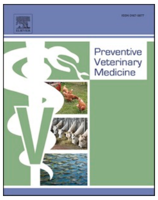
Preventive Veterinary Medicine 203 (2022) 105664

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

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



A comparison of logistic regression and classification tree to assess brucellosis associated risk factors in dairy cattle

Ameer Megahed[a,b,*], Sahar Kandeel[c], Dalal S. Alshaya[d,**], Kotb A. Attia[e], Muneera D.F. AlKahtani[d], Fatima M. Albohairy[f], Abdelfattah Selim[c,***]

[a] Department of Animal Medicine (Internal Medicine), Faculty of Veterinary Medicine, Benha University, Moshtohor-Toukh, Kalyobiya 13736, Egypt
[b] Department of Large Animal Clinical Sciences, College of Veterinary Medicine, University of Florida, FL 32610, USA
[c] Department of Animal Medicine (Infectious Diseases), Faculty of Veterinary Medicine, Benha University, Moshtohor-Toukh, Kalyobiya 13736, Egypt
[d] Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 102275, Riyadh 11675, Saudi Arabia
[e] Center of Excellence in Biotechnology Research, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia
[f] Extramural Research Department , Health Sciences Research Center, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

ARTICLE INFO

Keywords:
Brucellosis
Seroprevalence
Decision tree
Risk factors
Dairy cattle

ABSTRACT

Machine learning approaches have been increasingly utilized in the field of medicine. Brucellosis is one of the most common contagious zoonotic diseases with significant impacts on livestock health, reproduction, production, and public health worldwide. Therefore, our objective was to determine the seroprevalence and compare the logistic regression and Classification and Regression Tree (CART) data-mining analysis to assess risk factors associated with *Brucella* infection in the densest cattle populated Egyptian governorates. A cross-sectional study was conducted on 400 animals (383 cows, 17 bulls) distributed over four Governorates in Egypt's Nile Delta in 2019. The randomly selected animals from studied geographical areas were serologically tested for *Brucella* using iELISA, and the animals' information was obtained from the farm records or animal owners. Eight supposed risk factors (geographic location, gender, herd size, age, history of abortion, shared equipment, and disinfection post-calving) were evaluated using multiple stepwise logistic regression and CART machine-learning techniques. A total of 84 (21.0%; 95% CI 17.1–25.3) serum samples were serologically positive for *Brucella*. The highest seroprevalence of *Brucella* infection was reported among animals raised in herd size > 100 animals (65.5%), with no disinfection post-calving (61.7%), with a history of abortion (59.6%), and with shared equipment without thorough cleaning and disinfection (57.1%). The multiple stepwise logistic regression modeling identified herd size, history of abortion, and disinfection post-calving as important risk factors. However, CART modeling identified herd size, disinfection post-calving, history of abortion, and shared equipment as the most potential risk factors for *Brucella* infection. Comparing the two models, CART model showed a higher area under the receiver operating characteristic curve (AUROC = 0.98; 95% CI 0.95 – 1.00) than the binary logistic regression (AUROC = 0.89; 95% CI 0.73 – 0.92). Our findings strongly imply that *Brucella* infection is most likely to spread among animals raised in large herds (>100 animals) with a history of abortions and bad hygienic measures post-calving. The CART data-mining modeling provides an accurate technique to identify risk factors of *Brucella* infection in cattle.

1. Introduction

Brucellosis is one of the most prevalent and highly contagious but neglected zoonotic diseases worldwide, except in some developed countries that have managed to eradicate it. The disease is caused by *Brucella spp.*, gram-negative intracellular bacteria (Akhvlediani et al., 2017; Ducrotoy et al., 2018). Twelve species of *Brucella* have been identified to date, and most of them can infect several species of animals, including humans (Godfroid et al., 2010). In cattle, *Brucella* infection is primarily caused by *B. abortus*, less often by *B. melitensis,* and

[*] Correspondence to: 130 Deriso Hall, 2015 SW 16th Ave., Gainesville, FL 32610-0136, USA.
[**] Correspondence to: College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 102275, Riyadh 11675, Saudi Arabia.
[***] Correspondence to: Faculty of Veterinary Meddicine, Benha University, Moshtohor-Toukh, Kalyobiya 13736, Egypt.
E-mail addresses: ameermegahed@ufl.edu (A. Megahed), dsalshaya@pnu.edu.sa (D.S. Alshaya), Abdelfattah.selim@fvtm.bu.edu.eg (A. Selim).

https://doi.org/10.1016/j.prevetmed.2022.105664
Received 1 November 2021; Received in revised form 27 April 2022; Accepted 29 April 2022
Available online 4 May 2022

occasionally by *B. suis* (Díaz, 2013; OIE, 2018). In underdeveloped countries with low and moderate-income, the disease is often underreported with little or inefficient control program, resulting in significant health, economic, and livelihood burdens (McDermott et al., 2013a). This might be attributed to the misclassification of the disease as other reproductive diseases (Hegazy et al., 2011; Ducrotoy et al., 2017). In several countries, Bovine brucellosis is endemic, including Egypt, affects humans and animals of both gender, and is considered a significant risk to public health (Yu and Nielsen, 2010; Khurana et al., 2020).

In adult cattle, the infection localizes in the reproductive organs resulting in placentitis followed by abortion, causing production and reproduction losses, including a decrease in milk yield, chronic metritis, and decrease in fertility rate (McDermott et al., 2013b; Kothalawala et al., 2017; Franc et al., 2018). However, most infected animals abort once in their lifetime and retain the infection during their entire life (Godfroid et al., 2010). After the first abortion or in non-pregnant female cattle, the disease remains asymptomatic. The animals can shed the bacteria in their discharges, which is considered an essential source for spreading the infection between susceptible hosts (Hosein et al., 2018; Jamil et al., 2020). Therefore, the periodical monitoring of animals for brucellosis using serological tests would promote the detection of infected animals, contribute effective control measures, and decrease the spreading of brucellosis (Gwida et al., 2015).

Indirect enzyme-linked immunosorbent assay (iELISA) is the primary serological test used to screen brucellosis among susceptible animals and humans because of its high sensitivity and specificity (Nielsen, 2002; Chisi et al., 2017). According to available seroprevalence studies in Egypt, brucellosis is endemic in both animals and humans, despite a control program and strategic vaccination for animals (Abdelbaset et al., 2018; Hosein et al., 2018). In a small-scale study, the seroprevalence of brucellosis was 16.7% and 16.3% in cattle and sheep, respectively (Selim et al., 2019). Despite the effort attributed to control brucellosis in Egypt, the reasons behind its persistence are still poorly understood. However, the lack of adequate epidemiological data on the seroprevalence of *Brucella* and related risk factors can impede establishing efficient strategic control programs (Gwida et al., 2015; Eltholth et al., 2017).

Classification and Regression Tree (CART) is a machine-learning algorithm that has been used in clinical settings as an effective tool for clinical decision-making and risk factor assessment (Yakubu et al., 2015; Mburu et al., 2018; Selim et al., 2021). In the last decade, data mining modeling has started to be used in veterinary epidemiological studies. Recently, data mining techniques, including random forest, support vector machine, multivariate adaptive regression splines, and decision tree have been used to predict *Brucella* infection in cattle (Shirmohammadi-Khorram et al., 2019; Khan et al., 2020). However, to the best of our knowledge, no study compared the performance of both traditional logistic regression and CART modelings for identifying the most important risk factors of *Brucella* infection in cattle. Our objective was therefore to determine the seroprevalence of brucellosis and compare the binary logistic regression and CART machine-learning modelings to identify the risk factors associated with brucellosis in Egyptian dairy cattle.

2. Materials and methods

2.1. Ethical considerations

This study was permitted by the ethical committee of the Faculty of Veterinary Medicine, Benha University, and the blood samples were taken from cattle under owner̓ s consent.

2.2. Study area

The study was conducted in four Governorates (Gharbia, GB; Kafr El-Sheikh, KF; Menofia, MF; and Qalyoubia, QL) geographically situated at the Nile Delta of Egypt. The Delta region has a hot desert climate like the rest of Egypt and is located near Egypt's north coast. This region has a moderate temperature (average 25 °C) almost of year and increases only during the summer months. The average annual rainfall ranges from100 to 200 mm and mainly occurs during the winter months. Most of the cattle residing in these governorates are household cross-breeds and kept in a semi-grazing system. In Egypt, the cross-border movement of animals between governorates is active. Historically, agriculture, including livestock farming, is the backbone of Delta-people's income, and close human-animal interactions are its primary features.

2.3. Study design and sampling

A cross-sectional study was conducted between April 2018 to November 2019 in four Governorates located at the Nile Delta of Egypt. The selected areas have a high density of cattle population. The sample size needed for the present study was determined using the formula for descriptive studies:

$$[DEFF*Np(1-p)]/\ [(d_2/Z_{1-\alpha/2}*(N-1)+p*(1-p)],$$

Where, N = population size (150,000), p = prevalence of brucellosis (16.7; Selim et al., 2019), d = precision (1), DEFF = design effect (1.0), $Z_{1-\alpha/2} = 1.96$. Under these assumptions, a minimum of 370 animals was deemed necessary. Considering attrition of 8%, the final sample size of 400 was calculated to be enrolled from the four Governorates in this study. Governorates stratified the enrolled animals according to the approximate number of animals in each governorate obtained from the Animal Wealth Development Sector. All animals were randomly selected from different geographic locations within the governorate. Individual data, including location, herd size, age, shared equipment, history of abortion, and disinfection post-calving were recorded. Blood samples were obtained from the jugular vein of each examined animal using 20 G needles and 10 mL blood collection tubes. Serum was collected after centrifugation of the blood samples at 3000xg/min for 10 min. The serum samples were stored at − 20 °C until serological testing.

2.4. Serological analysis

All sera were tested for antibodies against *B. abortus* using a commercial iELISA kit (IDEXX Brucellosis Serum X2 Ab Test) according to the manufacturer's instructions. An ELISA reader measured the optical densities (ODs) of samples at 450 nm. The sample is considered positive if S/P ratio is ≥ 80%, and negative if S/P ratio is < 80%.

2.5. Data analysis

The seroprevalence of brucellosis was estimated with the exact binomial confidence intervals of 95% using PROC FREQ of SAS analytical procedure. The associations of *B. abortus* infection with different risk factors were evaluated using the Cochran-Armitage trend test, and the strength of associations was assessed through Phi coefficient value using PROC FREQ of SAS analytical procedure. Univariable logistic regression was used for the initial screening of investigated exposure factors associated with *B. abortus* infection. The logistic model, fitted with *B. abortus* infection as the outcome variable (present: 1, absent: 0), and history of abortion (2 levels: yes and no), sex (2 levels: male, female), age (3 levels: <4, 4–8, ≥8 years), herd size (3 levels: <30, ≥30–100, ≥100 animals), shared equipment (2 levels: yes and no), disinfection post-calving (2 levels: yes and no), and geographic location (5 levels: GB, KF, MF, and Qal) as exposure factors. The predicted probability curves for the most important risk factors were created using univariable logistic regression model-predicted probabilities.

Stepwise forward multivariable logistic regression was used to identify the most critical risk factor(s) associated with *B. abortus* infection based on the lowest value for the Akaike information criterion

(AIC). The stepwise forward multivariable logistic regression model was built by starting with no variables in the model. The 'stopping rule' for inclusion or exclusion of variables was based on the AIC. The logistic regression model predicts the log odds (logit) for the outcome as an additive function of the risk factors. The prevalence odds ratios (POR) were used as an approximate measure of relative risk (the likelihood of having a positive result for iELISA in an animal with a given risk factor compared with an animal without the risk factor). Confounding between risk factors retained in final models were examined by adding each of the variables to the model and assessing the changes in the POR (i.e., ≥ 20%) of the remaining variables in the model (Kiiza et al., 2021). Interaction between variables was tested by adding new terms to the model for every two variables for which interaction is being assessed. The coefficient of this term is then analyzed to see if the combination of these two variables affects the *Brucella* seropositivity (Harrell, 2015). Regression analysis was performed using SAS 9.4 (SAS Inst. Inc., Cary, NC), and $P < 0.05$ was considered significant.

A data mining technique, CART, was used to highlight the relationship between significant risk factors and their hierarchical classification in the tree diagram visualization. Briefly, *B. abortus* infection was scored on a binomial scale: 0 – absent, 1 – present. The initial dataset of 400 animals was divided into training (280, 70%) and validation (120, 30%) datasets using a stratified sampling method. The classification tree model was developed based on the assumption that the maximum depth of the tree (number of branches) is 6. Splitting (Gini index) and pruning (cross-validation) steps were used to build the classification tree (Breiman et al., 1984). The classification tree algorithm used the following variables: history of abortion, sex, age, herd size, shared equipment, disinfection post-calving, and geographic location to select the associated exposure factors. The tree model was assessed using the validation dataset through the following criteria: sensitivity (Se), specificity (Sp), misclassification rate, and the area under the receiver operating characteristics curves (AUROC). Decreasing values of the misclassification rate and increasing values of Se, Sp, and AUROC indicate higher quality of the information in the models. The ranking and significance of risk factors in terms of their importance were created based on the "Importance" measure, the percentage of agreeing cases when the main and surrogate splits are compared (SAS Institute Inc, 2014). Importance takes values in the range of 0–1, the higher the value, the greater the importance of a given measure in constructing the classification tree (i.e. generating splits). Importance measures are based on the reduction of the Gini score (Piwczyński et al., 2012). Finally, the model validation was performed on the same set of held-aside data for both modeling approaches and assessed through AUROC. All data mining modeling was performed with SAS® OnDemand for Academics (PROC HPSPLIT; SAS Inst. Inc., Cary, NC).

3. Results

3.1. Seroprevalence of brucellosis

The seroprevalence of brucellosis in dairy herds was defined in 400 serum samples obtained from dairy cattle (379 females and 21 males) with ages from < 4 to > 8 years old, raised in herd size between < 30 to > 100 animals, and located in four governorates (GB, KF, MF, and Ql) in Northern Egypt.

Overall, the seroprevalence of brucellosis was 84/400 at animal-level (21%; 95% CI 17.1–25.3). The results of univariable logistic regression showed that the seroprevalence of brucellosis was non-significant differed between localities and between males and females under the study. Gharbia governorate showed the highest seroprevalence of brucellosis (23.4%), while KF governorate showed the lowest seroprevalence for the disease (17.1%), as shown in Table 1 and Fig. 1. The distribution of *Brucella*-positive animals was differed between herd size ($P < 0.00$), age ($P < 0.001$), history of abortion ($P < 0.001$), disinfection post-calving ($P < 0.001$), and shared equipment ($P < 0.001$). The highest seroprevalence of brucellosis was present among animals raised in herd size > 100 animals (65.5%), with no disinfection post-calving (61.7%), with a history of abortion (59.6%), and with shared equipment without thorough cleaning and disinfection (57.1%) as shown in Table 1.

Our results showed moderate to high associations between the seroprevalence of *Brucella* infection and disinfection post-calving (Phi coefficient = 0.70), herd size (0.59), history of abortion (0.52), age in years (0.52), and shared equipment (0.48). However, no association was reported between *Brucella* seropositivity and geographic location (0.06) or gender (0.10).

3.2. Risk factors analysis

The findings of univariable logistic regression revealed that the animals raised in herds size > 100 animals, with a history of abortion and no disinfection post-calving, increased the probability of *Brucella* seropositivity by 64.1%, 62.3%, and 72.0%, respectively.

The final forward stepwise multivariable logistic regression model showed that herd size, history of abortion, and disinfection post-calving were significant risk factors for *Brucella*-infected animals (Table 2).

The CART model developed a decision tree for the most important risk factors of brucellosis with misclassification rate of 5.2% (Fig. 2). The sensitivity and specificity of the CART model were 81.0% and 98.4%, respectively. The first node in the tree diagram indicates the highest risk factor. The first node in the tree diagram was the herd size with an importance score of 10.7. The second node was the disinfection post-

Table 1
Univariable logistic regression analysis for identification of risk factors associated with *B. abortus* infection in 400 dairy cattle in the Nile Delta of Egypt.

Variable	Category	N	Positive	Prevalence (%)	POR (95% CI)	*P*-value
Geographic location	Gharbia	94	22	23.4	1.0 (Reference)	0.825
	Kafr El-Sheikh	111	19	17.1	0.7(0.3–1.3)	
	Menofia	96	22	22.9	1.0 (0.5–1.9)	
	Qalyoubia	99	21	21.1	0.9 (0.4–1.7)	
Gender	Male	17	1	5.9	1.0 (Reference)	0.151
	Female	383	83	21.6	4.4 (0.6–33.9)	
Herd size	< 30	178	5	2.8	1.0 (Reference)	< 0.001
	30–100	138	24	17.4	7.3 (2.7–19.6)	
	> 100	84	55	65.5	65.6 (24.2–177.7)	
Age (years)	< 4	162	7	4.3	1.0 (Reference)	< 0.001
	4–8	157	27	17.2	4.6 (1.9–10.9)	
	> 8	80	50	62.5	36.9 (15.3–89.2)	
Shared equipment	No	309	32	10.4	1.0 (Reference)	< 0.001
	Yes	91	52	57.1	11.5 (6.6–20.1)	
Abortion	No	306	28	9.2	1.0 (Reference)	< 0.001
	Yes	94	56	59.6	14.6 (8.3–25.8)	
Disinfection post-calving	Yes	267	2	0.8	1.0 (Reference)	< 0.001
	No	133	82	61.7	213.0 (50.8–894.0)	

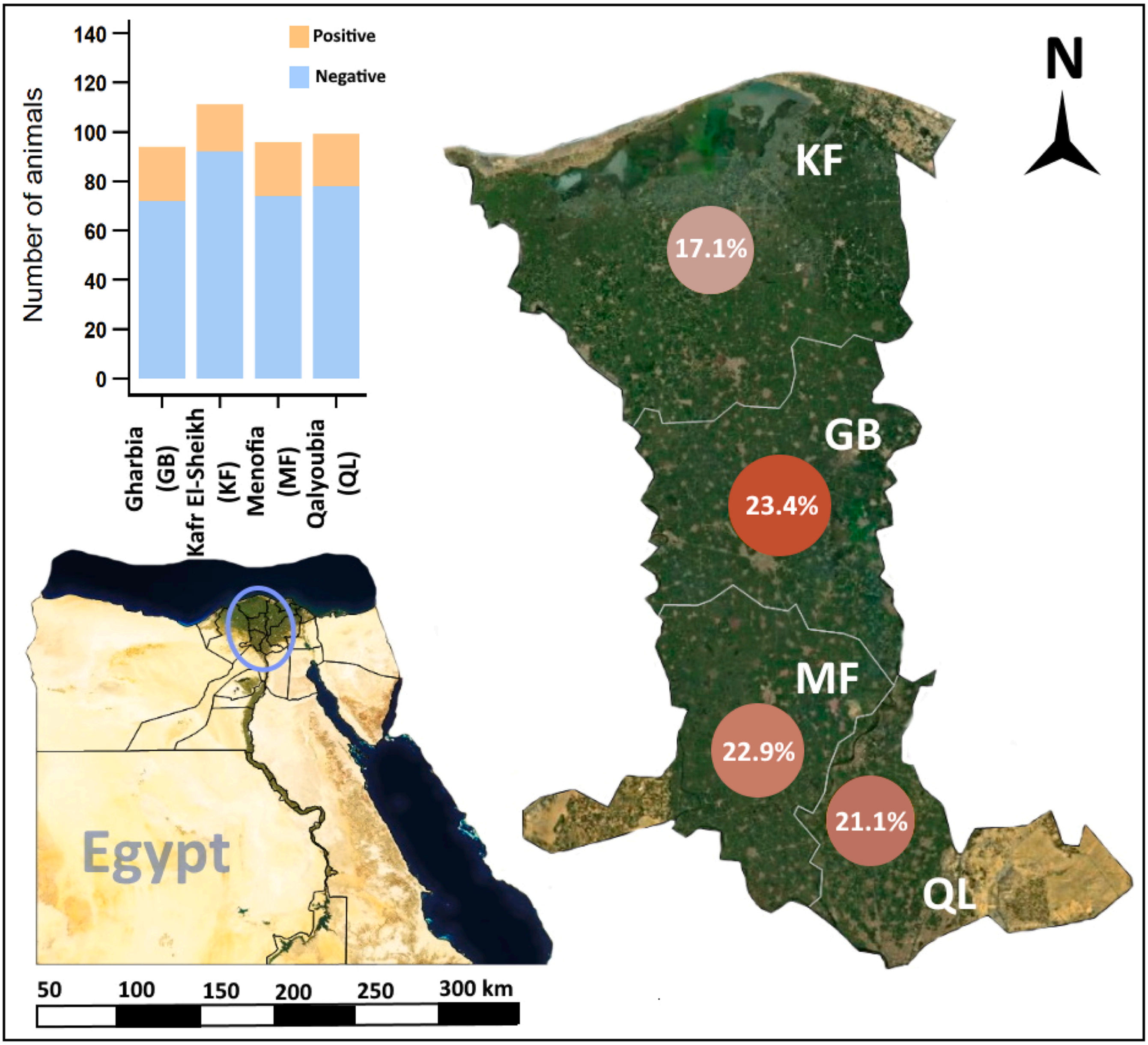


Fig. 1. Geographic distribution of *B. abortus* infection in dairy cattle of the Nile Delta of Egypt.

Table 2
Multiple stepwise logistic regression analysis of potential risk factors associated with *B. abortus* seropositivity in dairy cattle in the Nile Delta of Egypt.

Variable	Categories	Estimate	SE	P-value	POR_{adj}	95% CIOR
Intercept		-7.0	1.0	< 0.001	–	–
Herd size	< 30	Reference				
	30–100	2.6	1.1	0.016	4.4	1.39–14.3
	> 100	3.9	1.1	< 0.001	30.5	7.7–120.7
Abortion	No	Reference				
	Yes	1.3	0.5	0.010	3.9	1.5–10.4
Disinfection post-calving	Yes	Reference				
	No	5.4	0.8	< 0.001	212.3	42.4-> 999.9

calving (6.9) and the history of abortion (4.2). The fourth important risk factor was shared equipment between animals (2.4).

Comparing CART with binary logistic regression analyses, the CART model generally showed a superior model performance than logistic regression with an AUC= 0.98 (95% CI 0.95 – 1.00), compared to AUC of 0.89 (95% CI 0.73 – 0.92) for the logistic regression model.

4. Discussion

Brucellosis is considered one of the most dangerous zoonotic diseases that cause chronic debilitating illnesses in humans and substantial loss of productivity in livestock industries. For thousands of years, brucellosis has been an endemic disease in Egypt. Therefore, the main goals of

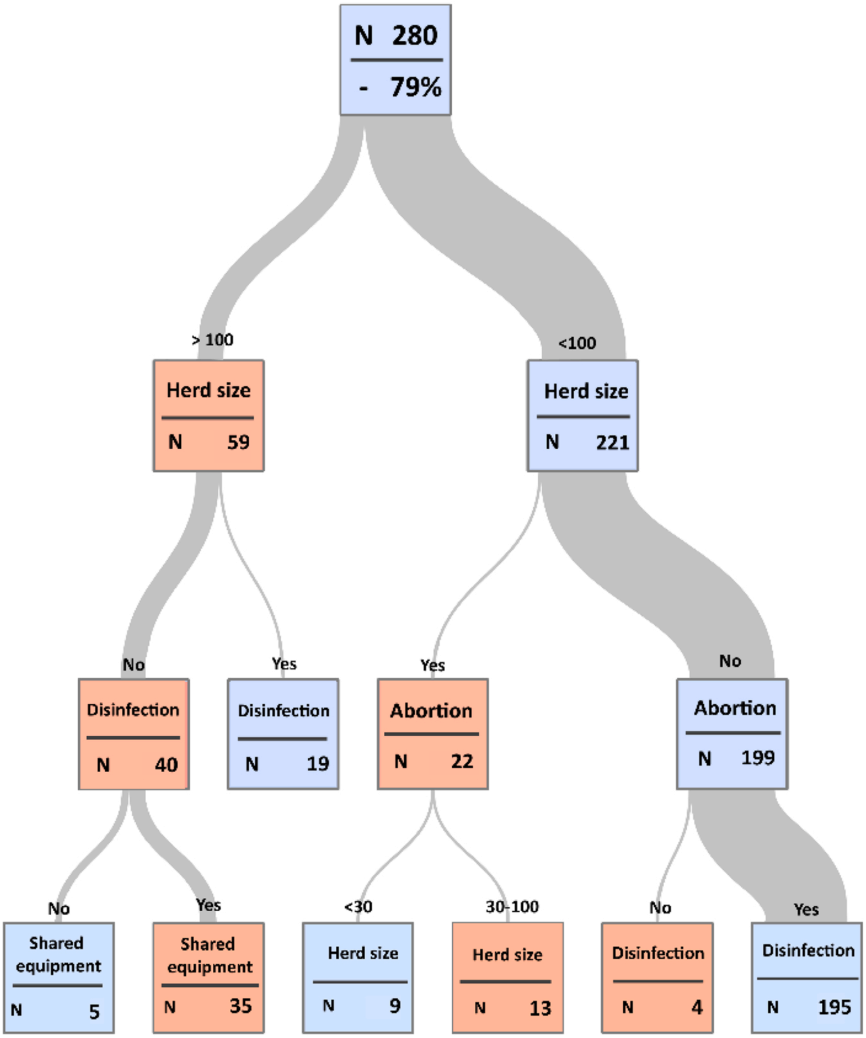


Fig. 2. Estimated classification tree determining the significant risk factors of *Brucella* in dairy cattle. This tree started with node 0 representing all observations. The color of the squares indicates the absence (blue) and the presence of *Brucella* (pink). Each node has a number of *Brucella* infected or non-infected animals. The width of the link between parent and child nodes is proportional to the number of observations in the child node. Classification tree analysis indicated that herd size, disinfection post-calving, history of abortion, and shared equipment are important risk factors of *Brucella*.

this research are to provide updated prevalence information on brucellosis in dairy cattle in Egypt and evaluate the logistic regression and decision tree data-mining analysis in identifying the risk factors associated with *Brucella* infection in Egyptian dairy cattle to provide an accurate guide for implementing effective prevention-control strategies. The main result of this study is that CART modeling showed an accurate analytical approach for identifying potential risk factors of *Brucella* infection in cattle. Accordingly, the herd size > 100 animals, no disinfection post-calving, history of abortion, and shared equipment might predict a higher risk for *Brucella* infection in dairy cattle. The CART machine-learning algorithm is a powerful, robust technique used in public health studies to assess risk factors (Yakubu et al., 2015; Mburu et al., 2018). Additionally, the decision tree is a powerful analytical method to identify interaction among independent variables, where the nature of the decision tree itself implies interacting variables without specifying a variable that is the interaction (Oh, 2019). The lower analytical performance of logistic regression compared to CART is might be due to the mathematical properties of logistic regression depending on the sample size and the data structure (Nemes et al., 2009).

The total seroprevalence of brucellosis in dairy cattle defined in this study is 21% that is higher than that reported in the earlier studies 5.4% (Samaha et al., 2008), 11.0% (Holt et al., 2011), 16.7% (Selim et al., 2019), confirming that brucellosis is endemic in the studied locations. However, comparing the estimated seroprevalence of brucellosis in our study with earlier Egyptian studies should be taken with caution because lack of unbiased studies that estimated brucellosis seroprevalence. Of interest, the reported seroprevalence of brucellosis in India ranged from 20% to 60% from 2004 to 2016, confirming our thought (Chand and Sharma, 2004; Jagapur et al., 2013; Pathak et al., 2016). However, this higher seroprevalence rate in our study might be due to the infected animals remaining infected the entire life (Godfroid et al., 2010; Deka et al., 2018). Therefore, the estimated seroprevalence in this study reflects the proportion of cows that are potentially shedding the *Brucella* organism. Additionally, if any animals in this study were vaccinated, the estimated seroprevalence might have been overestimated (Holt et al., 2011).

Our results showed that brucellosis is present in all studied governorates, with GB governorate recorded the highest seroprevalence among the studied areas. This might be attributed to cattle management differences and other agroecological factors that promoted or restricted contact between herds, in addition to the existence of the largest animal trading market in the Nile Delta region in the Gharbia governorate (Hegazy et al., 2011). However, the reported seroprevalence between the localities was not statistically significant, which is considered a sensible result because the four localities have a similar geographic and climatic nature.

The seroprevalence of brucellosis was not statistically different between male and female animals under the study. This result is in agreement with an earlier study (Segwagwe et al., 2018). However, most earlier studies reported higher prevalence in female animals than in male animals (Islam et al., 2013; Mangi et al., 2015; de Alencar Mota et al., 2016). Contrary, one study reported a higher prevalence in male animals than female animals (Mai et al., 2012).

In general, risk factors of brucellosis can be classified into four groups; (1) host factors such as age, sex, breed, history of abortion, retention of placenta, repeat breeding; (2) management factors including herd size, single/mixed herd, the introduction of the new animal; (3) agroecological factors such as geographical location and climate; (4) farmer factor includes farmer qualification, training, and experience (Deka et al., 2018). Among the observed factors that have been assessed in this study is the herd size. Herd size of > 100 animals has been identified as the most critical risk factor for higher *Brucella* seropositivity. This is consistent with earlier studies that found the large herds are more prone to *Brucella*-infection because the animals are kept closely together, resulting in increasing the risk of exposure that provides more opportunities for infection and the maintenance of *Brucella* in the cattle population (Coetzer et al., 1994; Tasiame et al., 2016; Kiiza et al., 2021). Additionally, keeping the optimal management practices is more challenging in the large herds than in small ones (Nicoletti, 1980; Matope et al., 2010; Segwagwe et al., 2018).

Our results showed associations between the seroprevalence of *Brucella* infection and animal age in years. Previous reports suggest that older animals are more likely to be seropositive than younger animals (Hassan et al., 2014; Mugizi et al., 2015). Increased susceptibility to infection with age could be assigned to the earlier exposure of older animals that may be possibly immune or perhaps persistent carriers (Segwagwe et al., 2018). Additionally, *Brucella* infection is more linked to sexual maturity due to the impact of sex hormones and placenta erythritol on the pathogenesis of the disease (Asmare et al., 2013). However, *Brucella* infection was found to be more common in younger calves in a previous investigation, suggesting that age is still an arguable risk factor (Kumar et al., 2016), supporting the results of logistic regression and CART approaches.

Our findings revealed that the history of abortion is a significant risk factor of cattle brucellosis that is consistent with earlier studies (Samaha et al., 2009; Lindahl et al., 2014; Alhaji et al., 2016). This finding corresponds with the biology of *Brucella* organisms as a significant cause of abortion due to the presence of erythritol in the uterus that constitutes the placental tropism for the development of *Brucella*, specifically in ruminants (O'Callaghan, 2013). However, other studies revealed no link between *Brucella* infection and abortion or placenta retention (Asmare et al., 2013; Mugizi et al., 2015). This might be because the infected animals remain infected the entire life (Holt et al., 2011).

Disinfection post-calving and avoiding sharing the equipment between animals seem to be the most imperative preventive measures to reduce the spread of *Brucella* infection among the dairy cattle population. This is a sensible result since the disinfection and cleaning of contaminated premises help kill and remove the causative pathogens. Avoiding sharing equipment decreases the opportunities to transfer *Brucella* organisms from infected animals to susceptible animals and therefore, the maintenance of *Brucella* spp. in cattle populations. It has been reported that the spread of *Brucella* organisms amongst animals is aided by increased contact between animals at shared feeding and watering places (Segwagwe et al., 2018).

The main limitations of this study include, first, the shortage of information about the vaccination history of enrolled animals makes the reported seroprevalence of brucellosis should be taken with caution. Second, being a cross-sectional study that cannot provide adequate evidence on cause and effect relationships. Therefore, longitudinal studies using a larger sample size and broader geographic representation are required to verify the associations obtained in this study. Third, in the present study, we used a hold-out validation strategy in order to obtain independent training and validation datasets. The reduced data and using a single train split can result in an enlarged variance; therefore, other validation approaches such as external validation multiple-fold cross-validation may achieve more accurate performance estimation.

5. Conclusion

The machine-learning classification tree provides a powerful, robust approach for identifying risk factors of *Brucella* infection in cattle. Brucellosis is most likely found in large herd sizes (>100) with a history of abortion and poor hygienic-managemental practices.

Funding

Princess Nourah Bint Abdulrahman University Researchers Supporting Project Number (PNURSP2022R20), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

CRediT authorship contribution statement

A.S., A.M., AK, D.A. and S.K. designed the study and conception of the research idea. **A.S.** sampling collection and performed the laboratory work. **A.M.** conducted data analysis and machine learning. **A.S., A.M., S.K., A.K., R.A. D.A. and I.K.** wrote and prepared the manuscript for publication and revision. All authors read and approved the final manuscript.

Acknowledgments

The authors extend their appreciation to Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2022R20), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflicts of Interest

The authors declare no conflict of interest.

References

Abdelbaset, A.E., Abushahba, M.F., Hamed, M.I., Rawy, M.S., 2018. Sero-diagnosis of brucellosis in sheep and humans in Assiut and El-Minya governorates, Egypt. Int. J. Vet. Sci. Med. 6, S63–S67.

Akhvlediani, T., Bautista, C.T., Garuchava, N., Sanodze, L., Kokaia, N., Malania, L., Chitadze, N., Sidamonidze, K., Rivard, R.G., Hepburn, M.J., 2017. Epidemiological and clinical features of brucellosis in the country of Georgia. PLoS One 12, e0170376.

Alhaji, N., Wungak, Y., Bertu, W., 2016. Serological survey of bovine brucellosis in Fulani nomadic cattle breeds (Bos indicus) of North-central Nigeria: potential risk factors and zoonotic implications. Acta Trop. 153, 28–35.

Asmare, K., Sibhat, B., Molla, W., Ayelet, G., Shiferaw, J., Martin, A., Skjerve, E., Godfroid, J., 2013. The status of bovine brucellosis in Ethiopia with special emphasis on exotic and cross bred cattle in dairy and breeding farms. Acta Trop. 126, 186–192.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. CRC press.

Chand, P., Sharma, A., 2004. Situation of brucellosis in bovines at organized cattle farms belonging to three different states. J. Immunol. Immunopathol. 6, 11–15.

Chisi, S.L., Marageni, Y., Naidoo, P., Zulu, G., Akol, G.W., Van Heerden, H., 2017. An evaluation of serological tests in the diagnosis of bovine brucellosis in naturally infected cattle in KwaZulu-Natal province in South Africa. J. S. Afr. Vet. Assoc. 88, 1–7.

Coetzer, J., Thomson, G., Tustin, R., 1994. Infectious Diseases of Livestock with Special Reference to Southern Africa. Oxford University Press, Cape Town; New York.

de Alencar Mota, A.L.A., Ferreira, F., Neto, J.S.F., Dias, R.A., Amaku, M., Grisi-Filho, J.H. H., Telles, E.O., Gonçalves, V.S.P., 2016. Large-scale study of herd-level risk factors for bovine brucellosis in Brazil. Acta Trop. 164, 226–232.

Deka, R.P., Magnusson, U., Grace, D., Lindahl, J., 2018. Bovine brucellosis: prevalence, risk factors, economic cost and control options with particular reference to India-a review. Infect. Ecol. Epidemiol. 8, 1556548.

Díaz, A., 2013. Epidemiology of brucellosis in domestic animals caused by Brucella melitensis, Brucella suis and Brucella abortus. Rev. Sci. Tech. 32.

Ducrotoy, M., Bertu, W.J., Matope, G., Cadmus, S., Conde-Álvarez, R., Gusi, A.M., Welburn, S., Ocholi, R., Blasco, J., Moriyón, I., 2017. Brucellosis in Sub-Saharan Africa: current challenges for management, diagnosis and control. Acta Trop. 165, 179–193.

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. Prev. Vet. Med. 151, 57–72.

Eltholth, M., Hegazy, Y.M., El-Tras, W.F., Bruce, M., Rushton, J., 2017. Temporal analysis and costs of ruminant brucellosis control programme in Egypt between 1999 and 2011. Transbound. Emerg. Dis. 64, 1191–1199.

Franc, K., Krecek, R., Häsler, B., Arenas-Gamboa, A., 2018. Brucellosis remains a neglected disease in the developing world: a call for interdisciplinary action. BMC Public Health 18, 1–9.

Godfroid, J., Nielsen, K., Saegerman, C., 2010. Diagnosis of brucellosis in livestock and wildlife. Croat. Med. J. 51, 296–305.

Gwida, M., El-Ashker, M., Melzer, F., El-Diasty, M., El-Beskawy, M., Neubauer, H., 2015. Use of serology and real time PCR to control an outbreak of bovine brucellosis at a dairy cattle farm in the Nile Delta region, Egypt. Ir. Vet. J. 69, 1–7.

Harrell, F.E., 2015. Regression Modeling Strategies, second ed. Springer Science + Business Media, New York City, NY.

Hassan, A.A., Uddin, M.B., Islam, M.R., Cho, H.-S., Hossain, M.M., 2014. Serological prevalence of brucellosis of cattle in selected dairy farms in Bangladesh. Korean. J. Vet. Res. 54, 239–243.

Hegazy, Y., Molina-Flores, B., Shafik, H., Ridler, A., Guitian, F., 2011. Ruminant brucellosis in upper Egypt (2005–2008). Prev. Vet. Med. 101, 173–181.

Holt, H.R., Eltholth, M.M., Hegazy, Y.M., El-Tras, W.F., Tayel, A.A., Guitian, J., 2011. Brucella spp. infection in large ruminants in an endemic area of Egypt: cross-sectional study investigating seroprevalence, risk factors and livestock owner's knowledge, attitudes and practices (KAPs). BMC Public Health 11, 1–10.

Hosein, H., Zaki, H.M., Safwat, N.M., Menshawy, A.M., Rouby, S., Mahrous, A., Madkour, B.E.-d, 2018. Evaluation of the General Organization of Veterinary Services control program of animal brucellosis in Egypt: an outbreak investigation of brucellosis in buffalo. Vet. World 11, 748.

Islam, M.R.U., Gupta, M.P., Filia, G., Sidhu, P.K., Shafi, T.A., Bhat, S.A., Hussain, S.A., Mustafa, R., 2013. Sero–epidemiology of brucellosis in organized cattle and buffaloes in Punjab (India). Age 3, 39.

Jagapur, R.V., Rathore, R., Karthik, K., Somavanshi, R., 2013. Seroprevalence studies of bovine brucellosis using indirectenzyme-linked immunosorbent assay (i-ELISA) at organized and unorganized farms in three different states of India. Vet. World 6, 550.

Jamil, T., Melzer, F., Saqib, M., Shahzad, A., Khan Kasi, K., Hammad Hussain, M., Rashid, I., Tahir, U., Khan, I., Haleem Tayyab, M., 2020. Serological and molecular detection of bovine brucellosis at institutional livestock farms in Punjab, Pakistan. Int. J. Environ. Res. Public Health 17, 1412.

Khan, A.U., Melzer, F., Hendam, A., Sayour, A.E., Khan, I., Elschner, M.C., Younus, M., Ehtisham-ul-Haque, S., Waheed, U., Farooq, M., Ali, S., Neubauer, H., El-Adawy, H., 2020. Seroprevalence and molecular identification of Brucella spp. in Bovines in Pakistan—investigating association with risk factors using machine learning. Front. Vet. Sci. 7, 594498.

Khurana, S.K., Sehrawat, A., Tiwari, R., Prasad, M., Gulati, B., Shabbir, M.Z., Chhabra, R., Karthik, K., Patel, S.K., Pathak, M., 2020. Bovine brucellosis–a comprehensive review. Vet. Q. 1–46.

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. Am. J. Trop. Med. Hyg. 104, 1241–1246.

Kothalawala, K.A.C., Makita, K., Kothalawala, H., Jiffry, A.M., Kubota, S., Kono, H., 2017. Association of farmers' socio-economics with bovine brucellosis epidemiology in the dry zone of Sri Lanka. Prev. Vet. Med. 147, 117–123.

Kumar, A., Gupta, V., Verma, A.K., Sahzad, S., Kumar, V., Singh, A., Reddy, N., 2016. Seroprevalence and risk factors associated with bovine brucellosis in western Uttar Pradesh, India. Indian J. Anim. Sci. 86.

Lindahl, E., Sattorov, N., Boqvist, S., Sattori, I., Magnusson, U., 2014. Seropositivity and risk factors for Brucella in dairy cows in urban and peri-urban small-scale farming in Tajikistan. Trop. Anim. Health Prod. 46, 563–569.

Mai, H.M., Irons, P.C., Kabir, J., Thompson, P.N., 2012. A large seroprevalence survey of brucellosis in cattle herds under diverse production systems in northern Nigeria. BMC Vet. Res. 8, 1–14.

Mangi, M., Kamboh, A., Rind, R., Dewani, P., Nizamani, Z., Mangi, A., Nizamani, A., Vistro, W., 2015. Seroprevalence of brucellosis in Holstein-Friesian and indigenous cattle breeds of Sindh Province. Pak. J. Anim. Health Prod. 3, 82–87.

Matope, G., Bhebhe, E., Muma, J., Lund, A., Skjerve, E., 2010. Herd-level factors for Brucella seropositivity in cattle reared in smallholder dairy farms of Zimbabwe. Prev. Vet. Med. 94, 213–221.

Mburu, J.W., Kingwara, L., Ester, M., Andrew, N., 2018. Use of classification and regression tree (CART), to identify hemoglobin A1C (HbA1C) cut-off thresholds predictive of poor tuberculosis treatment outcomes and associated risk factors. J. Clin. Tuberc. Other Mycobact. Dis. 11, 10–16.

McDermott, J., Grace, D., Zinsstag, J., 2013a. Economics of brucellosis impact and control in low-income countries. Rev. Sci. Tech. 32, 249–261.

McDermott, J., Grace, D., Zinsstag, J., 2013b. Economics of brucellosis impact and control in low-income countries. Rev. Sci. Tech. 32, 249–261.

Mugizi, D.R., Boqvist, S., Nasinyama, G.W., Waiswa, C., Ikwap, K., Rock, K., Lindahl, E., Magnusson, U., Erume, J., 2015. Prevalence of and factors associated with Brucella sero-positivity in cattle in urban and peri-urban Gulu and Soroti towns of Uganda. J. Vet. Med. Sci. 14–0452.

Nemes, S., Jonasson, J.M., Genell, A., Steineck, G., 2009. Bias in odds ratios by logistic regression modelling and sample size. BMC Med. Res. Methodol. 9, 1–5.

Nicoletti, P., 1980. The epidemiology of bovine brucellosis. Adv. Vet. Sci. Comp. Med. 24, 69–98.

Nielsen, K., 2002. Diagnosis of brucellosis by serology. Vet. Microbiol. 90, 447–459.

O'Callaghan, D., 2013. Novel Replication Profiles of Brucella in Human Trophoblasts Give Insights Into the Pathogenesis of Infectious Abortion. Oxford University Press.

Oh, S., 2019. Feature interaction in terms of prediction performance. Appl. Sci. 9, 5191.

OIE, T.M., 2018. Infection with Brucella abortus, Brucella melitensis and Brucella suis.

Pathak, A.D., Dubal, Z., Karunakaran, M., Doijad, S.P., Raorane, A.V., Dhuri, R., Bale, M., Chakurkar, E.B., Kalorey, D.R., Kurkure, N.V., 2016. Apparent seroprevalence, isolation and identification of risk factors for brucellosis among dairy cattle in Goa, India. Comp. Immunol. Microbiol. Infect. Dis. 47, 1–6.

Piwczyński, D., Sitkowska, B., Wisniewska, E., 2012. Application of classification trees and logistic regression to determine factors responsible for lamb mortality. Small Rumin. Res. 103, 225–231.

Samaha, H., Al-Rowaily, M., Khoudair, R.M., Ashour, H.M., 2008. Multicenter study of brucellosis in Egypt. Emerg. Infect. Dis. 14, 1916.

Samaha, H., Mohamed, T.R., Khoudair, R.M., Ashour, H.M., 2009. Serodiagnosis of brucellosis in cattle and humans in Egypt. Immunobiology 214, 223–226.

SAS Institute Inc. SAS/STAT® 9.4 User's Guide Cary. SAS Institute Inc. 2014.

Segwagwe, B.E., Samkange, A., Mushonga, B., Kandiwa, E., Ndazigaruye, G., 2018. Prevalence and risk factors for brucellosis seropositivity in cattle in Nyagatare District, Eastern Province, Rwanda. J. S. Afr. Vet. Assoc. 89, 1–8.

Selim, A., Attia, K., Ramadan, E., Hafez, Y.M., Salman, A., 2019. Seroprevalence and molecular characterization of Brucella species in naturally infected cattle and sheep. Prev. Vet. Med. 171, 104756.

Selim, A., Megahed, A., Kandeel, S., Alanazi, A.D., Almohammed, H.I., 2021. Determination of seroprevalence of contagious caprine pleuropneumonia and associated risk factors in goats and sheep using classification and regression tree. Animals 11, 1165.

Shirmohammadi-Khorram, N., Tapak, L., Hamidi, O., Maryanaji, Z., 2019. A comparison of three data mining time series models in prediction of monthly brucellosis surveillance data. Zoonoses Public Health 66, 759–772.

Tasiame, W., Emikpe, B., Folitse, R., Fofie, C., Burimuah, V., Johnson, S., Awuni, J., Afari, E., Yebuah, N., Wurapa, F., 2016. The prevalence of brucellosis in cattle and their handlers in North Tongu district of Volta region, Ghana. Afr. J. Infect. Dis. 10, 111–117.

Yakubu, A., Awuje, A., Omeje, J., 2015. Comparison of multivariate logistic regression and classification tree to assess factors influencing prevalence of abortion in Nigerian cattle breeds. J. Anim. Plant Sci. 25, 1520–1526.

Yu, W.L., Nielsen, K., 2010. Review of detection of Brucella spp. by polymerase chain reaction. Croat. Med. J. 51, 306–313.