, Pace, France
[c] INRAE, IGEPP, La Motte au Vicomte, 35640, Le Rheu, France

ARTICLE INFO

Keywords:
Bayesian deep learning
Stochastic mathematical epidemiological model
Uncertainty-aware parameter inference
Bovine Respiratory Disease

ABSTRACT

Bovine Respiratory Disease (BRD) is a prevalent infectious disease of respiratory tract in cattle, presenting challenges in accurate diagnosis and forecasting due to the complex interactions of multiple risk factors. Common methods, including mathematical epidemiological models and data-driven approaches such as machine learning models, face limitations such as difficult parameter estimation or the need for data across all potential outcomes, which is challenging given the scarcity and noise in observing BRD processes. In response to these challenges, we introduce a novel approach known as the Bayesian Deep Mechanistic method. This method couples a data-driven model with a mathematical epidemiological model while accounting for uncertainties within the processes. By utilising 265 lung ultrasound videos as sensor data from 163 animals across 9 farms in France, we trained a Bayesian deep learning model to predict the infection status (infected or non-infected) of an entire batch of 12 animals, also providing associated confidence levels. These predictions, coupled with their confidence levels, were used to filter out highly uncertain diagnoses and diffuse uncertainties into the parameterisation of a mathematical epidemiological model to forecast the progression of infected animals. Our findings highlight that considering the confidence levels (or uncertainties) of predictions enhances both the diagnosis and forecasting of BRD. Uncertainty-aware diagnosis reduced errors to 32 %, outperforming traditional automatic diagnosis. Forecast relying on veterinarian diagnoses, considered the most confident, had a 23 % error, whilst forecast taking into account diagnosis uncertainties had a close 27.2 % error. Building upon uncertainty-awareness, our future research could explore integrating multiple types of sensor data, such as video surveillance, audio recordings, and environmental parameters, to provide a comprehensive evaluation of animal health, employing multi-modal methods for processing this diverse data.

1. Introduction

Bovine Respiratory Disease (BRD) is a prevalent, multi-factorial affliction impacting cattle worldwide, involving infections in both the upper and lower respiratory tracts (infectious bronchopneumonia), influenced by various causal agents (Smith et al., 2020). Cattle across all ages and production types, including dairy, beef, and veal, are susceptible. Although research has primarily focused on beef or feedlot calves (Edwards, 2010; Woolums et al., 2018), as well as dairy replacement heifers or veal calves, due to BRD's significance in these sectors (Dubrovsky et al., 2019; Pardon et al., 2013), it remains a critical concern in all cattle populations. Untreated cases potentially leading to rapid performance decline and fatal pneumonia (Delabouglise et al., 2017; Engler et al., 2014). These statistics contribute significantly to economic losses, accounting for up to 20 % of farmers' annual incomes (Bareille et al., 2009). Diagnosing BRD is a complex and ambiguous process due to its potential interacting causes, including various pathogens such as bacteria and viruses, as well as non-infectious factors like the animal's genetics (breed, immunity, etc.) and environmental influences such as farming management practices (weaning, living conditions, treatments, etc.) (Gaudino et al., 2022; Murray et al., 2017). Antibiotics represent the primary treatment for BRD in the feeding

* Corresponding author at: Oniris, INRAE, BIOEPAR, 44300, Nantes, France.
E-mail address: loic.eyango@adventiel.fr (T.G.L. Eyango Tabi).
[1] Equal contributions

https://doi.org/10.1016/j.prevetmed.2024.106354
Received 12 March 2024; Received in revised form 30 August 2024; Accepted 30 September 2024
Available online 9 October 2024

period (Brault et al., 2019; Nickell and White, 2010), but their overuse raises concerns about the emergence of antibiotic resistance. Therefore, there is an urgent need to implement effective measures for studying the disease, controlling its spread, managing antibiotic usage and associated costs for farmers.

Various modelling approaches, encompassing data-driven methods such as machine learning algorithms, have been explored for diagnosing BRD. These algorithms, recognised for their effectiveness in applications like image recognition and natural language processing, have found multiple applications in veterinary research, predicting BRD (Cantor et al., 2022; Ramezani Gardaloud et al., 2022; Timsit et al., 2011). These methods extract BRD-related patterns by employing various sensors (ear-tag, intraruminal bolus, ultrasound scanner, etc) to record cattle behaviour, appearance and clinical signs. To minimise prediction errors, machine learning models adjust their internal parameters according to the given data. The performance and adaptability of these models heavily thus rely on the richness and diversity of collected data. However, the absence of a gold standard to diagnose BRD based on clinical signs (Timsit et al., 2016) makes early detection through sensor data alone susceptible to false alerts. Additionally, young beef cattle instinctively tend to conceal their clinical signs (Griffin, 2010), a behaviour that evolved as a survival tactics, further increasing the rarity of meaningful observations.

Unlike data-driven methods, knowledge-driven approaches, such as mechanistic models, are mathematical models designed to describe and represent the intricate relationships within a system. These models are constructed based on a set of rules derived from theoretical insights and empirical observations. In epidemiology, they find common application in predicting the dynamic behaviour of complex diseases like Covid (Plank et al., 2022), African swine fever (Muñoz et al., 2022), avian influenza (Lambert et al., 2023), and BRD (Picault et al., 2022). The primary strength of mechanistic models, in contrast to machine learning models, lies in their ability to simulate diverse scenarios without requiring extensive data. By simulating various disease management scenarios, such as surveillance, prevention, and intervention strategies, we can compare their effectiveness and offer evidence-based recommendations, guiding decision-makers to implement optimal control measures. Mechanistic models have demonstrated their efficiency for studying and controlling the spread of infectious diseases (Ezanno et al., 2020). However, it is crucial to note that their calibration presents a substantial challenge. It is a tedious work to re-calibrate these models to specific real-world outcomes.

Our research aimed to devise an approach that adeptly addresses the constraints of both data-driven and knowledge-driven methodologies in epidemiology. We sought to develop a method capable of extracting insights from limited data sources while conserving its reliability in the face of noisy observations. More specifically, we crafted multiple scenarios to compare various methods of automatically extracting insights from sensor data using a data-driven model, and subsequently repurposing these insights to forecast the progression of BRD through a mathematical epidemiological model.

2. Materials and methods

The methodology of this work is structured as follows (Fig. 1): In the Section 2.1, we provide a detailed overview of the data acquisition process, encompassing pulmonary ultrasound videos employed as sensor data with the clinical and biological examination, established by veterinarians, serving as the ground truth to distinguish infected animals expressing clinical signs. In Section 2.2, a baseline scenario was devised, where veterinarian expertise was used to reliably diagnose the total number of infected animals at several observation dates over a 30-day period. Crucial epidemiological parameters of a BRD mechanistic model were then inferred enabling the forecast of the optimal path of the evolution of the total number of infected animals. In subsequent sections, 2.3, 2.4, 2.5), we devise multiple methods to automatically estimate the total number of infected animals, to improve the diagnostics and forecast by incorporating uncertainty-based approaches to filter out noisy observations or propagate the uncertainty of the diagnostics to the forecasting. To conclude, the diagnostic and forecast performance of each method is assessed in comparison with the baseline scenario 2.2.

In our study, we use the term “diagnosis” to refer to predictions made instantaneously based on current data. In contrast, we use ’forecast’ to describe the prediction of future disease progression, utilizing current diagnostic data to project outcomes over time.

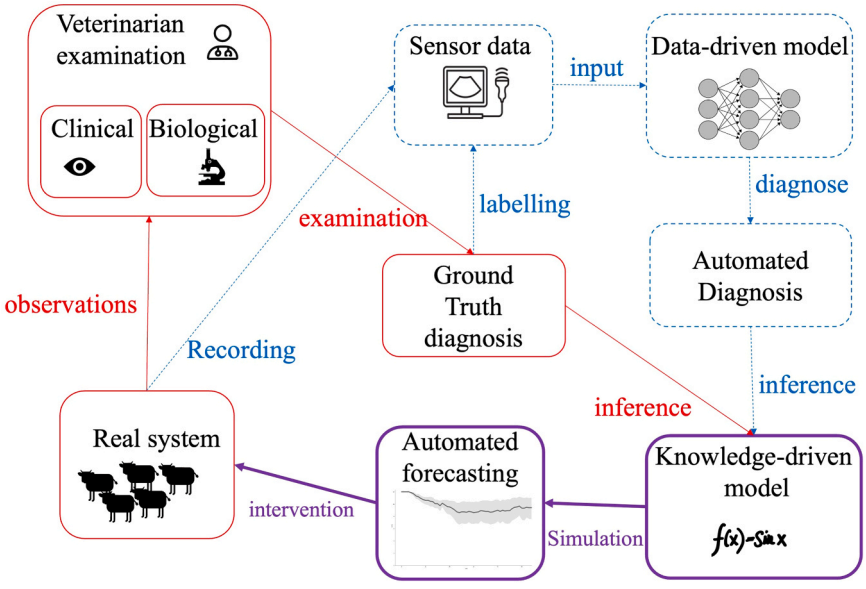


Fig. 1. Workflow of Bayesian deep mechanistic model. A real system is a farm containing animals in batches. Red path: Traditional approach involving manual diagnosis from veterinarian examinations, employed for parameter inference of a knowledge-driven forecast. Blue path: Integration of lung ultrasound sensor for automated data collection labelled with veterinarian-established ground truth. followed by a data-driven model trained to automate the diagnoses which are re-used along with their confidence levels in several ways to improve and automate the forecast through a knowledge-driven model enabling tailored interventions.

2.1. Data acquisition process

The data used in the experiments were gathered from nine fattening farms in western France. Each farm managed up to three batches simultaneously, with each batch consisting of an average of five to twelve calves. Of the total cattle, 78 % of Charolais breed, 12 % of Limousin, and 10 % mixed breeds. The Charolais breed was predominantly chosen because it is among the most prevalent beef breeds raised in France. The experimentation on each farm started on the day the cattle arrived (Day 0), for a duration of one month. The initial weeks following their arrival are considered the period when cattle are most susceptible to BRD (Babcock et al., 2009). Consequently, the data collection period spanned from January to June 2023. To simplify readability of the paper, let's denote by $t \in \{1, 5, 14, 21, 28\}$ the observation dates. The selection of examination dates varied among farms and depended on the availability of both farmers and a veterinarian.

2.1.1. Sensor data

Portable ultrasound scanners were employed to evaluate the animals' lung conditions on multiple days $t \in \{1, 5, 14, 21, 28\}$. The ultrasound scanner captured lung images in video format, featuring 28 frames per second in black and white, lasting up to 20 seconds each, with a resolution of 480×560 pixels. The animals' lungs were partitioned into eight inter- costal zones, following a standardised scanning protocol from the shoulder to the stomach: Ultrasound scanning was conducted on the 4th and 5th intercostal spaces. These spaces are the most cranial intercostal spaces accessible in animals of this size and age, as referenced by (Cuevas-Gómez et al., 2020). Each intercostal space was further divided into ventral and caudal portions, separated by the shoulder, resulting in four sections scanned per side for each animal. A veterinarian validated the dataset and, in some farms, ultrasound was only performed on one side of the animal because of the physical restraints we encountered. Within the local farming system, it was challenging to establish a recruitment criterion that mandated access to both sides of the thorax, alongside other criteria such as internet access and a sufficient number of animals purchased annually. The process of recording videos with an ultrasound scanner proved to be time-consuming, tedious, and challenging, requiring the animals to be kept in a fixed position. Consequently, some animals were prioritised for the study, only the ones having lesions larger than 1 cm2 were deemed meaningful (Masset et al., 2022). If multiple significant lesions ($> 1cm^2$) were observed in the same intercostal area, only one video was recorded, corresponding to the lesion with the largest surface area. On Day 0 ($t = 0$), one-third of the animals of the three batches underwent examination, while from Day 2 to Day 28 ($t > 0$), all animals in a selected batch were examined. To maintain a balanced dataset, videos of clinically healthy lungs, without ~~of~~ lesions, were also recorded. To simplify further explanation, we denote by $\mathscr{X}^{obs} = \{x_1^{obs}, x_2^{obs}, ..., x_i^{obs}, ..., x_n^{obs}\}$ the space containing the Lung UltraSound (LUS) video $x_{n,t}^{obs}$ for each individual i at time t. Note here that at a given time t, an animal having lesion (so scanned) in a given batch may vary since some lesion may disappear over time.

2.1.2. Ground truth data

Ground truth data consists in the identification of diseased animals, based on clinical and biological assessments. Multiple veterinarians participated in the examination of the animals, having undergone identical training to minimise annotation bias. Clinical assessments involved the examination of rectal temperature, lethargy, nasal/ocular discharge, and quantification of breathing rate/amplitude. Animals were categorised as clinically diseased if their rectal temperature exceeded 39.7°C (Timsit et al., 2011) and if they had at least one more clinical symptom, which helped to reduce false positives caused by non-infectious events. Biological assessments involved the use of nasal swabs to determine the presence or absence of BRD associated pathogens (*Mannheimia haemolytica*, *Pasteurella multocida*, *Mycoplasma bovis*, *Histophilus Somni*, Bovine Respiratory Syncytial Virus, Para-influenza Virus type 3, bovine coronavirus, Influenza D virus). This choice was primarily made due to restraint issues, the duration of the procedure, and the invasiveness of the alternative methods. Following collection, the swabs were immediately placed on ice and transported to the laboratory within three hours. There, the two swabs from each animal were suspended in 400–600 μL of phosphate-buffered saline (PBS) and stored at −80°C. PCR analyses were performed within a maximum of 6 months from freezing at the BIOEPAR laboratory (UMR BIOEPAR, France), using a range of commercial kits (BIOTK051, BIOTK052, BIOTK053, BIOTK054, Biosellal, France). An animal was considered biologically diseased if it had at least one of those pathogens. The final decision of whether an animal was considered infected or not was made using a combination of clinical and biological results (see Table 1). Clinical signs of BRD are often observed late in the disease course (Schaefer et al., 2007), this is partly caused by the fact that sick animals tend to hide their clinical signs. Therefore, in this study, we hypothesised that biological exams provide more informative insights than clinical exams. Consequently, when there was a disagreement between the two types of exams, the biological exam was considered the final decision. To simplify further explanation lets denote by $\mathscr{Y}^{obs} = \{y_1^{obs}, y_2^{obs}, ..., y_i^{obs}\}$ the space containing the infection

state $y_{i,t}^{obs} \in \{Infected, non-infected\}$ for each individual i at t ∈ {1, 5, 14, 21, 28}

Table 1 describes the data used for this study and Fig. 2 illustrated how we went from 5 to 12 animal per farm to 1254 individuals in total. In total, our study involved 9 farms, with each farm hosting up to 3 batches simultaneously. Each batch typically consisted of 5–12 animals. Veterinarians examined these animals over approximately 30 days, with multiple observation dates (e.g., t=1, 5, 14, etc.). We treated each observation of an animal as a distinct statistical individual. Although we initially had 12 unique animals, the total number of observations (and thus statistical individuals) increased to 1254. They were a total of 202 different bulls in the 1254 observations, the median number of observations per bull is 1 and the interquartile range is 4. It is important to clarify that these 1254 observations do not represent unique animals, as some were observed multiple times on different dates. A total of 1254 individual data points clustered in individual animals $\mathscr{Y}^{obs}$ were examined by veterinarians to determine their infection state $\mathscr{Y}^{obs}$. Note that, only a subset of individuals in $\mathscr{Y}^{obs}$ were recorded for LUS videos, specifically 163 individuals. And since each individual can be scanned on different zones (depending on the zones containing lesions), the total number of recorded LUS video $\mathscr{X}^{obs}$ amounts to 265. Each recorded LUS video in $\mathscr{X}^{obs}$ is associated to a unique infection state in $\mathscr{Y}^{obs}$ such that we have 265 independent pairs $(x_1^{obs}, y_1^{obs}), ..., (x_{265}^{obs}, y_{265}^{obs})$ This also means that 1091 individuals (87 %) in $\mathscr{Y}^{obs}$ did not have a recorded LUS video in $\mathscr{X}^{obs}$.

2.2. Baseline scenario

2.2.1. BRD mechanistic model

In this study, we employed a well-established epidemiological mathematical model (Fig. 3) for BRD developed by (Picault et al., 2019a). This model is a stochastic individual-based mechanistic approach and had been initially calibrated (Picault et al., 2022) to investigate the influence of farming practices, particularly pen size and associated risk levels, on the dynamics of BRD, including its morbidity, severity, lethality, and antimicrobial usage. Their results highlighted the significant impact of risk reduction strategies implemented during pen formation, as well as the effectiveness of collective treatments in mitigating BRD incidence within high-risk pens.

Implementation and simulation of BRD mechanistic model was

Table 1
Infection State assignment guidelines. Observations can either be clinical or biological. Decision of infection state of an animal is based on the combination of observations. Quantities are subdivided into 3 categories.

Observations		Decision	Quantities		
Clinically diseased	Biologically diseased	Infection state (Y_i^{obs})	No. of observations examined[1]	No. of observations scanned[2]	No. of LUS video recorded[3] ($\mathscr{I}^{obs}$)
Yes	Yes	Infected	68	15	31
Yes	No	Non-infected	15	2	2
No	Yes	Infected	128	9	19
No	No	Non-infected	30	4	6
NE*	Yes	Infected	0	0	0
NE*	No	Non-infected	0	0	0
Yes	NE*	Infected	93	15	29
No	NE*	Non-infected	910	118	178
NE*	NE*	Undefined	10	0	0

[1] Total number of animals examined by veterinarians
[2] Total number of animals having gotten scanned for LUS
[3] Total number of LUS Videos recorded
* (NE) Not Examined animals

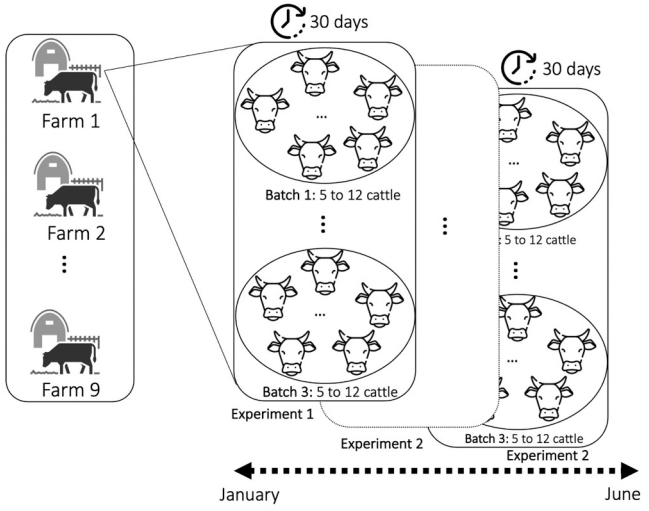


Fig. 2. Total number of individuals. Going from farm having 5–12 animals, we end up with over 1000 individuals if we consider each individual to be independent at each observation date.

facilitated by the EMULSION platform (Picault et al., 2019b), allowing the depiction of all model components in a human-readable, flexible structured text file processed by a generic simulation engine.

This facilitates collaboration and model refinement by scientists with diverse backgrounds, including veterinarians and epidemiologists.

2.2.2. Model calibration

The reliability and realistic performance of a mechanistic model depends on the behaviour of its internal parameters. Parameter inference refers to optimally adjusting the values of its parameters such that the mechanistic model can accurately represent real-world observations. In other words, inference fine-tunes these parameters, aligning the BRD mechanistic predictions closely with actual veterinarian diagnosis. The sensitivity analysis (Table 2) in (Picault et al., 2022), aiming to

Table 2
Definition of the three parameters of the BRD mechanistic model to be estimated.

Parameter definition	Notation	Min	Default	max
Initial Condition				
Initial Prevalence (No units)	θ_1	0.0	0.1	1.0
Epidemiological parameters				
Average duration in Infectious State (hours)	θ_2	0.0	120	+ inf
Pathogen transmission rate (h^{-1})	θ_3	0.0	0.008	1.0

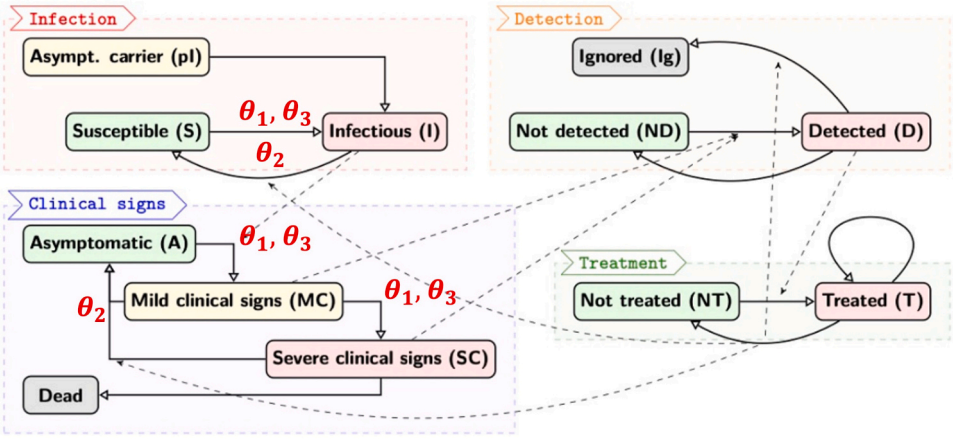


Fig. 3. Overview of the process's representation in the mechanistic BRD model (Picault et al., 2022). It integrates four processes (infection, clinical signs, detection, treatment). Animals transitioning to the infectious state (I) also exhibited mild clinical signs (MC), which can be detected (D), leading to initial treatment (T) that can be repeated. Successful treatment causes the animal to revert to susceptible (S) and asymptomatic (A) states. If successive treatments failed, the process will stop, and the animal will be 'Ignored' (Ig).

understand the importance of each parameter in the final prediction, emphasised the significance of parameters like the pathogen transmission rate and the average duration in the infectious state, critical for managing antimicrobial usage and mortality rates. In the present study, we focused on determining the optimal values of three crucial parameters, θ_1, θ_2, θ_3 as described below.

Table 2 outlines the meaning and permissible range of each parameter. The default values for these parameters are derived from the original publication of the mechanistic model by Sébastien Picault (Picault et al., 2019a): Initial Prevalence represents the initial proportion of infectious animals within the batch. Average Duration in Infectious State indicates the average time an animal remains in the infectious state. Pathogen Transmission Rate reflects the average rate at which a pathogen is transmitted from one animal to an- other. To estimate these parameters, we employed a likelihood-free inference method called Approximate Bayesian Computation (ABC) (Beaumont, 2019), specifically regression-based correction methods. This method entailed sampling randomly, in this particular case 100, parameter values within a biologically acceptable domain and utilising them to generate simulated datasets through the "average" pathogen model. Selected summary statistics, specifically, here the count of symptomatic animals, captured essential features of the observed data and related predictions of the deep learning model. The similarity between simulated and observed data was assessed using the Euclidean distance, and the top 1 % of sampled parameters were accepted. This process facilitated the estimation of the distribution of potential values for the selected parameters. A type of regression-based ABC method, ABC-NN (i.e., regression by neural network), was chosen.

It's noteworthy that the Euclidean distance, with equal weights assigned to every point, was used as the distance measure to compare observed and simulated data. This implies an assumption that all observations were made with the same confidence level.

2.2.3. Baseline implementation

In conclusion of this section, we created a scenario Fig. 4. Ground truth data (Table 1) is used to determine the infection status of each animal $y^{obs} \in \{Infected, non-infected\}$. By aggregating these results, we obtained the total counts of infected and non-infected animals over time, we refer to this information as diagnostics (Fig. 1). Similarly, since the mechanistic model (Fig. 3) can predict the numbers of infectious (I), susceptible (S), and asymptomatic carriers (pI), if we express the model's output as non-infectious being the sum of Susceptible Asymptomatic Carriers, then we can obtain the total counts of infected and non-infected animals over time, we refer to this information as forecasting (Fig. 1). Diagnostics are made during veterinarians' visits (once per week on average) whereas forecasting allows a continuous monitoring of the total number of infected animals because its predictions are computed more often (every 12 hours). The ABC parameter inference is employed to the estimate the optimal values of the three crucial parameters of the BRD mechanistic model such that its forecasting path can pass closest to the veterinarian diagnostics. This way we can better study the evolution of the disease. In theory, since the mechanistic model follows a set biological rule, its predictions could help inform future diagnostics of veterinarians. This scenario serves as our baseline, against which the diagnostic and forecast performance of subsequent methods presented in this paper will be discussed.

2.3. Automatic diagnostics from a punctual estimator

2.3.1. Deep learning

The baseline method outlined in the previous section is a classical approach used by modellers to construct and estimate the parameters of epidemiological models using real-world observations. This approach relies on veterinarians' diagnostics, necessitating clinical and biological examination of every animal during each visit. This implies that a farmer would have to schedule a veterinarian visit, round up all the animals, individually consult each one—a process that is laborious, tedious, costly and susceptible to human errors. Given the many obstacles to regular diagnosis, we propose in this section an approach aimed at automating the diagnosis. We developed in this section a spatiotemporal convolutional neural network (deep learning model) to predict the infection state $y_i^{obs} \in \{Infected, non-infected\}$ of individual animals using their LUS videos x_n^{obs} as input.

The processing of the input dataset $\mathscr{X}^{obs}$ (LUS videos) followed several steps. Initially, we addressed data imbalance in our LUS video database. Out of the 163 individuals recorded (Table 1), 24 % were diagnosed as infected animals by the veterinarians, and the remaining 76 % were non-infected animals. To create a balanced dataset, we applied a downsampling strategy to reduce the total number of non-infected individuals in the database. However, even with an equal number of individuals in each class, 70 % of LUS videos belonged to non-

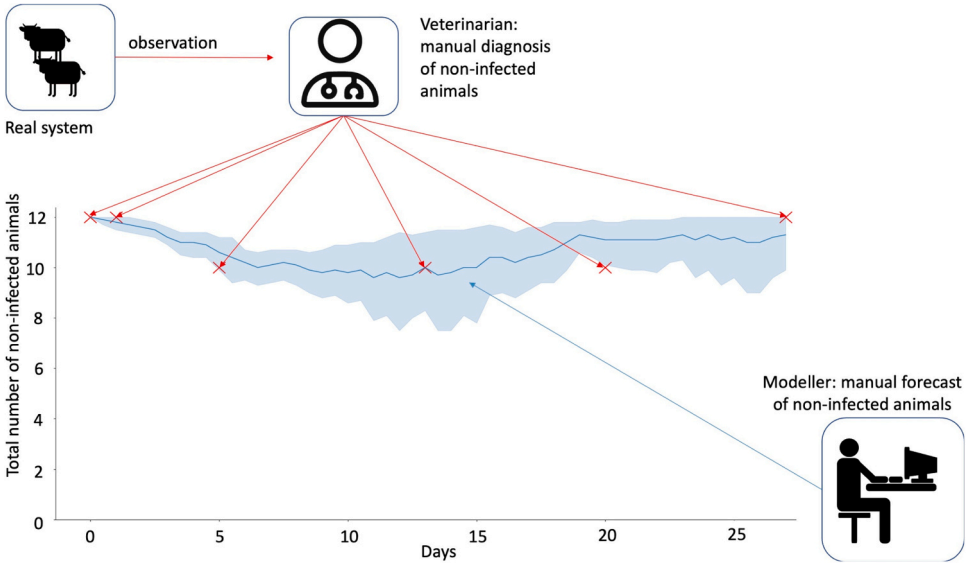


Fig. 4. Baseline Scenario: Traditional BRD study method - Veterinarians diagnose at different time steps animal infection states based on clinical and biological observations. Using these diagnoses, a modeller employs ABC inference to parameterise an average pathogen BRD mechanistic model (via Emulsion). This model enables the forecast of the total number of non-infected animals over a 30-day period.

infected animals due to some animals being scanned multiple times in different zones (see Sensor Data). Consequently, a second downsampling strategy was applied to reduce the number of LUS videos in the non-infected category. In the second step, the data obtained in the first step was split into a training set (comprising 70 % of videos), a validation set (comprising 20 % of videos), and a test set (comprising 10 % of videos). We ensured data integrity by preventing different videos of the same individual on the same day from appearing simultaneously in different sets, thus avoiding data leakage. For the test set, we used the complete data from the two farms that were set aside earlier when building the baseline method. In other words, the test dataset is both used to evaluated the performance of the deep learning model (automatic diagnostics) and mechanistic model (forecasting). Handling ultrasound videos posed challenges due to varying frame counts (technical issues when recording) and noisy images (moving animals). The solution employed, similar to handling text sequences, involved extracting images from the videos up to a maximum count and if a video had fewer images, the missing frames would be padded with zeros (Birnbaum et al., 2019).

A video encompasses both spatial information within individual frames and temporal information across the entire sequence. To effectively address both aspects in video analysis, we opted for a hybrid architecture, specifically a spatiotemporal convolutional neural network. In our approach, we integrated convolutional layers (Krizhevsky et al., 2017) with recurrent layers (Bengio et al., 1994). The convolutional layers focus on extracting spatial features, such as lesions, pleura lines, or other relevant anatomical details, while the recurrent layers capture temporal information related to the sequence or frequency of appearance of spatial features. For additional details about the architectures, including the chosen backbone, the number of layers, etc.

The code for implementation is written in Python (Version 3.8), utilising Tensorflow (Version 2.15.0) and Keras (Version 3.0.0) libraries to build, train, and evaluate the deep learning architecture. Optimisation was performed to find the best hyper-parameters using the Keras-Tuner library (Version 1.4.6). Additionally, experiment tracking was employed to monitor trained models through the use of the MLflow library (2.10.2).

2.3.2. Scenario simulation

Out of the nine farms, we set aside the data from two farms (testing set presented earlier), creating a sampling pool $\mathcal{A}$ of LUS videos x_i^{obs}. In a perfect world every examined individual $y_n^{obs} \in \{Infected, non-infected\}$ in those two farms should have been recorded for LUS video x_n^{obs}, however that is not the case here. To solve this issue, we had to artificially re-create a scenario where each examined individual has at least one corresponding LUS video. Algorithm 1 illustrates how a bootstrap sampling algorithm was used to generate a LUS video set x_n^{sim} for each examined individual over time. This enables us to create a proper set as input for an automated pipeline such as the one presented in this work.

Algorithm 1. Population simulation through bootstrap sampling

Input : a scalar of the observed animals over time y_n^{obs}
Output: a scalar of the simulated LUS x_n^{sim}
Initialisation;
  Create sampling pool of LUS video $\mathcal{A}$;
Re-sampling Process;
for *each Time Step (Examination Day)* **do**
  for *each animal examined* y_i^{obs} *in* $\mathcal{Y}^{obs}$ **do**
    Select randomly (with repetition) a LUS video from A such that $x^{sim} \leftarrow \pi(A)$;
  end
end

Out of the 9 farms 5 (Fig. 5), data from 2 farms were set aside specifically to evaluate the performance of the deep learning model, while the remaining 7 farms' data were used for training. To address the issue of missing data (38 videos), the test sample was reused as a sampling pool, allowing us to create a complete dataset. This sampling pool was ultimately used to assess the performance of the method we developed. The baseline accuracy represents the optimal outcome given that diagnoses are made by veterinarians and forecasts are based on their predictions. Therefore, our goal is to achieve an accuracy as close to this baseline as possible.

2.3.3. Method 1

In contrast to the baseline scenario where veterinarians manually provided diagnostics, in the current scenario, we trained a deep learning model to automatically diagnose $y_i^{obs} \in \{Infected, non-infected\}$ all individuals using only their LUS video x_i^{obs} (Fig. 6). In other words, we replaced the veterinarian diagnostics in Fig. 1 with predicted diagnostics. The subsequent steps mirrored the baseline approach, utilising ABC-NN to calibrate a BRD mechanistic model for optimal alignment with the predicted diagnosis. Regarding the diagnostics, it is important to note that we had to assign a virtual LUS video x_i^{sim} to every individual y_i^{obs}since they were not all scanned (as presented in the previous subsection "scenario simulation"). The evaluation of method 1 involves three steps: first, we compared its diagnostics performance with that of the baseline using the Relative Root Mean Square Error (RRMSE).

$$\mathrm{RRMSE}(y_i^{diag,pred}, y_i^{diag,obs}) = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i^{diag,pred} - y_i^{diag,obs})}{\sum_{i=1}^{n} y_i^{diag,obs}}} * 100$$

Secondly, we assessed its forecasting performance relative to the baseline by employing the same metric, RRMSE. Finally, we compared the posterior distribution of parameters obtained using method 1 with that obtained under the baseline scenario.

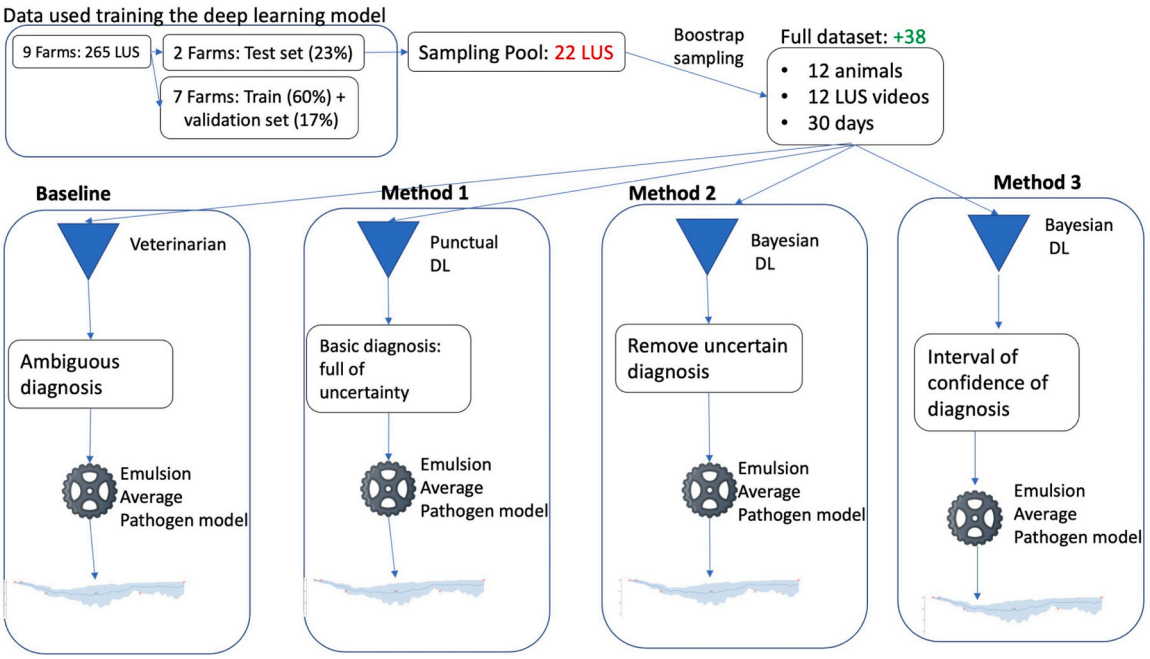


Fig. 5. Flow diagram depicting the data management process. One portion of the data was allocated for training and testing the deep learning model, while the remaining data was used to create a sampling pool. This pool was then utilized to generate a comprehensive dataset for evaluating the newly developed methods.

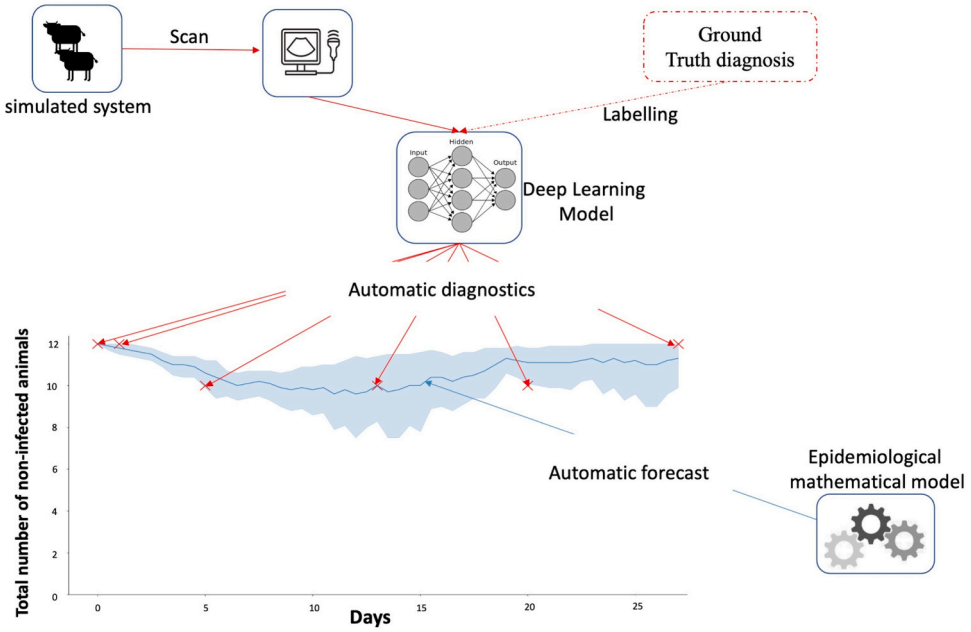


Fig. 6. Method 1 - Automating the traditional BRD study method. A deep learning model is trained on LUS videos and veterinarian examinations (utilized only during training) to provide punctual diagnoses of animals. These diagnoses are then used to parameterise three key epidemiological parameters of the average pathogen BRD mechanistic model using ABC inference. This approach enables automated diagnosis and forecast of the number of non-infected animal over a 30-day period.

2.4. Uncertainty-aware diagnostics

2.4.1. Uncertainty estimation

Method 1 integrates a data-driven model with a knowledge-driven model to automatically diagnose and forecast BRD. The output probabilities (softmax layer) from a deep learning model are often misinterpreted as indicative of model confidence, sometimes leading to wrong predictions with high-confidence; A situation with potentially significant real-world consequences. What if we could enhance Method 1 to distinguish between risky and confident diagnostics, allowing experts to exploit the most informative predictions for crucial decisions-making, while applying supplementary examinations to challenging diagnostics? Model uncertainty refers to the inherent lack of confidence or certainty that a model may have in its predictions, acknowledging

that the model may not always be unequivocally certain about the correct answer. Various factors, such as limited data, data ambiguity, model complexity, and noisy data, can contribute to model uncertainty. A common strategy to quantify uncertainty in deep learning predictions is the use of Bayesian Deep Learning (BDL) models. Unlike traditional models, BDL treats model parameters as probability distributions rather than point estimates. This approach enables the model to offer not just predictions but also a posterior distribution representing the range of possible values. Incorporating this element of uncertainty provides a more nuanced understanding of the model's output. In BDL, the posterior distribution captures the model's updated beliefs about parameters after observing the training data, offering a valuable measure of confidence and variability in predictions.

In this study, we incorporated Monte Carlo Dropout (MCD), a previously established method to estimate deep learning model uncertainty (Gal and Ghahramani, 2016). MCD involves employing dropout during both training and testing phases in neural networks. During training, dropout randomly deactivates certain neurons, compelling the network to learn robust features and avoid over-fitting. During testing, the model is assessed multiple times with dropout enabled, leading to varying probabilities in the softmax layer based on the dropout rate (proportion of deactivated neurons). Instead of yielding a prediction with a fixed softmax probability, MCD produces a probability distribution for each class, with the predicted class being the one with the highest average probability (Bayesian model averaging). The uncertainty of the prediction is then quantified here by computing the Shannon entropy over the probability distribution of the classes. A higher entropy value signifies greater uncertainty, as the probability mass is spread across multiple classes. The Shannon entropy, introduced in deep Bayesian active learning, usually serves as an acquisition function, guiding the selection of training data with the highest uncertainty in hopes of enhancing model generalisation.

Following the implementation of a Bayesian approximation in our deep learning model, the final step is to determine the uncertainty threshold that optimally separates uncertain predictions from confident ones. This is achieved iteratively, evaluating the F1_score (harmonic mean of precision and recall) for each threshold value and identifying the value that maximises performance in the training set. The precision is also called the positive predictive value and the recall can be called the sensitivity.

2.4.2. Diagnostic filtering

Being able to obtain a prediction and its uncertainty level, we decided in this section to re-use the uncertainty to filter out noisy observations. Algorithm 2 depict the given instructions to sample only noise-free observations. Understand here that by filtering out noisy observations, we are shrinking the original dataset. To simplify understanding, let's denote $I_t^{filtered} = \{N_1^{filtered,infected}, \ldots, N_t^{filtered,infected}\}$ the space containing the count of infected animals in the filtered dataset per examination date t. $C_t^{filtered}$ is the count of total animals ($infected + non - infected$) in the filtered dataset per examination date such that $N_t^{filtered,infected} \leq C_t^{filtered}$.

The code for the implementation is written in Python (Version 3.8), through the use of libraries of Tensorflow (version 2.15.0) and Keras (version 3.0.0) to build, train and evaluate the Bayesian deep learning architecture. Baal library (version 1.9.1) was used to implement filtering with the Entropy function. Optimization was also performed to find the best hyper- parameters using library KerasTuner (version 1.4.6).

2.4.3. Method 2

To conclude this section, we introduced a scenario (Fig. 7) where our Bayesian deep learning model performs batch diagnoses for all animals during each examination date. Following the batch diagnosis, we filtered out animals with excessively high uncertainty in the predictions.

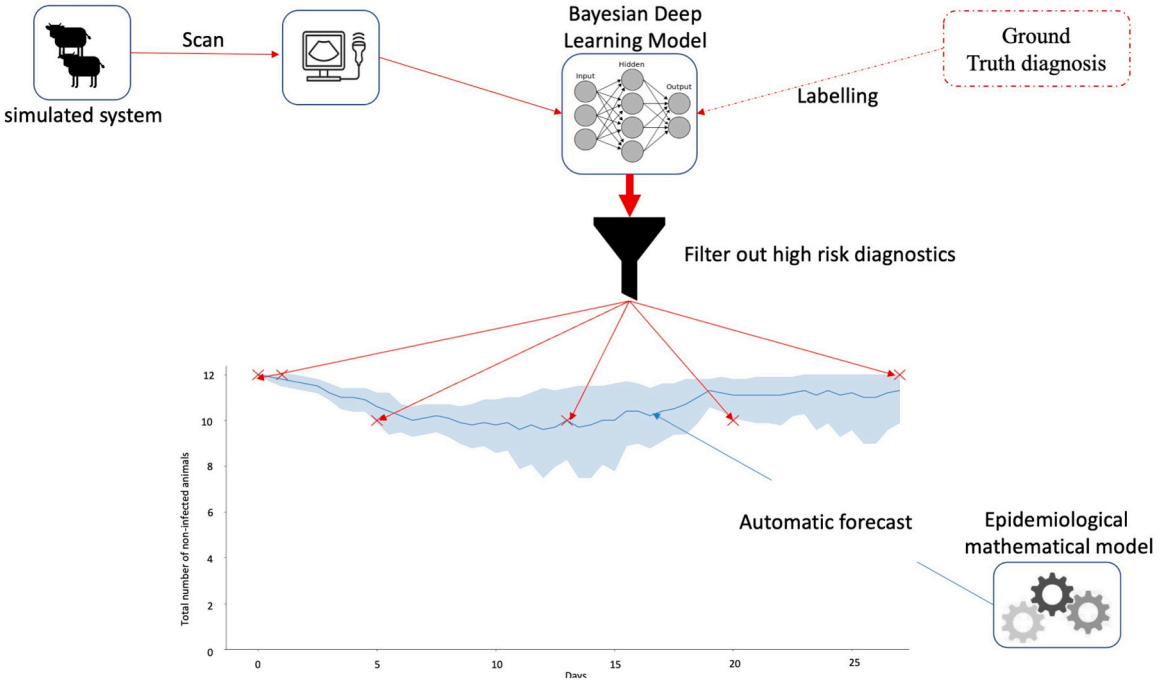


Fig. 7. Method 2 - Enhanced automated diagnostics through filtering. A Bayesian deep learning model is trained on LUS videos and veterinarian examinations (used only during training) to provide precise diagnoses with associated confidence levels. Low confidence diagnoses are flagged for further expert examination, while confident diagnoses are employed to parameterise the BRD mechanistic model. This approach ensures cautious diagnosis of the number of non-infected animal over a 30-day period.

Algorithm 2. Diagnostic adjustment through uncertainty-based filtering

Input : a scalar of the simulated LUS videos X_t^{sim} at each time step t
Output: a scalar of total approximated number of infected animals $\mathbf{I}_t^{estimated,infected}$ at every step t

Initialisation;
for *each time step t* **do**
  Get the total amount of initial animal $N_t^{observed,total}$;
  Count of total prediction kept $C_t^{filtered} = 0$;
  Count total infected animal filtered $N_t^{filtered,infected} = 0$;
  for *every* x^{sim} *in* X_t^{sim} **do**
    Diagnose using Bayesian deep learning $y^{predicted} =$
    Get uncertainty of prediction $y_t^{predicted}$ s.t. $\mathrm{U}(y_t^{predicted}) = \mathrm{Shannon.Entropy}(y_t^{predicted})$;
    if $\mathrm{U}(y_t^{predicted}) \leq \epsilon$ **then**
      Increment the total number of confident prediction, C_t += 1;
      if $y_t^{predicted} == infected$ **then**
        Increment number of infected prediction, $N_t^{filtered,infected}$ += 1;
      end
    end
  Get ratio of infected animals in filtered data, $\mathrm{R}_t = \frac{N_t^{filtered,infected}}{C_t^{filtered}}$
  Approximate the real number of infected animals, $I^{estimated,infected} \approx \mathrm{R}_t \times N_t^{observed,total}$;
  end
end

The ratio of infected animals in the filtered dataset was then used as an estimator for the true ratio of infected animals in the initial unfiltered dataset. This approach allowed us to obtain a revised diagnosis of the total number of infected animals per examination date. Similar to the scenarios outlined in the previous section, we used this new diagnostic to derive the epidemiological trajectory of the number of infected animals using the BRD mechanistic model. As in the previous section, we assessed the performance of this method (diagnosis, forecast, parameter inference) compared to the baseline, employing the same metrics as for method 1.

2.5. Uncertainty-aware deep mechanistic model

2.5.1. Ensemble estimator

Using a Bayesian deep learning model allows us to generate predictions along with as- sociated uncertainties. In this section, these uncertainties were leveraged to establish a confidence interval for the number of infected animals around the predicted value. For simplicity, let's denote $G(.)$ as the function that takes an LUS video input (x_t^{sim}) and predicts its infection state state ($y_t^{pred} \in \{infected, non-infected\}$) as a probability assigned to each class. Through the MCD technique, each execution of the Bayesian deep learning model on an input yields slightly different probabilities for each class. In the absence of noise in the input, the correct class will have the highest probability. However, with noisy inputs, the model may occasionally assign a higher probability to the wrong class.

In essance running a Bayesian deep learning model multiple times on a noisy input will, in most cases, predict the correct class, but occasionally, it may predict the wrong class. This implies that when $G(.)$ is applied multiple times to all animals (x^{sim}) in a batch ($\mathrm{X}^{\mathrm{sim}}$), it does not consistently diagnose the same number of infected animals. Running $G(\mathrm{X}^{\mathrm{sim}})$ multiple times outputs a distribution of the number of infected animals $P(N_t^{infected})$ in the batch. From this posterior distribution, the confidence interval of the number of infected animals is derived, and the uncertainty of the diagnostics can be interpreted either as the variance of the distribution or its entropy.

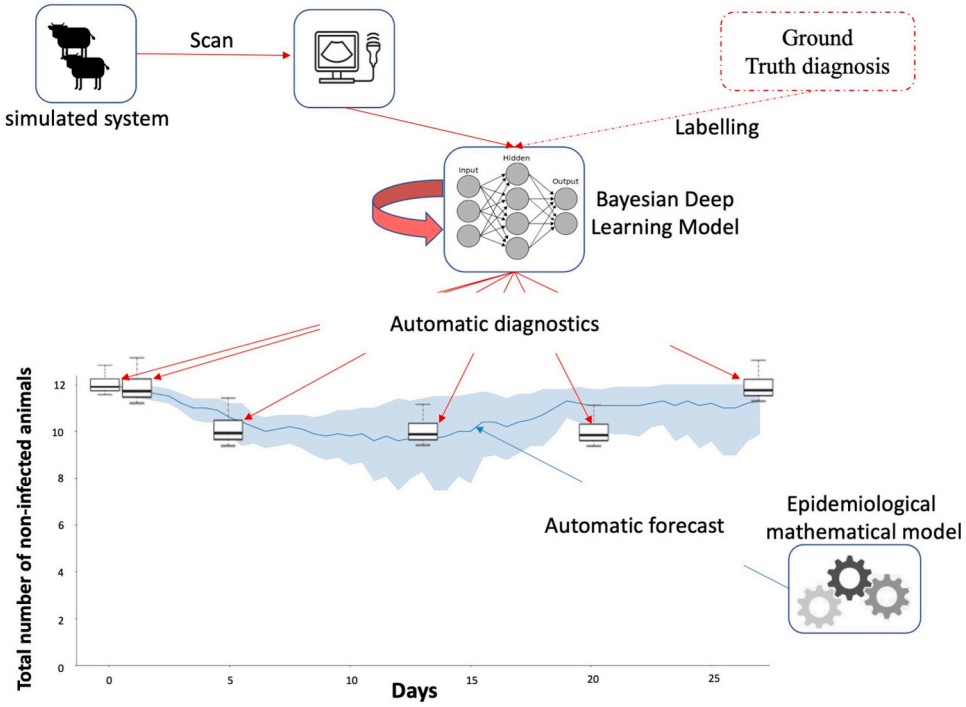


Fig. 8. Method 3 - Enhancing diagnosis and forecast through uncertainty propagation. A Bayesian deep learning model trained on LUS videos and veterinarian examinations (utilised only during training) provides an interval estimate of the total number of non-infected animals at different time steps. These diagnosis uncertainties are then propagated to the forecasting model (mathematical model), thereby enhancing its accuracy.

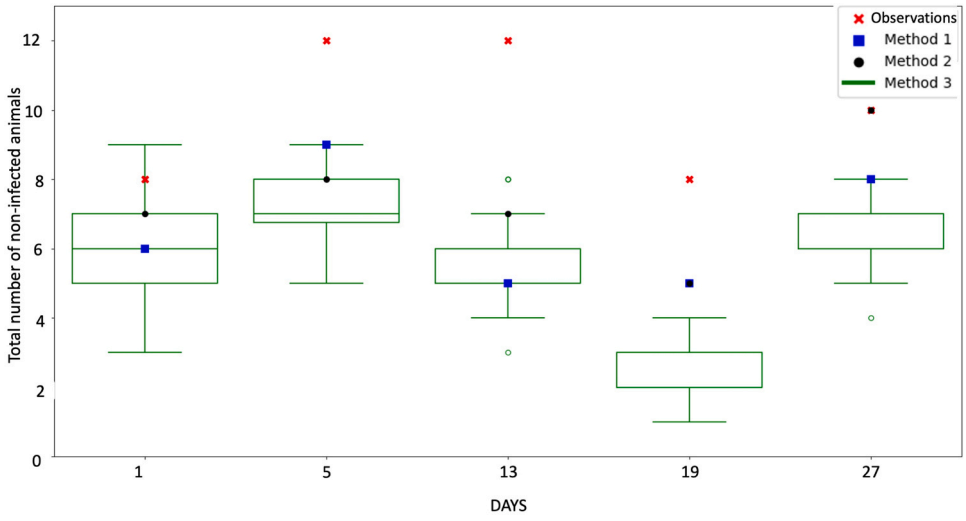


Fig. 9. Diagnostic performance comparison - Predicting the total number of non-infected animals in a farm batch at various observation dates. Methods 1, 2, and 3 employ Lung ultrasound videos as input, while baseline diagnoses are based on veterinarian observations, considered ground truth.

2.5.2. *Method 3 - uncertainty propagation*

In summary of this section, we formulated a scenario (Fig. 8) in which our Bayesian deep learning model forecasts a range of variation for the number of infected animals per examination date. The average prediction from the distribution of infected animals per examination date serves as the updated diagnostic. We compared this revised diagnostic with the baseline using the RRMSE. Furthermore, we introduced a step to incorporate the uncertainty from the revised diagnosis into the forecasting phase. During the parameter calibration of the mechanistic model, weights were assigned to each observation. The distance metric employed in the ABC-NN process is characterised by the weighted Euclidean distance, defined by the expression:

$$D_{\text{weighted}}(y^{obs}, y^{sim}) = \sqrt{\sum_{i=1}^{n} w_i (y_i^{sim} - y_i^{obs})}$$ Where, y^{obs} is the revised diagnostics, y^{sim} is the simulated diagnostics (by the mechanistic model) and w_i is the weight of the revised diagnostics expressed as the variance of distribution of repeated batch diagnostics. The forecast performance of the mechanistic model fitted with in this scenario was compared to that of the baseline using the RRMSE metric. Similarly, the estimated parameters were also compared with those of the baseline.

3. Results

Results are organised as follows: first, we compared the diagnostics performance of each method (1, 2, 3) to that of the baseline method.

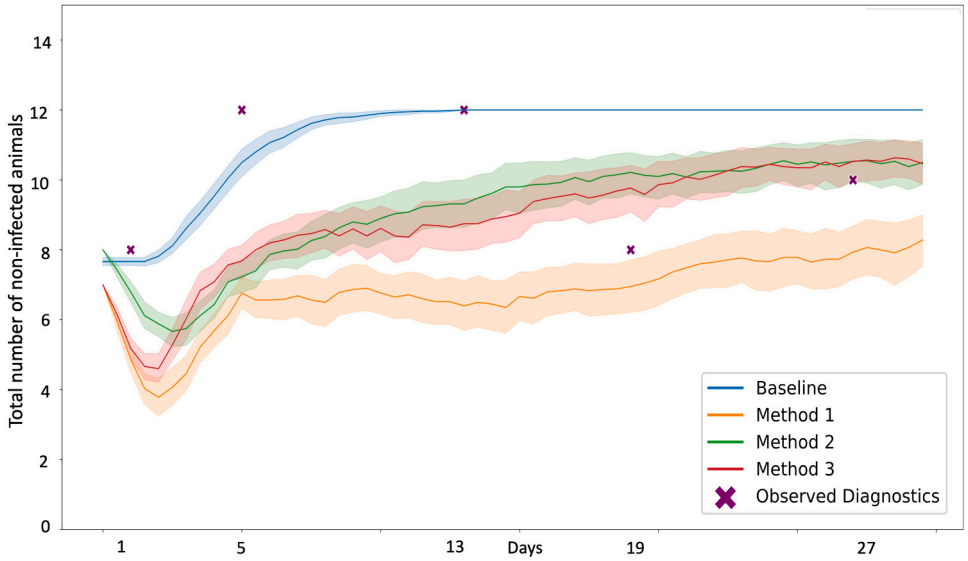


Fig. 10. Forecast Performance Comparison: predicting the progression of non-infected animals in a batch over 30 days. Ground truth diagnostics (cross) are based on veterinarian examination. Baseline forecast is fitted to ground diagnostics, while methods 1, 2, and 3 are fitted to their respective diagnoses (paragraph 3.1).

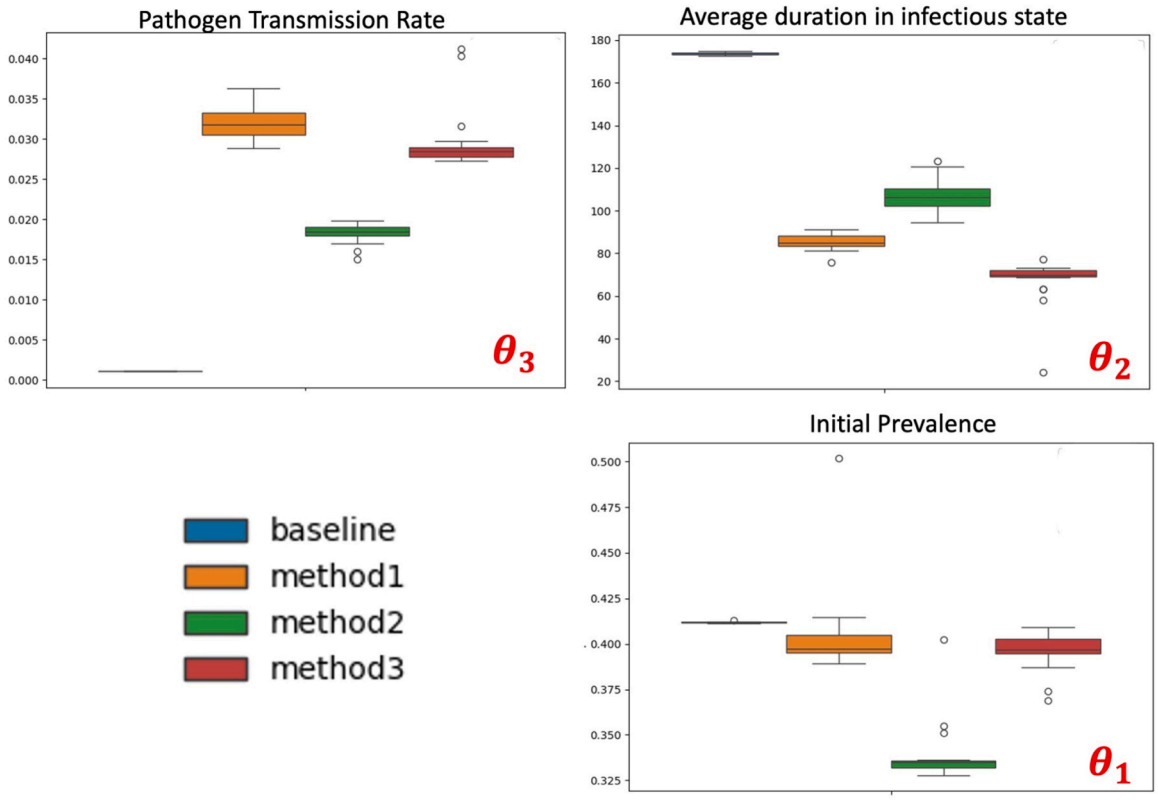


Fig. 11. Parameter inference performance comparison - Estimated distribution of pathogen transmission rate θ_3, average duration in infectious state θ_2, and initial prevalence θ_1 across methods. The goal is to be as close as possible to this baseline.

Secondly, we compared the forecast of each method to that of baseline. Third we compared the value of the epidemiological parameters of each method to that of the baseline.

3.1. Diagnostics

Applying each method presented in Section 3, 4, 5 to diagnose the total number of infected animals in a batch at different time step yield different results.Fig. 9 illustrates the diagnostic performance for each method compared to the baseline: baseline diagnostics (represented by crosses) are supposedly the most confident diagnostic that can be obtained, they were established based on veterinarians' examinations (clinical and biological). Method 1 (represented by squares), where a deep learning model is used as a punctual estimator, yields the highest errors compared to the baseline diagnostics with a RRMSE of 39 %. Note here that the deep learning model in itself after optimisation obtained a F1_score of 72 %. Method 2 (black dots), where the uncertainty of a Bayesian deep learning model is used to improve its predictive

performance, yields the lowest diagnostic errors compared to the baseline with a RRMSE of 32 %. In, method 3 (box-plots), where a bayesian deep learning model is used to diagnose in the form of an interval of variation, it is the mean of the interval that is used final value and which is compared to the baseline diagnostics. Method 3's performance is equivalent to that of method 2.

3.2. Forecasting

Applying method 1, 2, 3 to forecast using their previous diagnostics yield different results.

Fig. 10 illustrate the performance obtained for each method tested. The forecast of each method was fitted to their respective diagnostics (Fig. 9), note that for method 3, its forecast was adjusted to the mean diagnostic (from the box-plot). The ultimate goal here is to obtain forecasts that are the closest to the baseline. Method 1, where diagnostics are estimated using a deterministic predictor yields the worst results with an RRMSE of 38.4 %. method 2 and 3, where the uncertainty (or confidence) of each diagnostic is evaluated and re-used, yields the best results with an RRMSE of 27.2 %. The performance of method 2 and 3 is close to the baseline forecast (which was fitted on veterinarian diagnostics) the RRMSE is 23 %.

3.3. Parameter inference

The infectious dynamic (forecast) obtained for each method is characterised by different parameter values. These parameters represent epidemiological phenomenon's.

Fig. 11 illustrates the estimated parameter values for each method. The pathogen transmission rate, an epidemiological parameter indicating how rapidly a disease spreads among animals, is closest to our reference values (baseline) when using method 2. Similarly, the average duration in the infectious state, describing the period during which an infected animal can transmit the disease, aligns closer with our baseline when estimated using method 2. Initial prevalence refers to the proportion of infectious animals at the outbreak's onset. Interestingly, methods 1 and 3 demonstrate better estimation accuracy than method 2 in this regard. Despite similar forecast performances between methods 2 and 3, their estimated parameters differ significantly. BRD was an example of application of our methodology.

4. Discussion

4.1. Data acquisition process

In our experimental protocol, we structured our sampling periods ($t \in 1, 5, 14, 21, 28$) and selected animals for examination based on constraints posed by the availability of farmers and veterinarians. Due to these limitations, we were unable to collect data from every animal, leading us to employ bootstrap sampling to generate a comprehensive dataset. Despite this challenge, bootstrap sampling provided us with the most objective approach to ensure dataset completeness, given the circumstances. The encouraging results obtained might incite end-users to modify the experimental sampling protocol accordingly. deep learning performances are proportional to the size of the training set. Hence scaled up to commercials herds, if including a new training phase, should in theory lead to better performances. But if complete herds were not going to be scanned, how animals might be prioritised amongst the herd is a trickier question. By analogy with standard active learning techniques (Settles, 2009) like query-by-committee, one could rely on inter raters' disagreement on clinical signs for such animal selection. The idea for such selection is that the less the raters agree on the status of a certain individual, the more information this individual will give about the frontier between (health) states. Hence, animals could be chosen for biological and scan exams according to maximum inter raters' certainty (healthy or not) and uncertainty (to try to find the best discrimination between states).

Regarding the sensors used, we relied on lung ultrasound videos, a tool highlighted in the literature (Ollivett and Buczinski, 2016; Timsit et al., 2019) for its efficacy in quickly identifying caudal lung lobe consolidation, lung necrosis, and lung abscessation. However, certain limitations should be acknowledged. Only lesions adjacent to the pleura lines were visible, thereby restricting visibility of deeper lesions. Animals underwent TUS on both sides of the thorax in 6 out of 9 farms. However, in the remaining 3 farms, this was not feasible due to the type of restraint used. Within the local farming system, it was challenging to establish a recruitment criterion that mandated access to both sides of the thorax, alongside other criteria such as internet access and a sufficient number of animals purchased annually. We acknowledge this as a limitation, as previous studies in calves have demonstrated relatively low agreement between consolidations observed on one side of the thorax versus the other (kappa value = 0.33) (Buczinski et al., 2014). Similarly, in our study, when evaluating TUS results from the first day consolidations were detected in an animal, the agreement between the two sides of the thorax was low (kappa value = 0.30). Additionally, discerning whether observed consolidations are active or old (scarred) tissues (Masset et al., 2022) remains challenging. Furthermore, our access was restricted to a fairly caudal part of the lung (from the 4th intercostal space), but we know that lesions typically originate in the cranial part (Masset et al., 2022), potentially leading to delayed diagnosis. We hypothesised that an infected animal could have at least one pathogen detected in the cranial part, but depending on the pathogen, lung damage might not always be visible (as the pathogen may remain only in the upper respiratory tract) or could be delayed, potentially resulting in false negatives. To address these limitations, future research could explore integrating various sensor data types, such as video surveillance for behavioural observation, audio recordings for detecting acoustic events (e.g., coughing, sneezing), and environmental parameters like CO_2, NH_3, and temperature. By incorporating multi-modal data simultaneously, akin to a veterinarian's holistic evaluation (visual, auditory, environmental), we could provide a more comprehensive assessment of an animal's health. Video surveillance coupled with audio surveillance seems particularly promising due to their synchronised nature and ability to capture the entire batch of animals more frequently than lung ultrasound examinations. This would enable synergy between complementary sampling methods, relatively a specific but continuous (surveillance) and specific but only punctual (ultrasound).

In any case, whatever the sensor or sampling method, establishing a consensus on ground truth data remains challenging due to the absence of a clear gold standard (Timsit et al., 2016). In this study, we defined rules for diagnosing the infectious state of animals, considering an animal infectious if it harboured at least one pathogen during PCR examination. However, this approach carries inherent risks, as not all pathogens play equal roles (Gershwin et al., 2015), potentially leading to an elevated rate of false positives. Addressing this ambiguity in ground truth labelling, recent work has introduced conformal prediction under ambiguous ground truth (Stutz et al., 2023), leveraging multiple expert opinions for uncertainty quantification. Employing such methods with clinical and biological outcomes, even when they disagree, holds promise for enhancing diagnosis accuracy.

4.2. Automatic diagnostics

Pre-processing data for deep learning models poses inherent challenges, and in our study, we addressed this by adopting a downsampling strategy, randomly removing labels to achieve a balanced dataset. However, we recognise that this approach is sub-optimal as it may result in the loss of information. To overcome this limitation, we propose the application of active learning (Gal and Ghahramani, 2016), a technique where algorithms actively select the most informative inputs from a large pool of unlabelled data for annotation by a human annotator. This method has demonstrated effectiveness in enhancing the performance of

deep learning models, particularly when dealing with highly imbalanced datasets (Lee and Seo, 2022).

Our work proves the importance of evaluating the confidence level of predictions for diagnosis and forecast. As future research could integrate multiple types of sensor data, automatic diagnostic models must adapt to accommodate multi-modal approaches. To our opinion, the most promising way forward is to employ deep audio-visual learning model (Zhu et al., 2020). It could be a promising avenue for enhancing diagnosis accuracy and forecast capabilities in such multi-sensor environments as it treats audio-visual problems as a whole not as separate parts, audio and visual.

4.3. Forecasting

In our current understanding, only two epidemiological mathematical models have been published in the literature. The first model was introduced in 2019 (Picault et al., 2019a), followed by the second model in 2023 (Sorin-Dupont et al., 2023). In this study, we employed the first model, which was calibrated under the assumption of an "average pathogen" infection to maintain simplicity. Despite parameterising this model using veterinarian diagnoses (considered the most confident values), we observed a relative average forecast error of 23 %. This discrepancy may be attributed to the multi-pathogen nature of BRD, where the prevalence of each pathogen and their interactions can significantly influence infection dynamics. To address this complexity, future research could explore the adoption of a pathogen-specific model (Sorin-Dupont et al., 2023), offering a more nuanced understanding of BRD dynamics. In our study, the epidemiological mathematical model was parameterised through an inference process using Approximate Bayesian Computation algorithms. We incorporated uncertainties of observational points (diagnoses) into the process, enhancing forecasts by weighting each diagnosis according to its confidence level, expressed as variance. While the inclusion of rules within the mathematical model has contributed to error reduction, there remains a need to enhance the integration between observational data, their confidence levels, and the mathematical model. For instance, ABC is only one of many likelihood-free methods (Drovandi and Frazier, 2022) that we could adapt to take advantage of uncertainty- awareness in the deep learning predictions, hopefully leading to even more accurate and reliable forecast.

5. Conclusion

This study represents a pioneering effort in leveraging sensor data, particularly lung ultrasound videos, to cautiously diagnose and forecast the progression of BRD in a batch. Our method is versatile as it does not rely on specific sensor types or disease characteristics, making it applicable across various contexts where early detection is challenging. While acknowledging the imperfections in our data acquisition protocol, the findings of this study highlight a functional coupling method. To render it operational for the study of Bovine respiratory Disease, it is imperative to improve the data acquisition protocol.

Funding

This project has been financially supported by Adventiel and ANRT through CIFRE No. 2021/1805 and by a grant of the Carnot Institute France Futur Elevage (project SEPTIME). This Work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101000494 (DECIDE). This publication only reflects the author's views and the Research Executive Agency is not responsible for any use that may be made of the information it contains. This work was supported by the French region Pays de la Loire (PULSAR grant) and the ministry of agriculture (MultiPast grant). No funder was involved in the scientific or editorial choices made for this study.

CRediT authorship contribution statement

Maud Rouault: Writing – review & editing, Validation, Resources, Funding acquisition. **Victoria Potdevin:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Xavier L'hostis:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Sébastien Assié:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Sébastien Picault:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization. **Nicolas Parisey:** Writing – review & editing, Validation, Supervision, Project administration, Conceptualization. **Théophile Ghislain Loïc Eyango Tabi:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgements

We express our gratitude to the SEPTIME project for providing valuable data, and we extend our thanks to Louis Anthore, Alison Boffelli, Charlotte Giraud and Marianne Guy, the veterinary students and the MultiPast project for their dedicated efforts in data collection. Special appreciation goes to our colleagues from the BIOEPAR Dynamo team for their insightful comments provided during the development of this work and Gaël Beaunée for his guidance throughout the inference process. Additionally, we sincerely thank Carole Toczé (Idele), the agricultural cooperatives (Bovinéo, Terrena) and the participating farms for their invaluable contributions to the experimentation.

Ethical approval concerning the French legislation on experimental animal care was approved by the Ethics Committee in Animal Experimentation in Oniris, Nantes, France (authorization on living animals No. CERVO-2022-7-V).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.prevetmed.2024.106354.

References

2019. Picault, S., Ezanno, P., Assié, S., 2019a. Combining early hyperthermia detection with metaphylaxis for reducing antibiotics usage in newly received beef bulls at fattening operations: a simulation-based approach. https://hal.science/hal-01987110..

Babcock, A.H., White, B.J., Dritz, S.S., Thomson, D.U., Renter, D.G., 2009. Feedlot health and performance effects associated with the timing of respiratory disease treatment1. J. Anim. Sci. 87, 314–327. https://doi.org/10.2527/jas.2008-1201.

Bareille, N., Seegers, H., Quillet, J.-M., Assié, S., 2009. Impact of respiratory disorders in young bulls during their fattening period on performance and profitability. Impact of Respiratory Disorders in Young Bulls during Their Fattening Period on Performance and Profitability. Presented at the In: 25. World Buiatrics Congress.

Beaumont, M.A., 2019. Approximate Bayesian Computation. Annu. Rev. Stat. Its Appl. 6, 379–403. https://doi.org/10.1146/annurev-statistics-030718-105212.

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw. 5, 157–166. https://doi.org/10.1109/72.279181.

Birnbaum, S., Kuleshov, V., Enam, Z., Koh, P.W.W., Ermon, S., 2019. Temporal FiLM: Capturing Long-Range Sequence Dependencies with Feature-wise Modulations.

Brault, S.A., Hannon, S.J., Gow, S.P., Warr, B.N., Withell, J., Song, J., Williams, C.M., Otto, S.J.G., Booker, C.W., Morley, P.S., 2019. Antimicrobial Use on 36 Beef Feedlots in Western Canada: 2008–2012. Front. Vet. Sci. 6, 329. https://doi.org/10.3389/fvets.2019.00329.

Buczinski, S., Forté, G., Francoz, D., Bélanger, A. -M., 2014. Comparison of Thoracic Auscultation, Clinical Score, and Ultrasonography as Indicators of Bovine Respiratory Disease in Preweaned Dairy Calves. J. Vet. Intern. Med. 28, 234–242. https://doi.org/10.1111/jvim.12251.

Cantor, M.C., Casella, E., Silvestri, S., Renaud, D.L., Costa, J.H.C., 2022. 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. Front. Anim. Sci. 3, 852359. https://doi.org/10.3389/fanim.2022.852359.
Cuevas-Gómez, I., McGee, M., McCabe, M., Cormican, P., O'Riordan, E., McDaneld, T., Earley, B., 2020. Growth performance and hematological changes of weaned beef calves diagnosed with respiratory disease using respiratory scoring and thoracic ultrasonography. J. Anim. Sci. 98, skaa345. https://doi.org/10.1093/jas/skaa345.
Delabouglise, A., James, A., Valarcher, J.-F., Hagglünd, S., Raboisson, D., Rushton, J., 2017. Linking disease epidemiology and livestock productivity: The case of bovine respiratory disease in France. PLOS ONE 12, e0189090. https://doi.org/10.1371/journal.pone.0189090.
Drovandi, C., Frazier, D.T., 2022. A comparison of likelihood-free methods with and without summary statistics. Stat. Comput. 32, 42. https://doi.org/10.1007/s11222-022-10092-4.
Dubrovsky, S.A., Van Eenennaam, A.L., Karle, B.M., Rossitto, P.V., Lehenbauer, T.W., Aly, S.S., 2019. Epidemiology of bovine respiratory disease (BRD) in preweaned calves on California dairies: the BRD 10K study. J. Dairy Sci. 102, 7306–7319. https://doi.org/10.3168/jds.2018-14774.
Edwards, T.A., 2010. Control methods for bovine respiratory disease for feedlot cattle. Vet. Clin. North Am. Food Anim. Pract. 26, 273–284. https://doi.org/10.1016/j.cvfa.2010.03.005.
Engler, M., Defoor, P., King, C., Gleghorn, J., 2014. The impact of bovine respiratory disease: the current feedlot experience. Anim. Health Res. Rev. 15, 126–129. https://doi.org/10.1017/S1466252314000139.
Ezanno, P., Andraud, M., Beaunée, G., Hoch, T., Krebs, S., Rault, A., Touzeau, S., Vergu, E., Widgren, S., 2020. How mechanistic modelling supports decision making for the control of enzootic infectious diseases. Epidemics 32, 100398. https://doi.org/10.1016/j.epidem.2020.100398.
Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning..
Gaudino, M., Nagamine, B., Ducatez, M.F., Meyer, G., 2022. Understanding the mechanisms of viral and bacterial coinfections in bovine respiratory disease: a comprehensive literature review of experimental evidence. Vet. Res. 53, 70. https://doi.org/10.1186/s13567-022-01086-1.
Gershwin, L.J., Van Eenennaam, A.L., Anderson, M.L., McEligot, H.A., Shao, M.X., Toaff-Rosenstein, R., Taylor, J.F., Neibergs, H.L., Womack, J., Bovine Respiratory Disease Complex Coordinated Agricultural Project Research Team, 2015. Single Pathogen Challenge with Agents of the Bovine Respiratory Disease Complex. PLOS ONE 10, e0142479. https://doi.org/10.1371/journal.pone.0142479.
Griffin, D., 2010. Bovine pasteurellosis and other bacterial infections of the respiratory tract. Vet. Clin. North Am. Food Anim. Pract. 26, 57–71. https://doi.org/10.1016/j.cvfa.2009.10.010.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90. https://doi.org/10.1145/3065386.
Lambert, S., Bauzile, B., Mugnier, A., Durand, B., Vergne, T., Paul, M.C., 2023. A systematic review of mechanistic models used to study avian influenza virus transmission and control. Vet. Res. 54, 96. https://doi.org/10.1186/s13567-023-01219-0.
Lee, W., Seo, K., 2022. Downsampling for binary classification with a highly imbalanced dataset using active learning. Big Data Res 28, 100314. https://doi.org/10.1016/j.bdr.2022.100314.
Masset, N., Assié, S., Herman, N., Jozan, T., Herry, V., 2022. Ultrasonography of the cranial part of the thorax is a quick and sensitive technique to detect lung consolidation in veal calves. Vet. Med. Sci. 8, 1229–1239. https://doi.org/10.1002/vms3.774.
Muñoz, F., Pleydell, D.R.J., Jori, F., 2022. A combination of probabilistic and mechanistic approaches for predicting the spread of African swine fever on Merry Island. Epidemics 40, 100596. https://doi.org/10.1016/j.epidem.2022.100596.
Murray, G.M., More, S.J., Sammin, D., Casey, M.J., McElroy, M.C., O'Neill, R.G., Byrne, W.J., Earley, B., Clegg, T.A., Ball, H., Bell, C.J., Cassidy, J.P., 2017. Pathogens, patterns of pneumonia, and epidemiologic risk factors associated with respiratory disease in recently weaned cattle in Ireland. J. Vet. Diagn. Invest. 29, 20–34. https://doi.org/10.1177/1040638716674757.
Nickell, J.S., White, B.J., 2010. Metaphylactic antimicrobial therapy for bovine respiratory disease in stocker and feedlot cattle. Vet. Clin. North Am. Food Anim. Pract. 26, 285–301. https://doi.org/10.1016/j.cvfa.2010.04.006.
Ollivett, T.L., Buczinski, S., 2016. On-farm use of ultrasonography for bovine respiratory disease. Vet. Clin. North Am. Food Anim. Pract. 32, 19–35. https://doi.org/10.1016/j.cvfa.2015.09.001.
Pardon, B., Hostens, M., Duchateau, L., Dewulf, J., De Bleecker, K., Deprez, P., 2013. Impact of respiratory disease, diarrhea, otitis and arthritis on mortality and carcass traits in white veal calves. BMC Vet. Res. 9, 79. https://doi.org/10.1186/1746-6148-9-79.
Picault, S., Ezanno, P., Smith, K., Amrine, D., White, B., Assié, S., 2022. Modelling the effects of antimicrobial metaphylaxis and pen size on bovine respiratory disease in high and low risk fattening cattle. Vet. Res. 53, 77. https://doi.org/10.1186/s13567-022-01094-1.
Picault, S., Huang, Y.-L., Sicard, V., Arnoux, S., Beaunée, G., Ezanno, P., 2019b. EMULSION: transparent and flexible multiscale stochastic models in human, animal and plant epidemiology. PLOS Comput. Biol. 15, e1007342. https://doi.org/10.1371/journal.pcbi.1007342.
Plank, M.J., Hendy, S.C., Binny, R.N., Vattiato, G., Lustig, A., Maclaren, O.J., 2022. Using mechanistic model-based inference to understand and project epidemic dynamics with time-varying contact and vaccination rates. Sci. Rep. 12, 20451. https://doi.org/10.1038/s41598-022-25018-3.
Ramezani Gardaloud, N., Guse, C., Lidauer, L., Steininger, A., Kickinger, F., Öhlschuster, M., Auer, W., Iwersen, M., Drillich, M., Klein-Jöbstl, D., 2022. Early detection of respiratory diseases in calves by use of an ear-attached accelerometer. Animals 12, 1093. https://doi.org/10.3390/ani12091093.
Schaefer, A.L., Cook, N.J., Church, J.S., Basarab, J., Perry, B., Miller, C., Tong, A.K.W., 2007. The use of infrared thermography as an early indicator of bovine respiratory disease complex in calves. Res. Vet. Sci. 83, 376–384. https://doi.org/10.1016/j.rvsc.2007.01.008.
Smith, R.A., Step, D.L., Woolums, A.R., 2020. Bovine respiratory disease. Vet. Clin. North Am. Food Anim. Pract. 36, 239–251. https://doi.org/10.1016/j.cvfa.2020.03.009.
Sorin-Dupont, B., Picault, S., Pardon, B., Ezanno, P., Assié, S., 2023. Modeling the effects of farming practices on bovine respiratory disease in a multi-batch cattle fattening farm. Prev. Vet. Med. 219, 106009. https://doi.org/10.1016/j.prevetmed.2023.106009.
Stutz, D., Roy, A.G., Matejovicova, T., Strachan, P., Cemgil, A.T., Doucet, A., 2023. Conform. Predict. Ambiguous Ground Truth. https://doi.org/10.48550/arXiv.2307.09302.
Timsit, E., Assié, S., Quiniou, R., Seegers, H., Bareille, N., 2011. Early detection of bovine respiratory disease in young bulls using reticulo-rumen temperature boluses. Vet. J. 190, 136–142. https://doi.org/10.1016/j.tvjl.2010.09.012.
Timsit, E., Dendukuri, N., Schiller, I., Buczinski, S., 2016. Diagnostic accuracy of clinical illness for bovine respiratory disease (BRD) diagnosis in beef cattle placed in feedlots: a systematic literature review and hierarchical Bayesian latent-class meta-analysis. Prev. Vet. Med. 135, 67–73. https://doi.org/10.1016/j.prevetmed.2016.11.006.
Timsit, E., Tison, N., Booker, C.W., Buczinski, S., 2019. Association of lung lesions measured by thoracic ultrasonography at first diagnosis of bronchopneumonia with relapse rate and growth performance in feedlot cattle. J. Vet. Intern. Med. 33, 1540–1546. https://doi.org/10.1111/jvim.15483.
Woolums, A.R., Berghaus, R.D., Smith, D.R., Daly, R.F., Stokka, G.L., White, B.J., Avra, T., Daniel, A.T., Jenerette, M., 2018. Case-control study to determine herd-level risk factors for bovine respiratory disease in nursing beef calves on cow-calf operations. J. Am. Vet. Med. Assoc. 252, 989–994. https://doi.org/10.2460/javma.252.8.989.
Zhu, H., Luo, M., Wang, R., Zheng, A., He, R., 2020. Deep Audio-Vis. Learn.: A Surv. https://doi.org/10.48550/arXiv.2001.04758.