Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning-Based Prediction of Brucella Seropositivity and Associated Factors among Rwandan Dairy Cattle with Reproductive Disorders

Sarah Mwera
Rwanda Agriculture and Animal Resources development Board
Maximillian Manzi
Rwanda Agriculture and Animal Resources development Board
Jacques Munyemana
Rwanda Agriculture and Animal Resources development Board
Fabrice Ndayisenga
Rwanda Agriculture and Animal Resources development Board
Methode Gasana
Rwanda Agriculture and Animal Resources development Board
Pascal Nyabinwa
Rwanda Agriculture and Animal Resources development Board
Mupenzi Mutimura
Rwanda Agriculture and Animal Resources Development Board
Eugene Niyonzima
eugeneniyonzima@gmail.com

University of Rwanda https://orcid.org/0000-0001-9347-830X

Research Article

Keywords: Brucella abortus, Risk factors, Machine learning, Random Forest, Disease surveillance, Rwanda

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11495015/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Brucellosis poses significant economic and zoonotic challenges to Rwanda's dairy sector. This study assessed factors associated with *Brucella* seropositivity among high-risk dairy cattle and benchmarked machine learning (ML) algorithms for predictive surveillance. A cross-sectional survey was conducted from June to December 2025 across four Rwandan districts (Nyagatare, Kayonza, Bugesera, and Rubavu). Serum samples and survey data were collected from 274 purposively selected dairy cattle with documented reproductive disorders from 75 farms. Serum samples were screened using the Rose Bengal test (RBT), and RBT-positive samples were subsequently tested using an indirect ELISA for confirmation. Multivariable logistic regression was used to assess factors associated with seropositivity ($p < 0.05$), and five ML models (Random Forest, Logistic Regression, KNN, Decision Tree, SVM) were evaluated using 5 metrics: Precision, recall, F1 score, accuracy, and Area under the Curve (AUC). Overall seropositivity rate among cattle with reproductive disorders was 10.2% (28/274), with spatial clustering in Nyagatare (19.8%) and Kayonza (7.3%). Multivariate analysis identified district, farm position, farmer education, disease awareness, and vaccination as key independent predictors. Among ML algorithms, Random Forest performed best, achieving 89.09% accuracy, 89.74% precision, 94.59% recall, 92.11% F1-score, and an AUC of 86.19%. Non-linear models consistently ranked district, education, and farm position as top predictive features. Among the selected dairy cattle with reproductive disorders, *Brucella* seropositivity was associated socio-demographic factors, whereas vaccination and disease awareness exert strong protective effects. Random Forest demonstrated promising predictive performance for identifying seropositive cattle within the study population. The findings from this study provide field-based evidence to inform risk-based brucellosis surveillance in Rwanda and other countries with comparable epidemiological and dairy production settings.

1. Introduction

Brucellosis is a major reproductive disease prevalent globally that poses serious public health concerns and causes considerable economic damage due to abortion, decreased milk yield, and infertility in dairy cattle. As a major zoonotic bacterial disease, it accounts for over 500,000 new human cases reported annually worldwide (Pal et al., 2017).

In sub-Saharan Africa, and particularly within East African Community countries, brucellosis remains endemic. Cattle-level seroprevalence was reported to widely vary, ranging from 0.2% to 43.8%, reflecting a heterogeneous distribution heavily influenced by local husbandry systems and awareness levels (Djangwani, Abong', et al., 2021; Edao et al., 2020).

In Rwanda, *Brucella* antibodies have been identified in farm bulk milk samples at rates up to 19.7%, with significantly higher prevalence observed in open grazing systems (37.9%). This higher rate is closely linked to risk factors such as a history of abortion and retained placenta, highlighting a severe zoonotic risk associated with the human consumption of unpasteurized milk (Djangwani, Abong', et al., 2021). Another study at the wildlife-livestock-human interface in Rwanda reported a 7.4% bovine

seroprevalence, identifying advanced animal age, herd composition, and a history of abortion as significant risk factors while emphasizing that farmer education is vital to reducing disease occurrence (Ntivuguruzwa et al., 2020). The public health dimension of Brucellosis is further underscored by a documented 25% seropositivity rate among Rwandan women presenting with abortion or stillbirth, which was strongly linked to direct animal contact and the consumption of raw milk (Rujeni & Mbanzamihigo, 2014).

Given the importance of dairy farming for livelihoods and nutrition in Rwanda, understanding the factors associated with *Brucella* seropositivity among dairy cattle with reproductive disorders is important for informing targeted surveillance and control interventions, reducing the risk of zoonotic transmission, and improving cattle reproductive health.

To date, research on brucellosis in Rwanda has relied entirely on classical statistical models to assess prevalence and associated factors. Despite the endemic nature of bovine brucellosis in East Africa, predictive machine learning (ML) frameworks have not yet been deployed to evaluate infection risks or capture complex factor interactions in the region. This study addresses this knowledge gap by benchmarking supervised ML algorithms to predict Brucellosis seropositivity and characterize critical farm and animal-level associated factors in Rwandan dairy cattle. The resulting data-driven framework offers actionable insights to optimize targeted surveillance, reduce livestock productivity losses, and mitigate zoonotic risks.

The findings from this study will be instrumental in bridging advanced data science with veterinary public health in Rwanda. By providing high-accuracy predictive insights, these results will help design targeted, cost-effective control strategies and early-warning systems. Ultimately, this framework will optimize veterinary resources, guide farmer education, and safeguard both public health and the economic vitality of the dairy sector.

2. Materials and Methods

2.1. Study design

A cross-sectional study was conducted from June to December 2025 to determine the seropositivity rate and identify the factors associated with *Brucella* infection in Rwandan dairy cattle. The study targeted dairy herds with a history of reproductive disorders, particularly abortions, across four districts: Kayonza, Nyagatare, Bugesera in the Eastern Province, and Rubavu in the Western Province. These specific districts were purposively selected due to their intensive dairy production activities and high frequency of reported reproductive disorders (Ndazigaruye et al., 2018; Smith et al., 2021).

2.2. Study population

A total of 274 dairy cattle were sampled across 75 dairy farms in the study areas. The study population comprised dairy cattle from purposively selected herds with a documented history of reproductive

disorders. Eligible animals were defined as cows with recorded abortion events, identified at the herd level through farmer reports and veterinary records. This targeted sampling strategy focused on a population at a higher risk of exposure to *Brucella* spp., thereby enhancing the likelihood of detecting active infection and characterizing associated factors.

2.3. Data collection

In order to investigate epidemiological factors associated with *Brucella* infection in Rwandan dairy cattle, a structured questionnaire was developed in English, translated into Kinyarwanda, and digitized using KoboToolbox. Trained enumerators conducted face-to-face interviews with cow owners or herd managers during farm visits, collecting herd- and animal-level data, including reproductive history, abortion, and breeding practices (natural mating or artificial insemination).

Based on reproductive histories, 274 eligible cattle were selected for blood sampling by trained veterinarians. Blood was collected from the jugular vein into sterile, plain Vacutainer tubes. Samples were kept upright at ambient temperature for approximately 30 minutes to allow clot formation, transported in insulated cool boxes with ice packs to the National Veterinary Laboratory within two hours, and centrifuged at 14,000 rpm for 3 minutes. The separated serum was aliquoted into labelled cryovials and stored at − 20°C until serological analysis.

2.4. Serological analyses

Serological screening for brucellosis was initially performed using the Rose Bengal Test (RBT). All samples testing positive by the RBT were subsequently subjected to confirmatory testing via an enzyme-linked immunosorbent assay (ELISA). Specifically, the ID Screen® Brucellosis Serum Indirect Multi-species ELISA kit (IDvet, Grabels, France; BRUS-MS), which employs a purified *Brucella* lipopolysaccharide (LPS) antigen, was utilized to detect antibodies against *Brucella abortus*. The assays were conducted strictly in accordance with the manufacturer's protocols and the international standards established by the World Organization for Animal Health.

2.5. Statistical analysis

Data were cleaned in Microsoft Excel before analysis. All statistical analyses were performed using Stata and Python software. Univariate analyses were used to summarize socio-demographic, animal, and herd-level characteristics and to calculate the seropositivity rate of *Brucella* infection. Bivariate analyses assessed associations between potential factors and seropositivity. Variables were entered into a logistic regression model to identify independent risk factors, with results reported as odds ratios and 95% confidence intervals. Statistical significance was set at p-value ≤ 0.05.

Categorical variables were translated into numeric codes allowing statistical software (Stata and Python) to handle and analyze them. In this study, the outcome variable was defined by the recorded Rose Bengal and ELISA Tests for each cow. Class imbalance was assessed by examining the distribution of positive and negative outcomes. Stratified sampling was used to preserve class proportions in the training and testing datasets.

2.6. Machine Learning models

Five supervised machine learning classification techniques were deployed to predict brucellosis occurrence, namely Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). These models were implemented using the scikit-learn library in Python. These algorithms were selected due to the binary nature of the outcome variable of this study, Brucellosis status, which was categorized as "positive" for a positive case of Brucellosis and "negative" for a negative case of Brucellosis.

The dataset was partitioned into a training subset (80%) and an independent testing subset (20%) utilizing a stratified random sampling approach to preserve class proportions across both splits. The baseline dataset was determined to be adequately balanced across the target classes, eliminating the necessity for synthetic data resampling methods.

2.6.1. Logistic regression

Logistic regression analysis was applied to estimate model coefficients, and odds ratios were used to determine the factors associated with the prevalence of Brucellosis. The dependent variable was binary, coded as 1 for a positive Brucellosis case and 0 for a negative case. The model was expressed as $logit\,(p) = \ln(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_9 X_9$. Where logit (p) is the log of the odds $\frac{p}{1-p}$. This can also be expressed in terms of probability p, and $\beta 0 + \beta 1 + \beta 2 + \beta 3 + \cdots + \beta 9$ are the estimated coefficients for the predictors $X_1$to X_9. The predictors included in the model were district (X_1), education level of the farmer (X_2), age of the farmer (X_3), position at the farm (X_4), type of reared species (X_5), source of the animals (X_6), vaccination status against Brucellosis (X_7), prior testing for Brucellosis (X_8), and awareness of Brucellosis (X_9).

2.6.2. Decision Tree

A Decision Tree is a predictive modeling tool used in machine learning to relate input features of an observation to a target outcome. It operates by repeatedly splitting a dataset into smaller subsets based on feature values, creating a tree-like structure of decision rules. The Decision Tree algorithm is a data mining technique that partitions data into increasingly homogeneous groups with respect to the target variable. In this study, the prediction model for Brucellosis disease was created using a decision tree classifier.

2.6.3. Random Forest

Random Forest is a machine learning method that constructs multiple decision trees using training datasets to create a robust classification model. Instead of relying on a single tree, it makes predictions based on the majority vote from all trees, which improves accuracy, especially when handling large datasets. This algorithm integrates two key techniques, bagging and random feature selection, to build a stronger ensemble model. By using many trees, Random Forest reduces the risk of overfitting and

enhances generalization (Thanh Noi & Kappas, 2017). The study used this classifier to build a predictive random forest model for Brucellosis disease.

2.6.4. Support Vector Machine

The Support Vector Machine classifier finds the optimal hyperplane that separates two classes (typically labeled as + 1 and − 1) while maximizing the margin between the closest points from each class. The equation of the Support Vector Machine decision boundary was $w^T x + b = 0$ where w is weight vector, x is input feature vector b is bias term. A new data point is classified based on $f(x) = \mathrm{sign}(w^T x + b)$, If $f(x) > 0 \rightarrow$ Class + 1 or if $f(x) < 0 \rightarrow$ Class − 1.

2.6.5. K-Nearest Neighbors

The K-Nearest Neighbors classifier operates by assigning a class label to a new observation based on the labels of its nearest neighbors in the training dataset S. For any given point x, the classifier considers the set of k closest observations, denoted as $(y_{\pi\, i}(x) : 1 \leq k)$, where $\pi_1(x \ldots \ldots . x_m(x))$, where $\pi_1(x \ldots \ldots . x_m(x))$ represents the reordered training set ranked according to their distance from x. In this case, distance is measured using the Euclidean distance metric. The classifier computes the Euclidean distance between the new data point and each observation in the training set, then selects the $k$nearest neighbors. The class label assigned to the new data point was determined by majority voting among these $k$neighbors. Thus, if most of the K-Nearest Neighbors correspond to positive Brucellosis cases, the new observation is classified as a Brucellosis case; otherwise, it is classified as a negative case.

2.6.6. Evaluation Metrics

Model performance was evaluated using several metrics, including accuracy, precision, recall, F1-score, and the area under the curve (AUC). Together, these indicators offer a comprehensive evaluation of each model's ability to correctly classify both positive and negative cases of Brucellosis. To identify the best-performing model, a ranking approach was applied across all metrics. For each criterion, models were ordered based on their performance, with the highest score receiving rank 1. The individual ranks were then summed to produce an overall rank score, where a lower total indicates superior performance.

3. Results

3.1. Socio-Demographic and Herd Characteristics

The socio-demographic and operational characteristics of the studied farms are presented in Table 1. Most respondents were male (89.33%), aged 36–64 years (56.00%), and farm owners (62.67%). Primary education was the most common educational level (30.67%), while 22.67% had no formal education. Most cattle were female (94.89%) and crossbred (87.23%). Cattle-only production was reported by 65.33% of respondents. Nyagatare accounted for the largest proportion of surveyed farms (33.33%), followed by Rubavu (25.33%), Kayonza (22.67%), and Bugesera (18.67%).

Table 1
Socio-demographic characteristics of respondents and operational profiles of dairy cattle farms. Values represent absolute frequencies (n) and relative percentages (%) of respondents within each variable category.

Variables	Category	Frequency (n)	Percentage (%)
Farmer's age category (n = 75)	<=35	23	30.67
	36–64	42	56.00
	> 64	10	13.33
Farmer's gender (n = 75)	Female	8	10.67
	Male	67	89.33
Farmer's education level (n = 75)	University	17	22.67
	Secondary	18	24.00
	Primary	23	30.67
	Not educated	17	22.67
Respondent's position at the farm (n = 75)	Animal's keeper	12	16.00
	Farm Manager	16	21.33
	Farm Owner	47	62.67
Cattle breed category (n = 274)	Cross	239	87.23
	Cross Friesian	26	9.49
	Local	9	3.28
Cattle Sex (n = 274)	Female	260	94.89
	Male	14	5.11
Livestock species raised at the farm (n = 75)	Only cattle	49	65.33
	Cattle & Small ruminants	20	26.67
	Cattle & others livestock	6	8.00
Location of the studied farm (n = 75)	Bugesera	14	18.67
	Kayonza	17	22.67
	Nyagatare	25	33.33

Variables	Category	Frequency (n)	Percentage (%)
	Rubavu	19	25.33

3.2. Seropositivity and Spatial Distribution of *Brucella* infection

The distribution of *Brucella* seropositivity across the investigated study districts is presented in Table 2. Out of the 274 total collected samples, 28 tested positive for brucellosis, yielding an overall seropositivity rate of 10.22%. Spatial analysis revealed that positivity was highly clustered geographically. The highest district-level seropositivity rate was recorded in Nyagatare (19.83%), followed by Kayonza (7.27%). In contrast, no positive cases were detected among the samples collected from Bugesera, and Rubavu districts.

Table 2
Spatial distribution and seropositivity rates of *Brucella* infection in dairy cattle across the study districts. Values represent total number of studied farms (n), number of samples tested (n), number of seropositive cases (n), and district-level seropositivity rates (%).

District	Number of farms	Number of tested samples	Number of positive cases	Positivity rate (%)
Bugesera	14	29	0	0.00
Kayonza	17	55	4	7.27
Nyagatare	25	121	24	19.83
Rubavu	19	69	0	0.0
Total	75	274	28	10.22

3.3. Bivariable Analysis of Factors Associated with *Brucella* Seropositivity

The bivariable logistic regression analysis identified significant associations between *Brucella* seropositivity and farmer education ($p = 0.004$), age ($p = 0.042$), farm position ($p = 0.014$), animal source ($p < 0.001$), livestock species composition ($p = 0.009$), previous testing ($p = 0.011$), vaccination status ($p = 0.006$), district ($p < 0.001$), and disease awareness ($p = 0.038$). Most positive cases (85.71%) were recorded in Nyagatare, while all positive cases occurred among unvaccinated cattle. Nevertheless, no significant associations were observed for breed, sex, herd mixing, recent animal introductions, or breeding method. The detailed distribution frequencies for all variables are presented in Table 3.

Table 3
Bivariable logistic regression analysis of socio-demographic and farm-level potential predictors associated with *Brucella* seropositivity. Values represent number (percentage) of *Brucella* negative and *Brucella* positive cases within each variable category.

Variables	Category	*Brucella* Negative	*Brucella* positive	P-value
Education level of farmer				
	University	61(24.80)	3(10.71)	0.004
	Secondary	73(29.67)	17(60.71)	
	Primary	62(25.20)	7(25)	
	Not educated	50(20.33)	1(3.57)	
Age of the farmer				
	<=35	56(22.76)	1(3.57)	0.042
	36–64	153(62.2)	20(71.43)	
	> 64	37(15.04)	7(25)	
Position at farm				
	Animal's keeper	36(14.63)	0	0.014
	Farm Manager	46(18.70)	2(7.14)	
	Farm Owner	164(66.67)	26(92.86)	
Cattle breed				
	Cross	212(86.18)	27(96.43)	0.195
	Cross Friesian	26(10.57)	0	
	Local	8(3.25)	1(3.57)	
Animal Sex				
	Female	233(94.72)	27(96.43)	0.696
	Male	13(5.28)	1(3.57)	
Mix with other herds				
	Yes	60(24.39)	8(28.57)	0.627
	No	186(75.61)	20(71.43)	
Source of reared animals				
	Girinka	5(2.02)	1(3.57)	0.000
	Inherited	113(45.75)	25(89.29)	

Variables	Category	*Brucella* Negative	*Brucella* positive	P-value
	Purchased	128(51.82)	2(7.14)	
Reared livestock species				
	Cattle	147(59.76)	25(89.29)	0.009
	Cattle & others	11(4.47)	0	
	Cattle & small ruminants	88(35.77)	3(10.71)	
Introduction of new animal in last 12 months				
	Yes	134(54.47)	15(53.57)	0.928
	No	112(45.53)	13(46.43)	
Test for Brucellosis before				
	Yes	61(24.80)	1(3.57)	0.011
	No	185(75.20)	27.96.43)	
Vaccinate against Brucellosis				
	Yes	53(21.54)	0	0.006
	No	193 (78.46)	28(100.00)	
Breeding Method				
	AI & Own bull	26(10.57)	1(3.57)	0.148
	Artificial Insemination	32(12.60)	1(3.570	
	Own Bull	189(76.83)	26(92.86)	
District				
	Bugesera	29 (11.79)	0	0.000
	Kayonza	51(20.73)	4(14.29)	
	Nyagatare	97(39.43)	24(85.71)	
	Rubavu	69(28.05)	0	
Aware of Brucellosis				
	Yes	213(86.59)	20(71.43)	0.038
	No	33(13.41)	8(28.57)	

3.4. Multivariable Analysis of Factors Associated with *Brucella* Seropositivity

The multivariable logistic regression identified significant associations between *Brucella* seropositivity and district, farm role, education level, vaccination status, and disease awareness (Table 4). Compared with animal keepers, farm managers ($p = 0.017$) and farm owners ($p = 0.004$) had significantly higher odds of seropositivity. Primary education ($p = 0.031$) and university education ($p = 0.040$) were associated with lower odds compared with secondary education. Vaccination status ($p = 0.017$) and disease awareness ($p = 0.043$) were also significantly associated with seropositivity. Farmer age, livestock species composition, animal source, and previous testing history were not significantly associated with seropositivity.

Table 4

Multivariable logistic regression analysis of factors associated with *Brucella* seropositivity in the studied dairy cattle. Values represent Odds Ratios (OR), Standard Errors (SE), z-statistics, p-values, and 95% Confidence Intervals (CI) for each candidate predictor relative to its specified reference category.

Predictors	**Odds Ratio (OR)**	**Standard error(SE)**	**z**	**p>|z|**	**[95% CI]**	
District (Bugesera*)						
Kayonza	2.039	1.413	2.06	0.039	1.035	4.017
Nyagatare	1.755	1.284	2.25	0.025	1.075	2.866
Rubavu	1					
Age of farmer (36–64*)						
<=35	0.842	0.56	-0.26	0.796	0.229	3.099
> 64	0.416	0.315	-1.16	0.246	0.094	1.833
Education level of farmer (Secondary*)						
Not Educated	0.000	0.000	-0.01	0.994	0.000	0.000
Primary	0.089	0.100	-2.16	0.031	0.010	0.799
University	0.043	0.066	-2.06	0.04	0.002	0.861
Position at farm (Animal's keeper*)						
Farm Manager	2.033	1.346	2.39	0.017	1.136	3.640
Farm Owner	3.078	1.472	2.91	0.004	1.443	6.563
Reared livestock species (Cattle only*)						
Cattle & small ruminants	0.248	0.219	-1.58	0.114	0.044	1.395
Cattle & others	1					
Source of reared animals (Girinka*)						
Inherited	0.161	0.245	-1.2	0.23	0.008	3.17
Purchased	1	.	.	.	.	.
Vaccinated against Brucellosis (No*)						
Yes	0.117	0.104	-2.39	0.017	0.020	0.676
Test for Brucellosis before (No*)						
Yes	1.393	0.97	0.48	0.634	0.356	5.453
* Reference category						

Predictors	Odds Ratio (OR)	Standard error(SE)	z	p>|z|	[95% CI]	
Aware of Brucellosis (No*)						
Yes	0.043	0.067	-2.03	0.043	0.002	0.903
constant	0.6197	1.580	-0.19	0.851	0.004	91.9126
* Reference category						

3.5. Performance Evaluation of Machine Learning Models

The comparative predictive performance and ranking of the five evaluated machine learning models are summarized in Table 5. Random Forest (RF) achieved the best overall performance (rank sum = 6), with the highest accuracy (89.09%), precision (89.74%), F1-score (92.11%), and AUC (86.19%). Logistic regression ranked second (rank sum = 9) and achieved the highest recall (97.30%). K-nearest neighbours and decision tree showed intermediate performance, whereas support vector machine had the lowest overall ranking (rank sum = 24). Overall, RF demonstrated the strongest predictive performance among the five evaluated models.

Table 5
Comparative evaluation metrics and performance ranking of machine learning models for predicting *Brucella* seropositivity. Values represent classification performance scores in percentage (individual metric rankings). Rank sum reflects the cumulative rank across all five metrics, establishing overall model priority (1 = best performing).

Machine Learning Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)	Rank sum
Logistic Regression	83.64 (2)	81.82 (2)	97.30 (1)	88.89 (2)	76.43 (2)	9 (2)
Random Forest	89.09 (1)	89.74 (1)	94.59 (2)	92.11(1)	86.19 (1)	6 (1)
K-Nearest Neighbors	81.82 (3)	81.40 (3)	94.59 (2)	87.50 (3)	75.08 (3)	14 (3)
Decision Tree	80.00(4)	80.95 (4)	91.89 (3)	86.08 (4)	73.72 (4)	19 (4)
Support Vector Machine	78.18 (5)	80.49 (5)	89.19 (4)	84.62 (5)	72.37 (5)	24 (5)

3.6. Feature Importance and Selection

As shown in Fig. 1, district, farmer education, farm role, and brucellosis awareness were among the most influential predictors across the machine learning models. Logistic regression ranked vaccination status as the most influential factor, while other models emphasized district and socio-demographic characteristics. The consistent importance of district suggests geographic variation in *Brucella* seropositivity among the sampled cattle.

4. Discussion

Brucellosis constitutes an important zoonotic disease affecting livestock productivity and public health. In this study, 10.2% of the sampled dairy cattle with reproductive disorders were classified as seropositive following the RBT screening and indirect ELISA confirmation. This finding should be interpreted in the context of the high-risk population investigated and not as an estimate of seroprevalence in the general Rwandan dairy cattle population.

Within the studied population, geographical variation was evident, with seropositive cattle identified only in Nyagatare and Kayonza. This pattern is consistent with previous studies in Rwanda. Kiiza et al. (2021) reported an animal-level seroprevalence of 14% among 465 dairy cattle sampled across three agroecological zones, with a higher proportion in Nyagatare (22%) than in Muhanga (5%) and Nyabihu (6%). Similarly, Ntivuguruzwa et al. (2025) reported *Brucella* seropositivity in 11.9% of cattle sampled in Nyagatare District, with significant geographical variation between sectors. These findings support the importance of location-specific surveillance. However, differences in sampling strategies, study populations, and diagnostic approaches limit direct comparisons with the present study. The absence of positive animals in Bugesera and Rubavu should therefore not be interpreted as evidence that brucellosis is absent from these districts.

Farmer education was significantly associated with seropositivity, with lower odds observed among farmers with primary or university education than among those with secondary education. Although various studies have reported an association of farmer's higher education attainment with lower *Brucella* seropositivity in the reared cattle (Assenga et al., 2015; Ntivuguruzwa et al., 2020), this pattern does not support a simple relationship between increasing educational attainment and decreasing seropositivity. Differences in practical knowledge (Ting et al., 2025), access to veterinary information (Asakura et al., 2018), and adoption of preventive practices (Izadi et al., 2024) may contribute to the observed association, although these mechanisms were not directly assessed. The study by Ndazigaruye et al. (2018) in Nyagatare, reported that 57.5% of surveyed dairy farmers were unaware of brucellosis, while 85.8% did not screen newly introduced animals and 82.5% did not remove seropositive animals from their herds. More recently, (Ntivuguruzwa et al., 2025) identified lack of awareness as a factor associated with livestock seropositivity. Together, these findings highlight the need for farmer education and accessible veterinary extension services, alongside formal education.

Farm owners and managers had higher odds of seropositivity than animal keepers. This association may reflect differences in responsibilities related to animal purchasing, breeding decisions, herd movement, and disease management. Evidence from Rwanda indicates that herd management practices can influence brucellosis occurrence. The study by Ntivuguruzwa et al. (2025) reveiled that retaining aborted animals in herds was associated with substantially higher odds of seropositivity (OR = 10.0), while improper disposal of aborted fetuses was also associated with increased odds (OR = 3.15). Although these findings do not directly explain the association with farm role observed here, they demonstrate the

importance of management decisions in brucellosis prevention and support engaging both farm decision-makers and animal handlers in biosecurity interventions.

Vaccination status was significantly associated with *Brucella* seropositivity in the present study, with lower odds of seropositivity observed among cattle reported to have been vaccinated. Vaccination is an important component of brucellosis control, although its effectiveness depends on factors such as vaccine selection and implementation (Heidary et al., 2022). This association should be interpreted cautiously because the used cross-sectional design does not establish causality, and information on vaccine type, timing, and coverage was unavailable. Furthermore, vaccination can influence antibody responses and complicate serological interpretation, depending on the vaccine strain and diagnostic tests used (Ducrotoy et al., 2018; Heidary et al., 2022). Therefore, the observed association should not be interpreted as direct evidence of vaccine effectiveness. Together with the association between farmer awareness and seropositivity, these findings highlight the importance of an integrated brucellosis control approach combining appropriate vaccination strategies, farmer education, veterinary supervision, and improved farm biosecurity.

Random Forest achieved the best overall predictive performance among the five evaluated machine-learning models, with an AUC of 86.19%. This suggests useful discriminatory ability within the evaluated dataset, although independent validation is required before the model can be recommended for operational surveillance. Comparable research has demonstrated the potential of machine-learning approaches for identifying brucellosis-associated factors. Megahed et al. (2022), comparing logistic regression with a classification and regression tree model in 400 cattle from Egypt, reported an AUC of 0.98 for the tree model and 0.89 for logistic regression. These results support the potential value of non-linear approaches, although differences in study populations, predictors, and validation procedures preclude direct comparison of model performance. In the present study, district, farmer education, farm role, and disease awareness were among the prominent predictors. District may capture differences in husbandry systems, animal movement, and access to veterinary services, but the specific mechanisms underlying the geographical pattern remain to be established.

Overall, this study integrates serological testing, multivariable logistic regression, and machine learning to characterise *Brucella* seropositivity and its associated factors among dairy cattle with reproductive disorders in Rwanda. The findings identify potential priorities for targeted surveillance, farmer education, vaccination programmes, and farm-level biosecurity, while demonstrating the potential of machine learning to support risk prediction in veterinary epidemiology.

Declarations

Ethics approval:

Ethical approval for this study was obtained from the Institutional Review Board (IRB) of the University of Rwanda, College of Agriculture, Animal Sciences and Veterinary Medicine (UR-CAVM) (Reference No.

DRI012/EC/26).

Consent to participate:

All participants provided written informed consent before enrolment in the study.

Consent for publication:

Not Applicable

Competing interests:

*T*he authors have no relevant financial or non-financial interests to disclose

Author's Contribution

SM, EN and MM conceived the study and developed the study protocol; JM and MG carried out data collection, laboratory analysis, and interpretation of these data. SM, EN and MM drafted the manuscript; PN, FN and MMu critically revised the manuscript for intellectual content. All authors read and approved the final version of the manuscript.

Funding:

This study was supported by the Rwanda Dairy Development project (RDDP II).

Acknowledgements

We gratefully acknowledge the dairy farmers who participated in this study and extend our sincere appreciation to the staff of the Rwanda Agricultural Board for their invaluable support during the data collection. We also thank the Rwanda Dairy Development Project for their contribution to this study.

Data availability:

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Code availability:

Not applicable.

References

1. Asakura, S., Makingi, G., Kazwala, R., & Makita, K. (2018). Herd-level risk factors associated with Brucella sero-positivity in cattle, and perception and behaviours on the disease control among agro-pastoralists in Tanzania. *Acta Tropica*, *187*, 99–107. https://doi.org/10.1016/j.actatropica.2018.07.010
2. Assenga, J. A., Matemba, L. E., Muller, S. K., Malakalinga, J. J., & Kazwala, R. R. (2015). Epidemiology of Brucella infection in the human, livestock and wildlife interface in the Katavi-Rukwa ecosystem, Tanzania. *BMC Veterinary Research*, *11*(1), 189. https://doi.org/10.1186/s12917-015-0504-8
3. Djangwani, J., Abong', G. O., Njue, L. G., & Kaindi, D. W. M. (2021). Sero-prevalence and risk factors of *Brucella* presence in farm bulk milk from open and zero grazing cattle production systems in Rwanda. *Veterinary Medicine and Science*, *7*(5), 1656–1670. https://doi.org/10.1002/vms3.562
4. Djangwani, J., Ooko Abong', G., Gicuku Njue, L., & Kaindi, D. W. M. (2021). Brucellosis: Prevalence with reference to East African community countries – A rapid review. *Veterinary Medicine and Science*, *7*(3), 851–867. https://doi.org/10.1002/vms3.425
5. Ducrotoy, M. J., Muñoz, P. M., Conde-Álvarez, R., Blasco, J. M., & Moriyón, I. (2018). A systematic review of current immunological tests for the diagnosis of cattle brucellosis. *Preventive Veterinary Medicine*, *151*, 57–72. https://doi.org/10.1016/j.prevetmed.2018.01.005
6. Edao, B. M., Ameni, G., Assefa, Z., Berg, S., Whatmore, A. M., & Wood, J. L. N. (2020). Brucellosis in ruminants and pastoralists in Borena, Southern Ethiopia. *PLOS Neglected Tropical Diseases*, *14*(7), e0008461. https://doi.org/10.1371/journal.pntd.0008461
7. Heidary, M., Dashtbin, S., Ghanavati, R., Mahdizade Ari, M., Bostanghadiri, N., Darbandi, A., Navidifar, T., & Talebi, M. (2022). Evaluation of Brucellosis Vaccines: A Comprehensive Review. *Frontiers in Veterinary Science*, *9*, 925773. https://doi.org/10.3389/fvets.2022.925773
8. Izadi, S., Moghaddas, V., Feizi, A., Bahreinipour, A., & Barati, Z. (2024). Bovine brucellosis, associated risk factors and preventive measures in industrial cattle farms. *Heliyon*, *10*(22), e40180. https://doi.org/10.1016/j.heliyon.2024.e40180
9. Kerr, C. L., Patassi, A., Pato, P. S., Guitian, J., Diop, S. A., Musallam, I., Mangtani, P., & Nguipdop-Djomo, P. (2026). Seroprevalence and risk factors for Brucella species and Coxiella burnetii exposure in a cross-sectional serosurvey of occupationally exposed groups in peri-urban Lomé, Togo. *PLOS Neglected Tropical Diseases*, *20*(1), e0012657. https://doi.org/10.1371/journal.pntd.0012657
10. Kiiza, D., Biryomumaisho, S., Robertson, I. D., & Hernandez, J. A. (2021). Seroprevalence of and Risk Factors Associated with Exposure to Brucella Spp. In Dairy Cattle in Three Different Agroecological Zones in Rwanda. *The American Journal of Tropical Medicine and Hygiene*, *104*(4), 1241–1246. https://doi.org/10.4269/ajtmh.20-1426
11. Li, Z., Gui, D., Sun, Z., Zhang, J., Zhang, W., Zhang, H., Guo, F., & Chen, C. (2015). Immunization of BALB/c mice with *Brucella abortus* 2308Δ *wbkA* confers protection against wild-type infection. *Journal of Veterinary Science*, *16*(4), 467. https://doi.org/10.4142/jvs.2015.16.4.467

12. Megahed, A., Kandeel, S., Alshaya, D. S., Attia, K. A., AlKahtani, M. D. F., Albohairy, F. M., & Selim, A. (2022). A comparison of logistic regression and classification tree to assess brucellosis associated risk factors in dairy cattle. *Preventive Veterinary Medicine*, *203*, 105664. https://doi.org/10.1016/j.prevetmed.2022.105664
13. Mubanga, M., Mfune, R. L., Kothowa, J., Mohamud, A. S., Chanda, C., Mcgiven, J., Bumbangi, F. N., Hang'ombe, B. M., Godfroid, J., Simuunza, M., & Muma, J. B. (2021). Brucella Seroprevalence and Associated Risk Factors in Occupationally Exposed Humans in Selected Districts of Southern Province, Zambia. *Frontiers in Public Health*, *9*, 745244. https://doi.org/10.3389/fpubh.2021.745244
14. Ndazigaruye, G., Mushonga, B., Kandiwa, E., Samkange, A., & Segwagwe, B. E. (2018). Prevalence and risk factors for brucellosis seropositivity in cattle in Nyagatare District, Eastern Province, Rwanda. *Journal of the South African Veterinary Association*, *89*. https://doi.org/10.4102/jsava.v89i0.1625
15. Ntivuguruzwa, J. B., Kabalisa, E., Bosco Asifiwe, H. J., Uwamahoro, A. G., Nyampinga, I., Mukarurangwa, A., Rujeni, N., Habimana, J. P., Mukamuhirwa, M. L., Karemera, P., Umukunzi, P., Iradukunda, B., Mazimpaka, P., Hudson, P. J., Byukusenge, M., Buza, J. J., Kapur, V., Ndikubwimana, A. A., Mukagatare, I., … Katani, R. (2025). Seroprevalence and risk factors associated with brucellosis in humans and livestock in Nyagatare district of Rwanda. *Frontiers in Public Health*, *13*, 1665341. https://doi.org/10.3389/fpubh.2025.1665341
16. Ntivuguruzwa, J. B., Kolo, F. B., Gashururu, R. S., Umurerwa, L., Byaruhanga, C., & Van Heerden, H. (2020). Seroprevalence and Associated Risk Factors of Bovine Brucellosis at the Wildlife-Livestock-Human Interface in Rwanda. *Microorganisms*, *8*(10), 1553. https://doi.org/10.3390/microorganisms8101553
17. Pal, M., Gizaw, F., Fekadu, G., Alemayehu, G., & Kandi, V. (2017). Public Health and Economic Importance of Bovine Brucellosis: An Overview. *American Journal of Epidemiology and Infectious Disease*, *5*(2), 27–34. https://doi.org/10.12691/ajeid-5-2-2
18. Rujeni, N., & Mbanzamihigo, L. (2014). Prevalence of Brucellosis among Women Presenting with Abortion/Stillbirth in Huye, Rwanda. *Journal of Tropical Medicine*, *2014*, 1–3. https://doi.org/10.1155/2014/740479
19. Smith, L. J., Schurer, J. M., Ntakiyisumba, E., Shyaka, A., & Amuguni, J. H. (2021). Rift Valley fever knowledge, mitigation strategies and communication preferences among male and female livestock farmers in Eastern Province, Rwanda. *PLOS Neglected Tropical Diseases*, *15*(8), e0009705. https://doi.org/10.1371/journal.pntd.0009705
20. Thanh Noi, P., & Kappas, M. (2017). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors*, *18*(1), 18. https://doi.org/10.3390/s18010018
21. Ting, S., Freitas, M., Mantilo, M., França, A., Gomes, C., Pereira, A. J., Chen, W., Alves, L., Givney, R., Francis, J. R., Yan, J., Toribio, J.-A. L. M. L., Monteiro, M. A. A., Barnes, T. S., & Jong, J. B. D. C. (2025). Brucellosis in Timor-Leste: Knowledge, attitudes, and practices among cattle farmers and animal

health workers. *Preventive Veterinary Medicine*, *244*, 106621. https://doi.org/10.1016/j.prevetmed.2025.106621

Figures

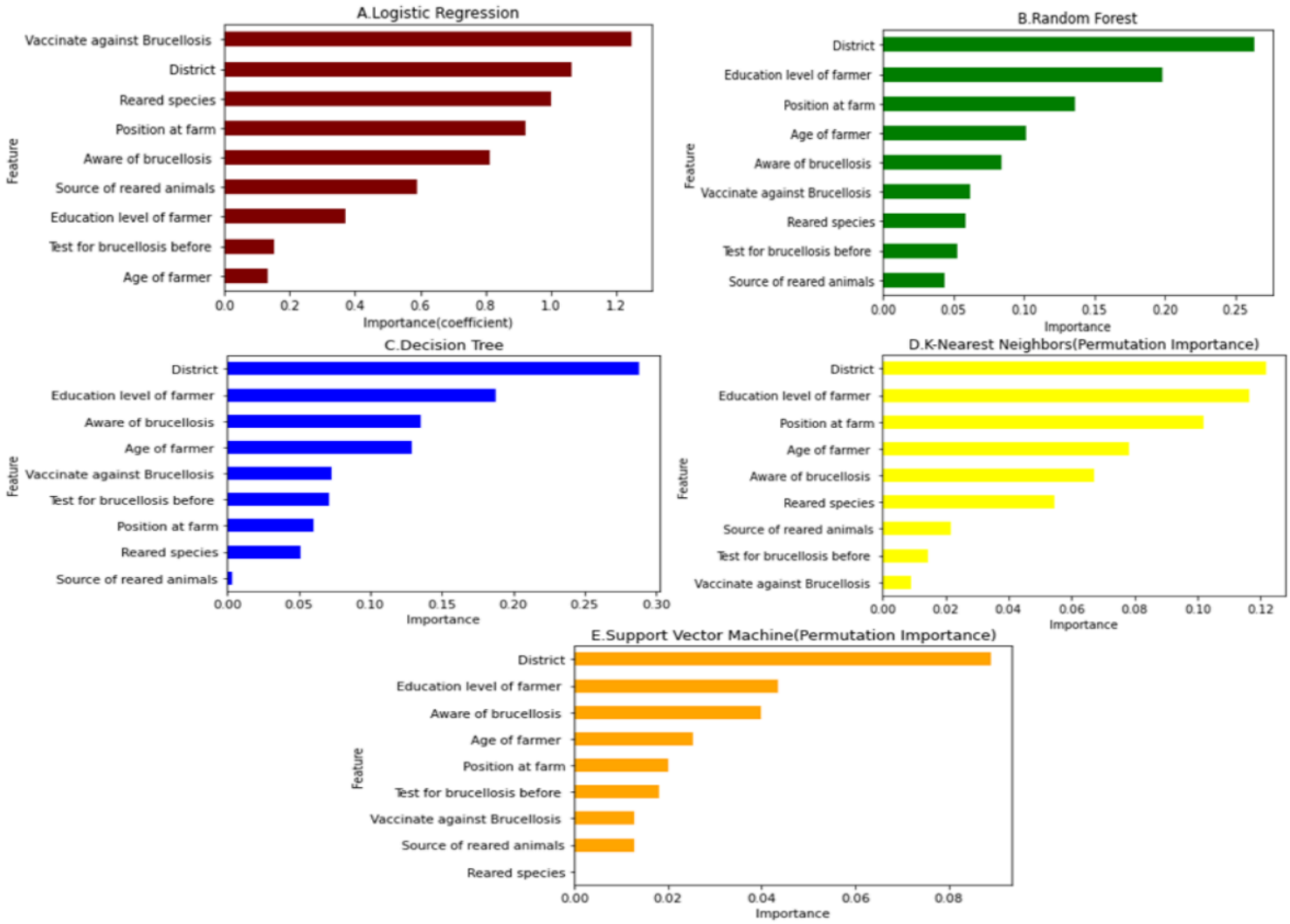


Figure 1

Comparative feature importance and coefficient rankings across the five evaluated machine learning models for predicting *Brucella* seropositivity.