Received: 16 November 2022 | Accepted: 3 February 2023

DOI: 10.1111/jvim.16664

STANDARD ARTICLE

Journal of Veterinary Internal Medicine Open Access

ACVIM American College of Veterinary Internal Medicine

A novel machine learning-based web application for field identification of infectious and inflammatory disorders of the central nervous system in cattle

Sara Ferrini[1] | Cesare Rollo[2] | Claudio Bellino[1] | Giuliano Borriello[1] | Giulia Cagnotti[1] | Cristiano Corona[3] | Giorgia Di Muro[1] | Mario Giacobini[1] | Barbara Iulini[3] | Antonio D'Angelo[1]

[1]Department of Veterinary Sciences, University of Turin, Turin, Italy

[2]Department of Medical Sciences, University of Turin, Turin, Italy

[3]Istituto Zooprofilattico del Piemonte Liguria e Valle d'Aosta, Turin, Italy

Correspondence
Giulia Cagnotti, Largo Paolo Braccini 2-10095, Grugliasco (TO), Italy.
Email: giulia.cagnotti@unito.it

Funding information
Ministero dell'Istruzione, dell'Università e della Ricerca

Abstract

Background: Central nervous system (CNS) infections in cattle are a major cause of economic loss and mortality. Machine learning (ML) techniques are gaining widespread application in solving predictive tasks in both human and veterinary medicine.

Objectives: Our primary aim was to develop and compare ML models that could predict the likelihood of a CNS disorder of infectious or inflammatory origin in neurologically-impaired cattle. Our secondary aim was to create a user-friendly web application based on the ML model for the diagnosis of infection and inflammation of the CNS.

Animals: Ninety-eight cattle with CNS infection and 86 with CNS disorders of other origin.

Methods: Retrospective observational study. Six different ML methods (logistic regression [LR]; support vector machine [SVM]; random forest [RF]; multilayer perceptron [MLP]; K-nearest neighbors [KNN]; gradient boosting [GB]) were compared for their ability to predict whether an infectious or inflammatory disease was present based on demographics, neurological examination findings, and cerebrospinal fluid (CSF) analysis.

Results: All 6 methods had high prediction accuracy (≥80%). The accuracy of the LR model was significantly higher (0.843 ± 0.005; receiver operating characteristic [ROC] curve 0.907 ± 0.005) than the other models and was selected for implementation in a web application.

Conclusion and Clinical Importance: Our findings support the use of ML algorithms as promising tools for veterinarians to improve diagnosis. The open-access web application may aid clinicians in achieving correct diagnosis of infectious and inflammatory

Abbreviations: CNS, central nervous system; CSF, cerebrospinal fluid; f_n, false negative; f_p, false positive; GB, gradient boosting; INF, infectious-inflammatory; IQR, interquartile range; KNN, K-nearest neighbors; LR, logistic regression; ML, machine learning; MLP, multilayer perceptron; *N*, total subjects; NL, neurolocalization; NON INF, noninfectious-inflammatory; RF, random forest; ROC-AUC, receiving operating characteristics area under curve; SVM, support vector machine; t_n, true negative; TNCC, total nucleated cell count; t_p, true positive.

neurological disorders in livestock, with the added benefit of promoting appropriate use of antimicrobials.

KEYWORDS
bovine neurology, central nervous system infections, clinical decision-making process, machine learning

1 | INTRODUCTION

Central nervous system (CNS) infections in cattle are a major cause of economic loss, mortality, and decreased productivity.[1,2] Certain neurological infections also may be zoonotic.[3] Achieving an etiological diagnosis may allow for accurate treatment and appropriate control and prevention measures. Doing so can be challenging however because clinical signs and hematological changes often are nonspecific. Cerebrospinal fluid (CSF) can be easily and safely collected in the field for diagnosis. It is the most direct antemortem method of diagnosing CNS disease, because advanced diagnostic imaging is much less feasible in large animals.[4]

Some nervous diseases are related to age, neuroanatomical localization often is associated with specific infectious disorders and CSF analysis for diagnosis of inflammation usually shows a moderate to marked increase in total nucleated cell count (TNCC) and protein concentration.[5-7] In the field, tentative diagnosis in a patient referred for neurological signs is arrived at by clinical reasoning and history taking for age, clinical findings including vital signs, clinical course, neuroanatomical localization, and CSF analysis. The cause of neurological signs rarely is confirmed however without necropsy examination.

On the other hand, techniques based on artificial intelligence are gaining widespread application in a large variety of predictive tasks both in human and veterinary medicine, as well as in medical imaging interpretation and the clinical decision-making process.[8-11] The ability to accurately predict the presence of an infectious or inflammatory disorder would be of considerable benefit for clinicians in selecting appropriate treatment.

Based on these premises, our primary objective was to develop and compare machine learning (ML) models that, based on demographic data and clinical and diagnostic findings, are able to predict the likelihood of a CNS disorder of infectious or inflammatory origin in neurologically-impaired cattle. Our secondary aim was to develop a user-friendly web application derived from the ML model that could be easily applied in clinical settings for the diagnosis of CNS disorders of infectious or inflammatory origin.

2 | MATERIALS AND METHODS

2.1 | Study population

The medical records of cattle presenting with signs suggestive of a CNS disorder to the neurology service of the Veterinary Teaching Hospital of Turin between July 2007 and March 2022 were reviewed. All animals underwent general and neurological examination by a board-certified specialist in neurology (ADA), CSF and blood analysis, and necropsy when possible.

The medical records were evaluated for data on age, sex, breed usage, neuroanatomical localization of the CNS disorder, CSF analysis (TNCC, microprotein concentration, final CSF interpretation), and final diagnosis expressed as the VITAMIN D mnemonic.[12] The final diagnosis was based on signalment, neurological examination, blood and CSF analysis, response to treatment, and necropsy histopathology when performed. Missing medical record data dictated exclusion from the study.

Two groups were formed according to the final diagnosis: patients with infectious or inflammatory disease of the CNS (INF group) and patients with a CNS disorder of other origin (eg, anomalous, vascular, neoplastic, degenerative, traumatic, or metabolic-toxic disorder; NON INF group).

This retrospective study was conducted in accordance with current animal welfare regulations (Directive 98/58/EC and Italian Decree Law 146/2001). Samples were collected during routine diagnostic evaluation. Written informed consent was obtained from the owners before veterinary assessment and treatment of their animals.

2.2 | Descriptive statistical analysis

Standard descriptive statistics were performed using commercially available software (Phyton version 3.8.8; Excel version 16.27). Continuous variables were tested for normal distribution using the Shapiro-Wilk test (alpha .05) and found to be not normally distributed. Standard descriptive statistics are reported as median and interquartile range (IQR) for continuous variables and percentage and frequency for categorical variables.

2.3 | Machine learning

Predicting a patient's classification in a diagnostic class (INF group or NON INF group) can be interpreted as a supervised binary problem. Within this framework, several ML models were trained and tested for their ability to detect the diagnostic class based on demographic and clinical and diagnostic data. A validation strategy then was applied to render the models generalizable for achieving accurate prediction in patients outside of the training set.

2.3.1 | Data preprocessing

Table 1 presents the variables and their measurements. Three were numerical (age, CSF TNCC, CSF microprotein concentration) and 4 were categorical (sex, breed usage, clinical neurolocalization, CSF interpretation). The numerical measurements were transformed by scaling each feature x to a range between 0 and 1 as follows:

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}.$$

Nonbinary categorical covariates were encoded using the one-hot encoding scheme (https://scikit-learn.org/stable/modules/preprocessing.html#preprocessing-categorical-features). The final preprocessed dataset included 20 features. Finally, because there were no missing values, no imputation of features was necessary.

2.3.2 | Validation strategy

The dataset was randomly divided into training and test sets in a proportion of 75% to 25%. The training set was used to train the models and perform 10-fold cross-validation. In general, K-fold cross-validation works by randomly splitting a dataset into K equally-sized subsets. The K-1 subsets are used for training the model, whereas the remaining subset is used as an internal test to measure a model's capabilities. This process is repeated until every subset has been employed in the validation phase. Finally, the K performances are averaged to obtain a unique cross-validation score. This procedure was applied to each possible set of hyperparameters to select those corresponding to the best cross-validation performance. Details on the hyperparameters for each model are given in File S1 and Table S1 in the Supporting Materials. Finally, the model presenting optimal hyperparameters was retrained on the entire training set and final performance was calculated on the blind test set.

We repeated the experiment 100 times with different random seeds of the training test split to prevent anomalies in dataset division and obtain a statistically robust evaluation. The average and SEM of the results are reported.

TABLE 1 Demographic, clinical, and diagnostic attributes of the dataset.

Attributes		Detail
Demographic data	Age	Months
	Sex	Male or female
	Breed usage	Beef, dairy, or both
Clinical findings	Neurolocalization	Forebrain, brainstem, central vestibular system, cerebellum, spinal, multifocal, diffuse intracranial
Laboratory findings (CSF)	TNCC	Number of cells/μL
	Microprotein concentration	mg of microprotein/dL
	Interpretation	Unremarkable, mononuclear pleocytosis, neutrophilic pleocytosis, mixed pleocytosis, lymphocytic pleocytosis, albuminocytologic dissociation

Abbreviations: CSF, cerebrospinal fluid; TNCC, total nucleated cell count.

2.3.3 | Classification algorithms

The ML algorithms for diagnosis prediction were logistic regression (LR), support vector machine (SVM), random forest (RF), multilayer perceptron (MLP), K-nearest neighbors (KNN), and gradient boosting (GB). Briefly, LR is a statistical model that estimates the probability of an event occurring (which in our case is the INF vs NON INF diagnosis), based on a given dataset of independent variables (the clinical features). Because the outcome is a probability, it is bounded between 0 and 1. Support vector machine predicts the outcome by identifying the curve that best separates samples belonging to the 2 classes in the data points space, where each sample is a data point for which the axes are the clinical features. The curve is required to be as distant as possible from data points of both classes. Random forest is an ensemble of decision trees (ie, tree-like structures in which each internal node represents a test on a feature, each branch represents the outcome of the test, and each leaf node represents a class label [INF vs NON INF]). Multilayer perceptron is a type of artificial neural network. This model can find complex relationships among clinical features by learning nonlinear functions to predict outcome. K-nearest neighbors algorithm, similar to SVM, maps the samples into their features' space. However, instead of finding the best separator of the 2 classes, it labels a sample on the basis of the class of the nearest samples. The GB algorithm combines simple models, called weak learners, into a single strong learner in a multistep fashion. The idea of GB is that, for each step, the weak learner learns to fix the errors of the previous learner, and this procedure is repeated for a certain number of stages.

A further detailed description of the ML algorithms is given in File S1 in the Supporting Materials. The analyses were performed using the Scikit-Learn package (version 1.0.2) in Python 3.8.

2.3.4 | Metrics and model comparison

The standard ML metrics to measure algorithm performance on the test set were: accuracy, recall (sensitivity), precision (positive predictive value), and F1 score. The receiver operating characteristics area under the curve (ROC-AUC) was calculated for each ML model.

Accuracy refers to the ratio between correctly identified cases and the total number of cases (Equation 1). For our study, accuracy is a measurement of the model's capability to determine whether the CNS disorder is of infectious or inflammatory origin or not by

assigning equal importance to both classes. Recall, also named sensitivity, is the ratio between correctly identified positive cases and the total number of positive cases (Equation 2). It measures the percentage of correctly identified infectious or inflammatory cases. Precision, or the positive predictive value, refers to the number of animals correctly predicted as having an infectious or inflammatory disorder out of the total number predicted (Equation 3). The F1 score (Equation 4) is the harmonic mean of precision and recall. It is particularly useful in imbalanced datasets (the number of cattle with infection or inflammation of the CNS is not comparable to the number of cattle with a non-infectious inflammatory disorder). The ROC curve is a probability curve that plots the true positive rate against the false positive rate at various classification thresholds. The ROC-AUC provides an aggregate measure of the ability of the classifier to distinguish between classes across all classification thresholds. Unlike other metrics, the ROC-AUC directly considers the probabilistic output of the predictor (between 0 and 1) and also quantifies how good the model is at ranking predictions. When the AUC is 1, the model can distinguish perfectly between positive and negative class points, whereas an AUC 0.5 means that the classifier is predicting either random class or constant class for all data points.

$$\text{Accuracy} = \frac{t_p + t_n}{N}, \tag{1}$$

$$\text{Recall (sensitivity)} = \frac{t_p}{t_p + f_n}, \tag{2}$$

$$\text{Precision (positive predictive value)} = \frac{t_p}{t_p + f_p}, \tag{3}$$

$$\text{F1 score} = \frac{2t_p}{2t_p + f_p + f_n}, \tag{4}$$

where t_p is true positive; t_n is true negative; f_n is false negative; f_p is false positive; N is total subjects. A major property of precision, recall, and F1 score is that, by definition, they are calculated only on the positive class (INF group), whereas accuracy and the ROC-AUC take into account positive and negative classes equally and usually are a better choice when the dataset is balanced and there is equal interest in predicting both classes correctly. For this reason, only accuracy was applied as an evaluation metric for selecting optimal hyperparameters.

Finally, the post hoc Friedman test and the Nemenyi test for multiple comparisons were performed to determine the most suitable ML classifier, in which we compared accuracy and ROC-AUC metrics of all 6 models. Statistical significance was set alpha .05.

2.3.5 | Web user interface

The model with the best average performance in the 100 repeated trials was implemented in the web application. Because each trial resulted in a different set of optimal hyperparameters, there were 100 versions of the selected model. To address this issue, we took the median value of each hyperparameter for implementation of the final model. This classifier then was retrained on the entire dataset to exploit the full capacity of the data.

The web application was built using the Streamlit Python-based framework (https://streamlit.io/). The application can operate only when all input variables are given. It returns the probability of a patient with infection or inflammation of the CNS. Probabilities >50% are indicative of a CNS disorder of infectious or inflammatory origin, whereas probabilities <50% are predictive of a CNS disorder not of infectious or inflammatory origin.

TABLE 2 Demographic, clinical, and diagnostic attributes of the INF and NON INF groups.

Clinical attributes	INF group (n = 98)	NON INF group (n = 86)
Age	4 (IQR, 0-12)	4 (IQR, 1.50-6.50)
Sex—no. (%)		
Female	57/98 (58)	42/86 (49)
Male	41/98 (42)	44/86 (51)
Breed usage—no. (%)		
Beef	75/98 (77)	67/86 (78)
Dairy	13/98 (13)	11/86 (13)
Both	10/98 (10)	8/86 (9)
Neurolocalization—no. (%)		
Forebrain	27/98 (28)	54/86 (63)
Multifocal	24/98 (24)	6/86 (7)
Brainstem	23/98 (24)	1/86 (1)
Focal spinal	3/98 (3)	19/86 (22)
Central vestibular system	16/98 (16)	0
Cerebellum	3/98 (3)	4/86 (5)
Diffuse intracranic disorder	2/98 (2)	2/86 (2)
CSF—TNCC (cells/μL)	33.2 (IQR, 12.20-107.4)	6.50 (IQR, 2.10-12.15)
CSF—microprotein (mg/dl)	61.00 (IQR, 38.85-198.31)	30.00 (IQR, 22.85-48.09)
CSF—interpretation—no (%)		
Mononuclear pleocytosis	51/98 (52)	25/86 (29)
Unremarkable	10/98 (10)	53/86 (62)
Neutrophilic pleocytosis	27/98 (28)	0
Mixed pleocytosis	5/98 (5)	2/86 (2)
Albuminocytologic dissociation	4/98 (4)	3/86 (3.5)
Lymphocytic pleocytosis	1/98 (1)	3/86 (3.5)

Abbreviations: CSF, cerebrospinal fluid; TNCC, total nucleated cell count.

3 | RESULTS

3.1 | Study sample characteristics

The study sample was 184 cattle (85/184 [46%] males and 99/184 [54%] females). The median age was 4 months (IQR, 1-9 months). Most were beef cattle (142/184, 78%), 24/184 (13%) were dairy cattle, and 18/184 (9%) belonged to breeds of both beef and dairy usage. Forebrain neurological localization was identified in 81/184 (44%), multifocal localization in 30/184 (16%), the brainstem was involved in 24/184 (13%), focal spinal localization in 22/184 (12%), the central vestibular system in 16/184 (9%), the cerebellum in 7/184 (4%), and diffuse intracranial disorder was identified in 4/184 (2%). The median TNCC was 12.5 cells/μL (IQR, 3.9-44.2) and the median microprotein concentration was 42.9 mg/dL (IQR, 27.5-95.9). The final CSF analysis showed mononuclear pleocytosis in 76/184 (41%), unremarkable results in 63/184 (34%), neutrophilic pleocytosis in 27/184 (15%), mixed pleocytosis in 7/184 (4%), albuminocytological dissociation in 7/184 (4%), and lymphocytic pleocytosis in 4/184 (2%).

An infectious or inflammatory disorder was diagnosed in 98/184 (53%) animals (INF group) and a CNS disorder of other origin in the remaining 86/184 (47%; NON INF group). Among the latter, a metabolic-toxic disorder was diagnosed in 50/86 (58%), an anomalous congenital condition in 14/86 (16%), trauma in 12/86 (14%), degenerative disease in 6/86 (7%), vascular disorder in 3/86 (4%), and CNS neoplasm in 1/86 (1%). Table 2 presents the characteristics of the 2 groups.

3.2 | Algorithms

We trained and evaluated the binary classifiers: LR, SVM, RF, MLP, KNN, and GB. The dataset was balanced between the 2 groups: 53% in the INF group and 47% in the NON INF group.

Table 3 presents the average evaluation metrics obtained from the 6 classification algorithms trained on the training set and evaluated on the test set. The LR classifier had the highest average accuracy and ROC-AUC (0.843 ± 0.005 and 0.907 ± 0.005, respectively), whereas the RF classifier returned the lowest average accuracy and ROC-AUC (0.802 ± 0.005 and 0.801 ± 0.005, respectively). The Friedman test and the post hoc Nemenyi comparisons were performed to compare the ROC-AUC and the accuracy metrics of the 6 models. The accuracy of the LR classifier was statistically superior to the other models ($P \leq .01$), except for the SVM ($P = .28$); the ROC-AUC of the LR classifier was statistically higher than that of all the other models ($P = .001$). Tables S2 and S3 in the Supplementary Materials present the P values for all comparisons, whereas Figures 1 and 2 present the results of the ROC-AUC and the accuracy for each model. Based on these results, the LR algorithm was implemented in a free-use web application (https://cnsprediction.streamlit.app/) in the Streamlit Python-based framework (Figure 3).

3.3 | Importance of clinical attributes

We identified the major factors predicting the diagnostic class and their contributions to prediction. Indeed, not all features contributed equally to the decision-making process in the classification models. We examined the LR coefficients to explain the relative contribution of each feature.

Figure 4 reports the LR coefficients attributed to the top 10 encoded features. Those associated with the INF group were neutrophilic pleocytosis in CSF and neurolocalization: brainstem, central vestibular system, and multifocal localization. Patient age was positively associated with this group as well.

The features associated with the NON INF group were unremarkable CSF analysis and neurolocalization: spinal, forebrain, diffuse intracranial and cerebellar localization. Other CSF results, TNCC, microprotein concentration, sex, and breed usage were retained as being less informative by the model for prediction.

4 | DISCUSSION

We applied advanced computational modeling to predict the diagnostic class in cattle with CNS disorders. The large dataset size and

TABLE 3 Performance of the classification algorithms in predicting diagnostic class for accuracy, precision, recall, F1 score, and ROC-AUC metrics computed on the test set.

Model	Accuracy ±SEM	Precision (sensitivity) ±SEM	Recall (positive predicted value) ±SEM	F1-score ±SEM	ROC-AUC ±SEM
LR	0.843 ± 0.005	0.904 ± 0.006	0.794 ± .007	0.843 ± 0.005	0.907 ± 0.004
SVM	0.832 ± 0.005	0.890 ± 0.006	0.788 ± 0.006	0.833 ± 0.005	0.835 ± 0.005
RF	0.802 ± 0.005	0.817 ± 0.006	0.818 ± 0.007	0.814 ± 0.005	0.801 ± 0.005
MLP	0.810 ± 0.006	0.847 ± 0.008	0.797 ± 0.007	0.818 ± 0.005	0.811 ± 0.006
KNN	0.823 ± 0.004	0.887 ± 0.006	0.772 ± 0.008	0.822 ± 0.005	0.827 ± 0.004
GB	0.805 ± 0.005	0.820 ± 0.006	0.821 ± 0.008	0.817 ± 0.005	0.804 ± 0.005

Note: The average of 100 trials and the SEM are reported.
Abbreviations: GB, gradient boosting; KNN, K-nearest neighbors; LR, logistic regression; MLP, multilayer perceptron; RF, random forest; SVM, support vector machine.

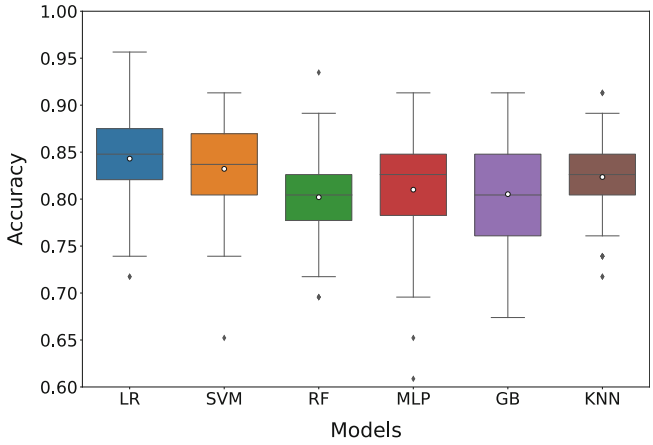

FIGURE 1 Average accuracies of the 6 machine learning methods (LR, logistic regression; SVM, support vector machine; RF, random forest; MLP, multilayer perceptron; KNN, K-nearest neighbors; GB, gradient boosting) run on the test set for predicting diagnostic class (INF or NON INF group).

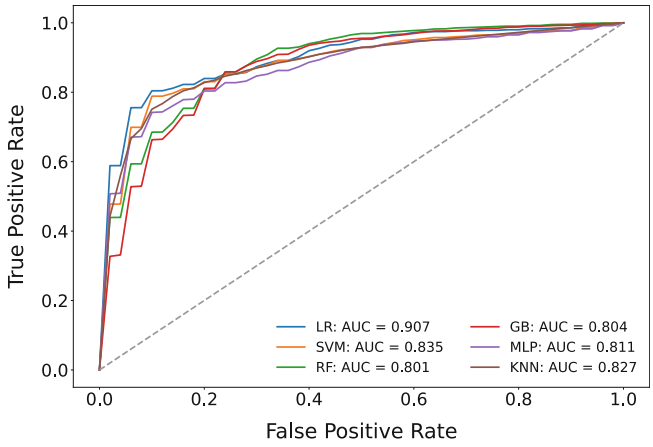

FIGURE 2 Receiver operating characteristic (ROC) curves of the 6 machine learning methods (LR, logistic regression; SVM, support vector machine; RF, random forest; MLP, multilayer perceptron; KNN, K-nearest neighbors; GB, gradient boosting) computed on the test set for predicting diagnostic class (INF or NON INF group).

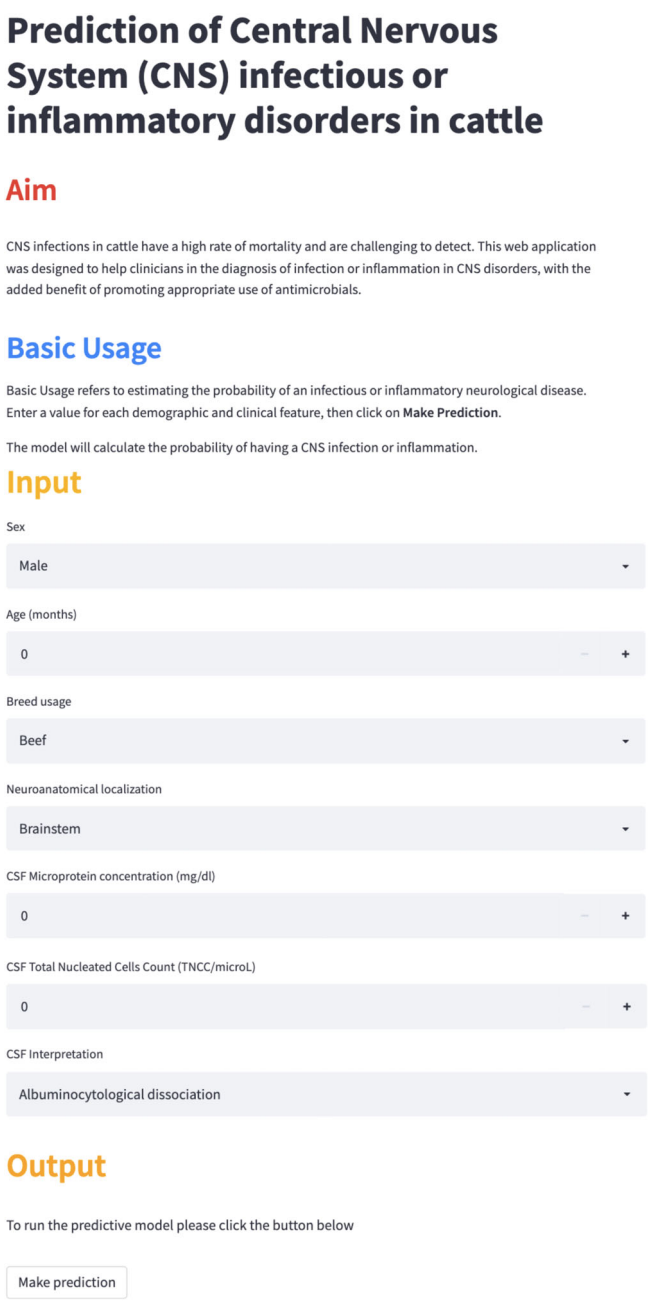

FIGURE 3 Web application Graphic Interface. From https://cnsprediction.streamlit.app/.

balance in diagnostic classes allowed us to perform robust predictive analyses using ML techniques. One of the advantages of using ML is that, by dividing the dataset into training and test groups, the learned model can be validated on unseen data, thus decreasing the risk of overfitting and improving generalizability as a result.

In our study, all 6 ML models performed well to predict diagnostic class (approximately 80%) according to the evaluation metrics. The average precision was higher than (LR, SVM, MLP, KNN) or equal to (RF and GB) the recall scores. Higher precision indicates that the classifier is better able to minimize false positives than false negatives.

Two points were evaluated for selecting which metrics to consider. The study dataset was balanced for the number of patients in each of the 2 diagnostic classes, and thus accuracy and ROC-AUC were robust and reliable metrics. Also, we were interested in accurate classification of both positive (INF group) and negative (NON INF group) cases. This information potentially decreases inappropriate use of antimicrobials, thus helping control the spread of antibiotic resistance without compromising animal health.

For these reasons, neutral metrics such as accuracy and ROC-AUC were chosen to determine the optimal model. Comparison of these 2 metrics showed that LR outperformed all of the others and therefore was selected for implementation in the web application. The reason why LR, which unlike the other models is linear, showed the best performances may be related to dataset size and the

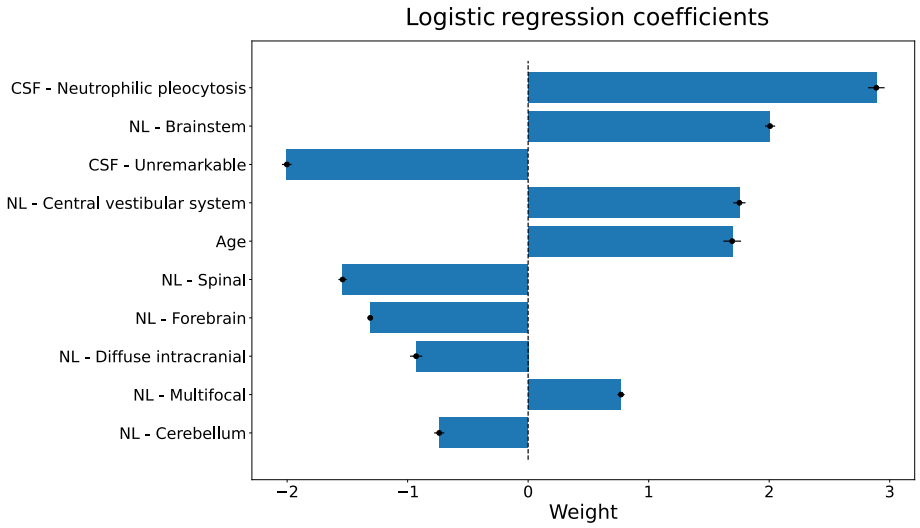

FIGURE 4 Plot of logistic regression coefficient ranking of prediction of diagnostic class. Coefficients with a positive weight were associated with the INF group, whereas coefficients with a negative weight were associated with the NON INF group. CSF, cerebrospinal fluid; NL, neurolocalization.

possibility that complex and nonlinear interaction between variables was not sufficiently intense. In fact, because all of other models were intrinsically built to learn nonlinear functions for outcome prediction, their performance could suffer if the available dataset is not large enough to encompass the potential complexity.

The variables used to train the model and their coefficients were consistent with the current literature and our clinical experience. The LR coefficient most strongly associated with the INF group was neutrophilic pleocytosis in the CSF. Indeed, detection of neutrophilic pleocytosis usually is interpreted as indicating a bacterial CNS infection in cattle.[7,13]

Brainstem, central vestibular and multifocal neurolocalization also were strongly associated with the INF group. In ruminants, the main differential diagnoses for brainstem disease are listeriosis and brainstem abscess. Similarly, central vestibular system involvement often occurs consequent to the spread of otogenic intracranial infection.[14] Multifocal localization is considered highly suspicious of infectious or inflammatory disease in both human and veterinary medicine.[2]

Older age was associated with increased risk of belonging to the INF group. Infectious and inflammatory diseases of the CNS can occur in animals of any age.[2] The NON INF group consisted mostly of animals with metabolic disease (60% of those in the NON INF group), which affects animals of any age.[15] However, our study population included several patients with congenital anomalies (16% of the NON INF group), which usually are referred in the first days of life, which partially could explain this result.

In contrast, the LR coefficient most often associated with the NON INF group was unremarkable CSF analysis, as noted in previous studies.[13] Spinal localization associated with the NON INF group may be explained by the fact that neurological disorders of the spine often are of traumatic origin,[16] whereas CNS infection or inflammation frequently is associated with intracranial signs.[2] In addition, neoplasia was reported to be the most common cause of spinal cord lesions in a recent study on recumbent dairy cows.[17]

Forebrain localization was associated with the NON INF group. It is a common in metabolic-toxic disorders in young and adult cattle,[15] which made up approximately 60% of diagnoses in the NON INF group. Hypocalcemia and hypomagnesemia have been reported as the most common causes of seizures in cattle,[18] whereas seizures caused by infection occurred in <9% of cases. Finally, the main differential diagnosis of diffuse intracranial and cerebellar localization, which was associated with the NON INF group, includes congenital and genetic anomalies and metabolic-toxic disorders.[19,20]

Variables that were less useful for the predictive model were TNCC and microprotein concentration. Although this finding was somewhat unexpected because CSF analysis for infectious and inflammatory conditions usually is characterized by a moderate to marked increase in both TNCC and protein concentration, most cases of infection in our study had only mild or moderate pleocytosis, which does not rule out other neurological disorders.[13,21]

Our study had some limitations. The predicted diagnostic class was either infectious inflammatory or noninfectious inflammatory because sample size did not allow for differentiation of the predictive output for all classes of the VITAMIN D mnemonic. A larger representation of each etiological class and a larger dataset would be necessary to train the model adequately to predict such classes with reasonable confidence. Similarly, our dataset size did not allow the models to subdifferentiate each infectious case from the others. Increasing sample size by involving other study centers with strict inclusion criteria to minimize bias could overcome this problem and allow for higher predictive power. Finally, we cannot exclude a geographical bias that could influence the prevalence of different disease pathogens because all cases came from a particular Italian area (Piedmont). The inclusion of other research centers in different geographical areas could help achieve more representative disease prevalence, potentially making our model more effective worldwide.

Overall, our findings and user-friendly web application may be a useful tool in the clinical decision-making process. Although the web application cannot replace the experience of a veterinarian, it can serve as a guide to diagnosis, with the added benefit of promoting more responsible use of antimicrobials.

ACKNOWLEDGMENT

This study was supported by Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) under the programme "Dipartimenti di Eccellenza ex L.232/2016" to the Department of Veterinary Science, University of Turin.

CONFLICT OF INTEREST DECLARATION

Authors declare no conflict of interest.

OFF-LABEL ANTIMICROBIAL DECLARATION

Authors declare no off-label use of antimicrobials.

INSTITUTIONAL ANIMAL CARE AND USE COMMITTEE (IACUC) OR OTHER APPROVAL DECLARATION

This retrospective study was conducted in accordance with current animal welfare regulations (Directive 98/58/EC and Italian Decree Law 146/2001). Samples were collected during the routine analysis required to perform the diagnostic procedure. Informed, written consent was obtained from the owners to authorize veterinary assessment and treatment of their animals.

HUMAN ETHICS APPROVAL DECLARATION

Authors declare human ethics approval was not needed for this study.

ORCID

Giulia Cagnotti https://orcid.org/0000-0003-1287-6723

REFERENCES

1. Konradt G, Bassuino DM, Prates KS, et al. Suppurative infectious diseases of the central nervous system in domestic ruminants. *Pesq Vet Bras*. 2017;37:820-828.
2. Fecteau G, George LW. Bacterial meningitis and encephalitis in ruminants. *Vet Clin North Am Food Anim Pract*. 2004;20(2):363-377.
3. Clarke LL, Hawkins IK, Rissi DR. Central nervous system diseases of cattle in Georgia, 2001-2017. *J Vet Diagn Invest*. 2019;31(4):588-593.
4. Nagy DW. Diagnostics and ancillary tests of neurologic dysfunction in the ruminant. *Vet Clin North Am Food Anim Pract*. 2017;33(1):9-18.
5. Dore V, Smith G. Cerebral disorders of calves. *Vet Clin North Am Food Anim Pract*. 2017;33(1):27-41.
6. Boileau MJ, Gilliam J. Brainstem and cranial nerve disorders of ruminants. *Vet Clin North Am Food Anim Pract*. 2017;33(1):67-99.
7. Stokol T, Divers TJ, Arrigan JW, McDonough SP. Cerebrospinal fluid findings in cattle with central nervous system disorders: a retrospective study of 102 cases (1990-2008). *Vet Clin Pathol*. 2009;38(1):103-112.
8. Fraiwan MA, Abutarbush SM. Using artificial intelligence to predict survivability likelihood and need for surgery in horses presented with acute abdomen (colic). *J Equine Vet Sci*. 2020;90:102973.
9. Biercher A, Meller S, Wendt J, et al. Using deep learning to detect spinal cord diseases on thoracolumbar magnetic resonance images of dogs. *Front Vet Sci*. 2021;8:721167.
10. Cihan P, Gökçe E, Kalipsiz O. A review of machine learning applications in veterinary field. *Kafkas Univ Vet Fak Der*. 2017;23:673-680.
11. Zhu G, Jiang B, Tong L, Xie Y, Zaharchuk G, Wintermark M. Applications of deep learning to neuro-imaging techniques. *Front Neurol*. 2019;10:869.
12. Jaggy A. Classification of neurological diseases: VITAMIN D. In: Jaggy A, ed. *Small Animal Neurology. An Illustrated Text*. Hannover: Schutersche; 2009:48-52.
13. Scott PR. Diagnostic techniques and clinicopathologic findings in ruminant neurologic disease. *Vet Clin North Am Food Anim Pract*. 2004;20(2):215-230.
14. Morin DE. Brainstem and cranial nerve abnormalities: listeriosis, otitis media/interna, and pituitary abscess syndrome. *Vet Clin North Am Food Anim Pract*. 2004;20:243-273.
15. Cebra CK, Cebra ML. Altered mentation caused by polioencephalomalacia, hypernatremia, and lead poisoning. *Vet Clin North Am Food Anim Pract*. 2004;20:287-302.
16. Divers TJ. Acquired spinal cord and peripheral nerve disease. *Vet Clin North Am Food Anim Pract*. 2004;20:231-242.
17. Achard D, Francoz D, Grimes C, et al. Cerebrospinal fluid analysis in recumbent adult dairy cows with or without spinal cord lesions. *J Vet Intern Med*. 2017;31(3):940-945.
18. D'Angelo A, Bellino C, Bertone I, et al. Seizure disorders in 43 cattle. *J Vet Intern Med*. 2015;29(3):967-971.
19. Scarratt WK. Cerebellar disease and disease characterized by dysmetria or tremors. *Vet Clin North Am Food Anim Pract*. 2004;20(2):275-286.
20. Washburn KE, Streeter RN. Congenital defects of the ruminant nervous system. *Vet Clin North Am Food Anim Pract*. 2004;20(2):413-434.
21. Tvedten HW. Clinical pathology of bovine neurologic disease. *Vet Clin North Am Food Anim Pract*. 1987;3(1):25-44.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ferrini S, Rollo C, Bellino C, et al. A novel machine learning-based web application for field identification of infectious and inflammatory disorders of the central nervous system in cattle. *J Vet Intern Med*. 2023;37(2):766-773. doi:10.1111/jvim.16664