

Application of Machine Learning for Predicting Brucellosis Disease in Dairy Cattle

S. M.A.K. Hussaini

Bangladesh Agricultural University

M. H. Tito

Bangabandhu Sheikh Mujibur Rahman Science and Technology University

M. M. Hasan

Bangladesh Agricultural University

Md. Siddiqur Rahman (✉ siddique.medicine@bau.edu.bd)

Bangladesh Agricultural University

S. T. Sharmy

Bangladesh Agricultural University

R. R. Sarker

Bangladesh Agricultural University

A. K. M. A. Rahman

Bangladesh Agricultural University

H. Neubauer

30IE Reference Laboratory for Brucellosis, Institute of Bacterial Infections and Zoonoses, Friedrich-Loeffler-Institute, Jena, Germany.

Article

Keywords: SMOTE, ROC Curve, MLP, J48 tree, Association, Brucellosis Prediction

Posted Date: November 1st, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3480234/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Brucellosis, a zoonotic disease caused by *Brucella* bacteria, has emerged as a significant concern in Bangladesh, affecting both animals and humans, with economic repercussions and health risks. This paper addresses the urgent need for effective disease management in a country where agriculture and livestock are pivotal to the economy. The study focuses on identifying risk factors for brucellosis in dairy cattle and their correlation with factors collected from Central Cattle Breeding and Dairy Farm (CCBDF) and Military Dairy Farms (MDF) in Bangladesh. Serum and milk samples, along with demographic data, were collected and analyzed. To address the challenge of imbalanced data, the study introduces the application of the "SMOTE" Weka filter for the first time in Bangladesh's veterinary profession. This balances the dataset, enhancing the accuracy of predictive modeling. Machine learning models, specifically Multilayer Perceptron (MLP) and J48 decision tree algorithms, are employed for brucellosis prediction. MLP achieved a Correct Classification (CC) rate of 95.0801%, while J48 achieved 94.0503%, demonstrating their efficacy. Sensitivity analysis identifies key attributes, with retained placenta being the most significant, aiding in prioritizing control measures. Association Rules reveal patterns and relationships among disease-related factors, enhancing understanding and proactive prevention.

INTRODUCTION

Brucellosis, a debilitating zoonotic disease caused by bacteria of the genus *Brucella*, has emerged as a significant concern in many parts of the world, including Bangladesh ¹. The disease poses a threat to both animal and human health, leading to economic losses, reduced productivity in livestock, and potential transmission to humans through direct contact or consumption of contaminated animal products ².

In the context of Bangladesh, where agriculture and livestock farming are integral to the economy and livelihoods of a substantial population, the prevalence of brucellosis in animals demands urgent attention ³. Traditional approaches to disease surveillance and control often fall short due to their inherent limitations in handling the complexities of large-scale data, diverse variables, and non-linear relationships ⁴. Herein lies the potential of machine learning to revolutionize disease prediction and management ⁵.

Machine learning, a subset of artificial intelligence, has demonstrated remarkable capabilities in predictive analytics and data-driven decision-making across various domains ⁶. In the context of disease prediction, machine learning models can harness the power of complex data relationships to forecast disease outbreaks based on diverse variables, such as climate patterns, animal demographics, and environmental factors ⁷. By leveraging historical data and identifying patterns that might elude traditional statistical methods, machine learning approaches offer a fresh perspective on tackling the challenges posed by brucellosis in Bangladesh ⁸.

The aim of this study is to identify the risk factors associated with brucellosis positivity in dairy cattle. Additionally, the study aims to establish a correlation between the factors collected from aborted dairy cows at the Central Cattle Breeding and Dairy Farm (CCBDF) and Military Dairy Farms (MDF) in Bangladesh.

MATERIALS AND METHODS

The research was carried out at both the Central Cattle Breeding and Dairy Farm, as well as the Military Farm. We collected serum samples from a total of 460 lactating cows, alongside their corresponding milk samples, in a random selection process. To gain insights into the demographics and risk factors at both the farm and animal levels, we administered a comprehensive questionnaire.

We collected approximately 5 to 10 ml of blood from each of the 460 lactating cows by carefully accessing the jugular vein. The blood-filled tubes were then stored upright at room temperature for one hour before being refrigerated at 4°C overnight. Subsequently, they go through centrifugation at 3,000 rpm for a duration of 10 minutes, which allowed for the separation of the serum samples. These separated serum samples were then carefully transferred into sterile Eppendorf tubes and preserved at an astonishingly low temperature of -20°C until the time of testing.

Simultaneously, we also obtained milk samples from the same group of cows, which were essential for conducting the milk ring test. In cases involving abortion or other reproductive complications, such as retained placenta, vaginal swabs were collected, or in instances of abortion, fetal samples were gathered, whenever such incidents were reported to the relevant authorities.

To ensure a comprehensive dataset, we administered a standardized questionnaire designed to capture crucial data pertaining to the history, production, and breeding records of the lactating cows. Additionally, we recorded pertinent information such as age, sex, breed, location, calving status, and any relevant disease history at the time of sample collection. Furthermore, for each sample, we meticulously documented details such as (a) history of abortion, (b) presence or absence of retained placenta, (c) history of repeat breeding, (d) age group categorization [≤ 6 years, ≥ 6 years], and (e) parity. The data was randomly divided into two sets: 80% for training and 20% for validation. The dataset exhibited an imbalance, and to rectify this issue, we generated a synthetic data to achieve balance. This was accomplished through the utilization of WEKA 3.8.6 (The Waikato Environment for Knowledge Analysis, 2022) software according to January 25, 2022 WEKA manual ⁹.

BALANCING OF DATA

Within the realm of veterinary medicine, the significance of a balanced dataset cannot be overstated, as it is integral to ensuring the accuracy and reliability of predictive modeling and diagnostic outcomes. Recent years have witnessed an increasing acknowledgment of the pivotal role that data-driven methodologies play in augmenting clinical decision-making and disease management for animal

patients⁹. Nonetheless, the efficacy of such approaches is intricately tied to the availability of well-balanced datasets, encompassing sufficient representation for each class or category of animals and their corresponding medical conditions¹⁰. In cases of skewed data distributions, the ensuing biases and suboptimal outcomes can significantly impede the applicability and generalizability of machine learning algorithms and predictive models (Fig. 1).

To address this, the current study introduces a methodology to rectify imbalanced datasets through the utilization of the SMOTE (Synthetic Minority Over-sampling Technique) Weka filter¹¹. Notably, this is the first time the SMOTE Weka filter has been applied to dataset balancing in Bangladesh's veterinary profession, up to our knowledge (Fig. 2). Adopting balanced datasets can improve therapy effectiveness, diagnostic accuracy, and overall animal healthcare practices¹⁰.

MACHINE LEARNING MODELS

In this paper we have utilized total of two different ML's to explore, explain and model the behavior of brucellosis disease with respect to several input parameters¹². The ML's are selected bases on their suitability and acceptability are given below:

Multilayer Perceptron

The Multilayer Perceptron (MLP) neural network can be used as an effective method for detecting and predicting (Table 1) brucellosis^{13,14}. Multilayer Perceptron (MLP) predicts brucellosis by utilizing historical data and analyzing various factors associated with the disease¹⁵. The MLP algorithm is trained on a dataset that contains numerous information. This paper focuses on the detection of brucellosis, utilizing the potent capabilities of MLP's adaptive learning mechanisms¹⁶. These mechanisms thrive on a diverse range of veterinary data, encompassing critical factors like Abortion, Calving, Age, Repeat Breeding, and Retained Placenta.

During the training process, the MLP algorithm is learned the complex patterns and relationships within the data. It adjusts the weights and biases of its interconnected layers to optimize its ability to predict brucellosis incidence accurately¹⁷. This training enables the MLP to recognize patterns and correlations between the input factors and the occurrence of brucellosis¹⁶.

After training the MLP, we have used it to make predictions on new, unseen data. By inputting the relevant factors into the trained MLP model, it has generated predictions on the likelihood of brucellosis outbreaks. The MLP takes into account the learned patterns and relationships to provide accurate predictions based on the input data.

The visual representation of the Multilayer Perceptron (MLP) how it works is shown in the (Figure 3) shortly.

Through this approach, we have the opportunity to achieve a heightened degree of precision in the classification of diseases, early detection, and prognosis of appropriate treatments. This transformative

integration of MLP-based methodologies not only enriches the analytical processes but also ushers in the potential to revolutionize the field by offering nuanced insights into the complex interplay of factors influencing animal health and reproduction ¹⁸. This study delves into the realm of harnessing MLP's computational prowess, it contributes significantly to the expanding landscape of AI-powered advancements in veterinary medicine, ultimately fostering re-abled methodologies for tackling critical reproductive challenges in animals ¹⁹.

Table 1
Generated from Multilayer Perceptron (MLP) Classifier

Model	Evaluation Criteria					
	<i>CC</i>	<i>RMSE</i>	<i>MAE</i>	<i>RAE</i>	<i>KS</i>	Processing <i>Time (s)</i>
Multilayer Perceptron	95.0801%	0.2075	0.0768	15.3701%	0.9017	0.79

In Table 1, by using Multilayer Perceptron (MLP) algorithm we have yielded promising results. The MLP algorithm demonstrated high accuracy with a Correct Classification (CC) of 95.0801%, as well as precise predictions indicated by low Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values (0.2075 and 0.0768, respectively). The Relative Absolute Error (RAE) of 15.3701% underscored its ability to provide accurate predictions across varying data scales. Additionally, the Kappa Statistics value of 0.9017 showcased the algorithm's robustness in handling complex veterinary datasets. Notably, the MLP algorithm's swift processing time of 0.79 seconds highlighted its efficiency. These findings emphasize the potential of MLP through Weka in enhancing animal care through improved diagnostics.

J48 Tree

J48, a powerful decision tree algorithm, finds numerous applications in the prediction of brucellosis ²⁰. By analyzing relevant factors, we have created a decision tree model which can be assisted in detecting and predicting the occurrence of this bacterial infection. With a dataset containing attributes and corresponding class labels indicating brucellosis cases, J48 recursively partitions the data based on attribute values, creating decision rules that accurately classify instances (Fig. 4). By traversing the decision tree and following the decision rules, J48 effectively predicts the presence or absence of brucellosis in new, unseen instances.

The accuracy of J48 in brucellosis prediction is contingent upon the quality of the dataset and the selection of relevant attributes ²⁰. With its ability to learn and classify based on patterns and relationships, J48 provides valuable insights for early detection and control of brucellosis, assisting in the prevention and management of this infectious disease ²¹.

Table 2
Generated from J48 Classifier

Model	Evaluation Criteria					
	<i>CC</i>	<i>RMSE</i>	<i>MAE</i>	<i>RAE</i>	<i>KS</i>	Processing <i>Time (s)</i>
J48	94.0503%	0.2322	0.096	19.2109%	0.8811	0

In the presented Table 2, the performance evaluation of a predictive model, specifically the J48 decision tree algorithm, is outlined. The model's predictive accuracy is assessed through various evaluation criteria. The CC column represents the Correct Classification Rate, indicating that the J48 model achieved an accuracy of 94.0503%, implying that nearly 94.05% of instances were correctly classified by the model. The "RMSE" (Root Mean Squared Error) value of 0.2322 reflects the average deviation between the predicted and actual values, with lower values indicating better predictive accuracy.

The MAE (Mean Absolute Error) of 0.096 represents the average magnitude of errors between predicted and actual values, providing insights into the model's precision. The "RAE" (Relative Absolute Error) of 19.2109% indicates the average percentage difference between predicted and actual values, which gives an understanding of the model's performance relative to the actual values. The reported Kappa Statistic value of 0.8811 suggests that there is a high level of agreement between the observed classifications and the classifications. It indicates that the model's predictions are significantly better than what would be expected by random chance.

Lastly, the Processing Time (s) column indicates the time taken by the J48 model to process the given dataset, shown as "0" seconds. This processing time is important to consider when assessing the model's feasibility for real-time or time-sensitive applications. Overall, this table provides a comprehensive overview of the J48 model's performance in terms of accuracy, error metrics, discrimination capability, and processing efficiency.

RESULT AND DISCUSSION

ROC CURVE

The ROC curve has been widely applied in the field of brucellosis for various purposes, including diagnostic test evaluation and the assessment of biomarkers²². We have utilized ROC curves to determine the optimal cutoff points for these tests, maximizing sensitivity and specificity. Additionally, ROC curve analysis also been employed to assess the discriminatory ability of clinical biomarkers in differentiating brucellosis cases from controls²³. The area under the ROC curve (AUC) is often calculated as a measure of the discriminatory power of the test or biomarker²². A higher AUC indicates better diagnostic accuracy, with values closer to 1 indicating excellent performance²⁴.

The area under the ROC curve (AUC) has been used as a measure of the diagnostic accuracy of different tests and biomarkers, aiding in the selection and validation of reliable diagnostic tools for brucellosis²². Overall, the ROC curve has proven to be a valuable tool in the field of brucellosis for evaluating diagnostic tests and identifying effective biomarkers for early detection and management of the disease²⁵.

The "yes" output on the ROC curve signifies accurate identification of brucellosis cases (true positives). It indicates effective detection of affected individuals, contributing to the true positive rate²⁶. A sharp upward shift of the "yes" curve reflects the model's strong discrimination between positives and non-cases, achieving high true positives while minimizing false positives (Fig. 5).

The "no" output signifies the predictive model correctly identifying instances as non-cases of brucellosis (true negatives). This indicates successful identification of individuals without brucellosis, boosting the true negative rate²⁶. The ROC curve's "no" class portrayal highlights the model's capacity to uphold a strong true negative rate and reduce false positives, enhancing overall accuracy in predicting non-cases (Fig. 6).

The "yes" and "no" outputs on the ROC curve play a crucial role in revealing the predictive model's performance in distinguishing brucellosis cases from non-cases. Their positions and shapes provide insights into the model's sensitivity and specificity, aiding researchers and clinicians in assessing the model's utility for brucellosis prediction.

SENSITIVITY ANALYSIS

Sensitivity analysis plays a crucial role in predicting and understanding the dynamics of brucellosis. By systematically varying input parameters and observing how they impact the model's output, sensitivity analysis helps assess the sensitivity of the prediction model to different variables²⁷. In the context of brucellosis prediction, sensitivity analysis allows to identify which factors have the most significant influence on disease occurrence and transmission and we have found that retained placenta is the most important factor in this regard. This insight helps in prioritizing control measures and interventions. Additionally, sensitivity analysis aids in quantifying the uncertainty associated with model predictions, providing valuable information for taking decision. By conducting sensitivity analysis, it is possible to improve the accuracy and reliability of brucellosis prediction models, enabling more effective prevention and control strategies in areas at risk of brucellosis outbreaks²⁷.

Table 3
Generated from Ranked Attributes

Attribute	Significance
Retained Placenta	0.046126
Repeat breeding	0.025818
Age	0.023025
Abortion	0.016488
Calving	0.001

In the Table 3 presented values represent the significance of each attribute in contributing to the classification outcomes. The numeric values next to the attribute names indicate their respective importance scores. The attribute retained placenta holds the highest importance with a score of 0.046126, suggesting that it significantly influences the prediction of this specific condition. Similarly, the attributes repeat breeding, age, abortion, and calving exhibit importance scores of 0.025818, 0.023025, 0.016488, and 0.001, respectively. These scores provide insights into the relative impact of each attribute on predicting the associated reproductive conditions. The analysis underscores the potential of these attributes, particularly retained placenta, in contributing to accurate classification models for identifying prevalent reproductive issues in dairy cattle.

Attributes are ranked according to their impact factors are followed by-

Retained Placenta > Repeat breeding > Age > Abortion > Calving.

ASSOCIATION RULES

Association rules are a valuable tool in predicting and understanding the dynamics of brucellosis. By analyzing patterns and relationships within a dataset, we have made meaningful associations between different variables related to the disease²⁸. From the association, we have developed some interventions and control strategies to mitigate the spread of the disease. Association rules also provide insights into the risk factors and potential transmission pathways of brucellosis, enabling proactive measures to be taken to prevent outbreaks. The application of association rules in brucellosis prediction enhances our understanding of the disease and aids in the development of effective preventive measures.

Association Rules offer a nuanced yet potent approach to predict brucellosis. By uncovering hidden data relationships, they unveil vital patterns missed by traditional methods. They take the form of "if-then" statements, using observed patterns to anticipate brucellosis. We have made the following rules by using association rules are-

Rule 1

Abortion = No Repeat breeding = No Retained Placenta = No ==> RBT = No < conf:(1) > lift:(1.97) lev:(0.15) [132] conv:(66.97)

This rule suggests that if an observation does not have instances of abortion, repeat breeding, and retained placenta, then it is highly likely (100% confidence) that the Rose Bengal Test result will be negative (No). The lift value of 1.97 indicates that this rule is 1.97 times more likely to occur compared to random chance. The support value of 132 indicates that this rule is applicable to 132 instances within the dataset. The conviction value of 66.97 suggests a strong dependency between the antecedent and the consequent.

Rule 2

Abortion = Yes Repeat breeding = Yes Retained Placenta = Yes ==> RBT = Yes < conf:(0.91) > lift:(1.84) lev:(0.07) [64] conv:(5.26)

This rule suggests that if an observation has instances of abortion, repeat breeding, and retained placenta, then there is a high likelihood (91% confidence) that the Rose Bengal Test result will be positive (Yes). The lift value of 1.84 indicates that this rule is 1.84 times more likely to occur compared to random chance. The support value of 64 indicates that this rule is applicable to 64 instances within the dataset. The conviction value of 5.26 suggests a moderate dependency between the antecedent and the consequent.

Rule 3

Age = 9 Yrs Calving = Four Abortion = Yes Repeat breeding = No Retained Placenta = Yes ==> RAPID KIT TEST = Yes < conf:(1) > lift:(1.98) lev:(0.08) [72] conv:(72.35)

This rule suggests that if an observation has an age of 9 years, fourth calving, and instances of abortion and retained placenta, with no instances of repeat breeding, then it is highly likely (100% confidence) that the RAPID KIT TEST result will be positive (Yes). The lift value of 1.98 indicates that this rule is 1.98 times more likely to occur compared to random chance. The support value of 72 indicates that this rule is applicable to 72 instances within the dataset. The conviction value of 72.35 suggests a strong dependency between the antecedent and the consequent.

Rule 4

Area = CCBDF Abortion = No Repeat breeding = No Retained Placenta = No ==> RAPID KIT TEST = No < conf:(0.99) > lift:(2.01) lev:(0.1) [87] conv:(44.39)

This rule suggests that if an observation is from the Central Cattle Breeding and Dairy Farm (CCBDF) area and does not have instances of abortion, repeat breeding, and retained placenta, then it is highly likely (99% confidence) that the RAPID KIT TEST result will be negative (No). The lift value of 2.01 indicates that this rule is 2.01 times more likely to occur compared to random chance. The support value of 87

indicates that this rule is applicable to 87 instances within the dataset. The conviction value of 44.39 suggests a moderate dependency between the antecedent and the consequent.

Rule 5

Calving = 2nd Abortion = No Repeat breeding = No Retained Placenta = No ==> MRT = No < conf:(0.99) > lift:(2.03) lev:(0.06) [50] conv:(25.83)

This rule suggests that if an observation is from the second calving, does not have instances of abortion, repeat breeding, and retained placenta, then it is highly likely (99% confidence) that the Milk Ring Test result will be negative (No). The lift value of 2.03 indicates that this rule is 2.03 times more likely to occur compared to random chance. The support value of 50 indicates that this rule is applicable to 50 instances within the dataset. The conviction value of 25.83 suggests a moderate dependency between the antecedent and the consequent.

Rule 6

Calving = 1st Abortion = Yes Repeat breeding = Yes Retained Placenta = Yes ==> MRT = Yes < conf:(0.99) > lift:(1.93) lev:(0.13) [119] conv:(30.53)

This rule suggests that if an observation is from the first calving, has instances of abortion, repeat breeding, and retained placenta, then it is highly likely (99% confidence) that the Milk Ring Test result will be positive (Yes). The lift value of 1.93 indicates that this rule is 1.93 times more likely to occur compared to random chance. The support value of 119 indicates that this rule is applicable to 119 instances within the dataset. The conviction value of 30.53 suggests a moderate dependency between the antecedent and the consequent.

CONCLUSION

In conclusion, the paper highlights the growing concern of brucellosis, a zoonotic disease, in Bangladesh, which poses significant threats to both animal and human health, as well as economic losses in the livestock industry.

The utilization of the SMOTE Weka filter to balance the dataset within the field of veterinary medicine in Bangladesh represents a pioneering endeavor.

The study also discusses the use of machine learning models, specifically the Multilayer Perceptron (MLP) and J48 decision tree, to predict and manage brucellosis in cattle. These models demonstrate promising results in terms of accuracy, precision, and efficiency. The MLP model achieved a Correct Classification (CC) rate of 95.0801%, while the J48 model achieved a CC rate of 94.0503%. These high accuracy rates suggest that machine learning can effectively contribute to disease prediction and management.

Furthermore, sensitivity analysis identified key attributes influencing brucellosis prediction, with retained placenta being the most significant factor. This information can help prioritize control measures and interventions in managing the disease. Additionally, the paper discusses the use of Association Rules to identify patterns and relationships between various factors related to brucellosis. These rules can aid in understanding risk factors and transmission pathways, enabling proactive measures for prevention.

In summary, this paper demonstrates the application of SMOTE, the potential of machine learning, sensitivity analysis, and Association Rules in improving the prediction and management of brucellosis in dairy cattle in Bangladesh. These methods offer valuable insights and tools for addressing this important public health and economic concern. Further research and application of these techniques may lead to more effective disease control strategies in the future.

Declarations

ACKNOWLEDGEMENTS

This research work was supported by the Bangladesh Agricultural University (BAURES) under the Project No. 2021/1019/BAU-Development of Treatment Strategy against Brucellosis In Dairy Cattle and LDDP, DLS, Ministry of Fisheries and Livestock under the Project No. 2022/07/LDDP- Transmission dynamics and abortion risk factors of brucellosis in large dairy herds in Bangladesh.

ETHICAL APPROVAL

The study protocol of Ethical statement was peer reviewed and approved by the Animal Welfare and Experimentation Ethics Committee of Bangladesh Agricultural University authority [Ethical approval number – AWEEC/BAU/2023(55)] and individual respective all personnel and informed written consent prior to the collection of blood. Animal research was approved by the Faculty of Veterinary Science of Bangladesh Agricultural University and concern Military authority of Bangladesh Army.

We confirm that all methods were performed in accordance with the relevant guidelines and regulations by including a statement in the methods <https://www.nature.com/srep/journal-policies/editorial-policies#experimental-subjects>. In accordance with the editorial policies (<https://www.nature.com/srep/journal-policies/editorial-policies#experimental-subjects>), we included a statement in our manuscript confirming the study is reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>).

FUNDING

This research work was supported by the Bangladesh Agricultural University (BAURES) under the Project No. 2021/1019/BAU, and the title was “Development of Treatment Strategy against Brucellosis In Dairy Cattle” and LDDP, DLS, Ministry of Fisheries and Livestock under the Project No. 2022/07/LDDP, and the title was Transmission dynamics and abortion risk factors of brucellosis in large dairy herds in Bangladesh.

AVAILABILITY OF DATA AND MATERIALS

All data are used here which collected from the Central Cattle Breeding and Dairy Farm (CCBDF) and Military Dairy Farms (MDF) in Bangladesh. Here is the link of dataset that used in this paper (https://drive.google.com/file/d/1BetmtCQsl9KA9sCj75bn-2j_KUR6Gk5n/view?usp=drive_link)

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests.

References

1. Ahasan, Md. S., Rahman, Md. S., Rahman, A. K. M. A. & Berkvens, D. Bovine and Caprine Brucellosis in Bangladesh: Bayesian evaluation of four serological tests, true prevalence, and associated risk factors in household animals. *Trop Anim Health Prod* 49, 1–11 (2017).
2. Pokharel, S., Shrestha, P. & Adhikari, B. Antimicrobial use in food animals and human health: time to implement 'One Health' approach. *Antimicrobial Resistance & Infection Control* 9, 181 (2020).
3. Tithy, N. *et al.* PREVALENCE AND ASSOCIATED RISK FACTORS OF BOVINE BRUCELLOSIS IN SMALLHOLDER DAIRY COWS OF MYMENSINGH DISTRICT IN BANGLADESH. *Journal of Veterinary Medical and One Health Research* 4, (2022).
4. Rahman, M. T. *et al.* Zoonotic Diseases: Etiology, Impact, and Control. *Microorganisms* 8, 1405 (2020).
5. Santangelo, O. E., Gentile, V., Pizzo, S., Giordano, D. & Cedrone, F. Machine Learning and Prediction of Infectious Diseases: A Systematic Review. *Machine Learning and Knowledge Extraction* 5, 175–198 (2023).
6. Sarker, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* 2, 160 (2021).
7. Fregoso-Aparicio, L., Noguez, J., Montesinos, L. & García-García, J. A. Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. *Diabetology & Metabolic Syndrome* 13, 148 (2021).
8. Dematheis, F. *et al.* Machine Learning Algorithms for Classification of MALDI-TOF MS Spectra from Phylogenetically Closely Related Species *Brucella melitensis*, *Brucella abortus* and *Brucella suis*. *Microorganisms* 10, 1658 (2022).
9. wekadocs/WekaManual.pdf · stable-3-8 · WEKA / weka · GitLab. GitLab <https://git.cms.waikato.ac.nz/weka/weka/-/blob/stable-3-8/wekadocs/WekaManual.pdf> (2022).
10. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *jair* 16, 321–357 (2002).
11. Waqar, M. *et al.* An Efficient SMOTE-Based Deep Learning Model for Heart Attack Prediction. *Scientific Programming* 2021, e6621622 (2021).

12. Elreedy, D., Atiya, A. F. & Kamalov, F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Mach Learn* (2023) doi:10.1007/s10994-022-06296-4.
13. Fenta, H. M., Zewotir, T. & Muluneh, E. K. A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones. *BMC Medical Informatics and Decision Making* 21, 291 (2021).
14. Bagheri, H. *et al.* Forecasting the monthly incidence rate of brucellosis in west of Iran using time series and data mining from 2010 to 2019. *PLOS ONE* 15, e0232910 (2020).
15. Memarian, H. & Balasundram, S. Comparison between Multi-Layer Perceptron and Radial Basis Function Networks for Sediment Load Estimation in a Tropical Watershed. *Journal of Water Resource and Protection* 04, 870–876 (2012).
16. Bokhare, A., Bhagat, A. & Bhalodia, R. Multi-layer Perceptron for Heart Failure Detection Using SMOTE Technique. *SN COMPUT. SCI.* 4, 182 (2023).
17. Bikku, T. Multi-layered deep learning perceptron approach for health risk prediction. *Journal of Big Data* 7, 50 (2020).
18. Sonawane, J. S. & Patil, D. R. Prediction of heart disease using multilayer perceptron neural network. in *International Conference on Information Communication and Embedded Systems (ICICES2014)* 1–6 (2014). doi:10.1109/ICICES.2014.7033860.
19. Ezanno, P. *et al.* Research perspectives on animal health in the era of artificial intelligence. *Veterinary Research* 52, 40 (2021).
20. Zhai, M. *et al.* Research on the predictive effect of a combined model of ARIMA and neural networks on human brucellosis in Shanxi Province, China: a time series predictive analysis. *BMC Infectious Diseases* 21, 280 (2021).
21. Shah, W. *et al.* A Machine-Learning-Based System for Prediction of Cardiovascular and Chronic Respiratory Diseases. *Journal of Healthcare Engineering* 2021, e2621655 (2021).
22. Luu, C., Nguyen, D.-D., Van Phong, T., Prakash, I. & Pham, B. T. Using Decision Tree J48 Based Machine Learning Algorithm for Flood Susceptibility Mapping: A Case Study in Quang Binh Province, Vietnam. in *CIGOS 2021, Emerging Technologies and Applications for Green Infrastructure* (eds. Ha-Minh, C., Tang, A. M., Bui, T. Q., Vu, X. H. & Huynh, D. V. K.) 1927–1935 (Springer Nature, 2022). doi:10.1007/978-981-16-7160-9_195.
23. Yang, Y. *et al.* Unravelling potential biomarkers for acute and chronic brucellosis through proteomic and bioinformatic approaches. *Frontiers in Cellular and Infection Microbiology* 13, (2023).
24. Fosgate, G. T. *et al.* Estimation of receiver-operating characteristic curves to determine accuracy of a competitive enzyme-linked immunosorbent assay for the serodiagnosis of *Brucella* infection in domestic water buffalo (*Bubalus bubalis*) and cattle. *American Journal of Veterinary Research* 64, 57–64 (2003).
25. Yao, M. *et al.* Evaluation of the Combined Use of Major Outer Membrane Proteins in the Serodiagnosis of Brucellosis. *Infect Drug Resist* 15, 4093–4100 (2022).

26. Janssens, A. C. J. W. & Martens, F. K. Reflection on modern methods: Revisiting the area under the ROC Curve. *International Journal of Epidemiology* 49, 1397–1403 (2020).
27. Hoo, Z. H., Candlish, J. & Teare, D. What is an ROC curve? *Emerg Med J* 34, 357–359 (2017).
28. Grange, Z. L. *et al.* Ranking the risk of animal-to-human spillover for newly discovered viruses. *Proceedings of the National Academy of Sciences* 118, e2002324118 (2021).
29. Khare, S. & Gupta, D. Association rule analysis in cardiovascular disease. in *2016 Second International Conference on Cognitive Computing and Information Processing (CCIP)* 1–6 (2016). doi:10.1109/CCIP.2016.7802881.

Figures

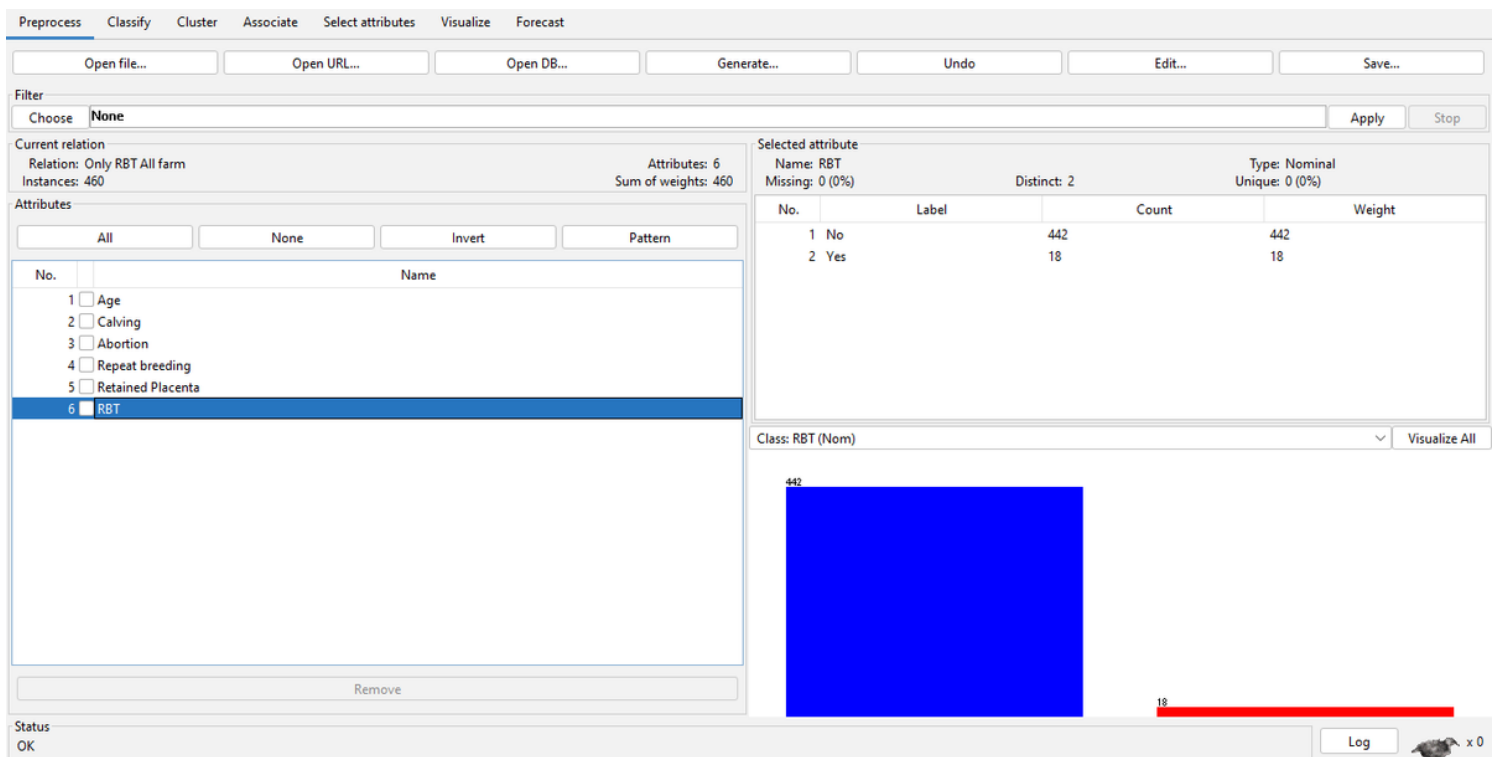


Figure 1

Disparities in Diagnosed Results between Positive and Negative Cases Prior to Employing SMOTE where 'No – 442' and 'Yes – 18' (Imbalanced Dataset)

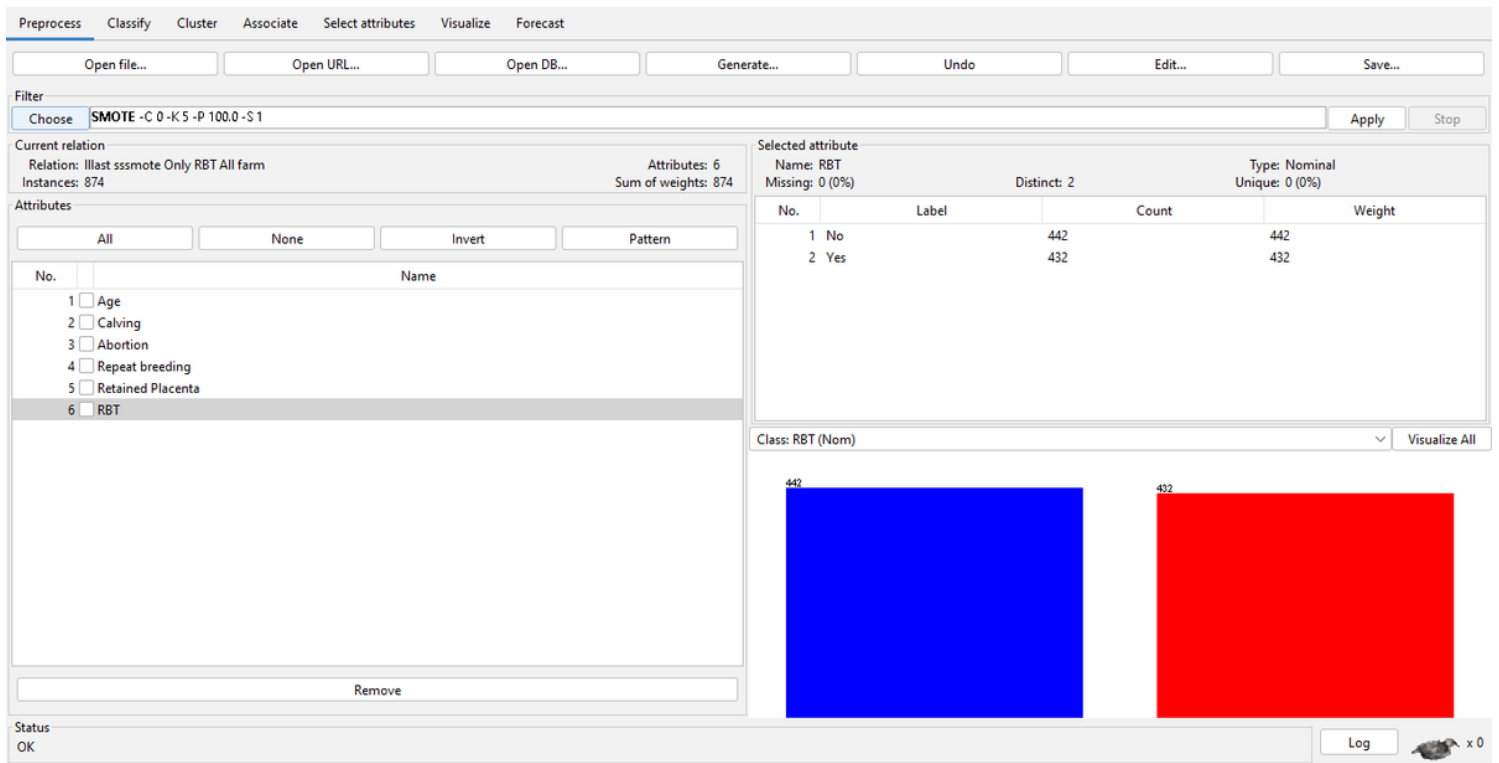


Figure 2

After applying SMOTE, the diagnosed positive and negative results achieve a nearly equal balance where 'No – 442' and 'Yes – 432'(Balanced dataset)

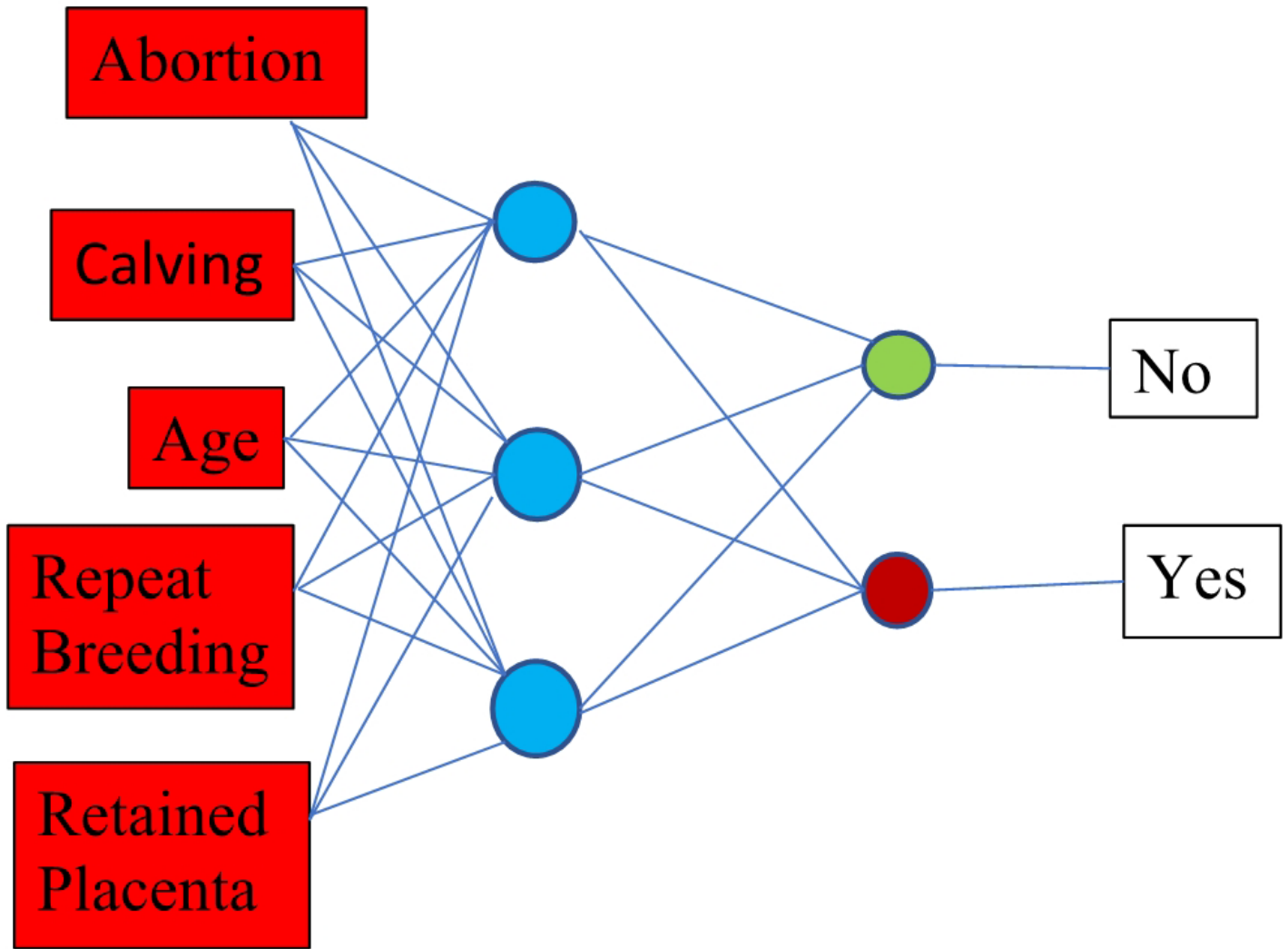


Figure 3

Operational Workflow of the Multilayer Perceptron (MLP)

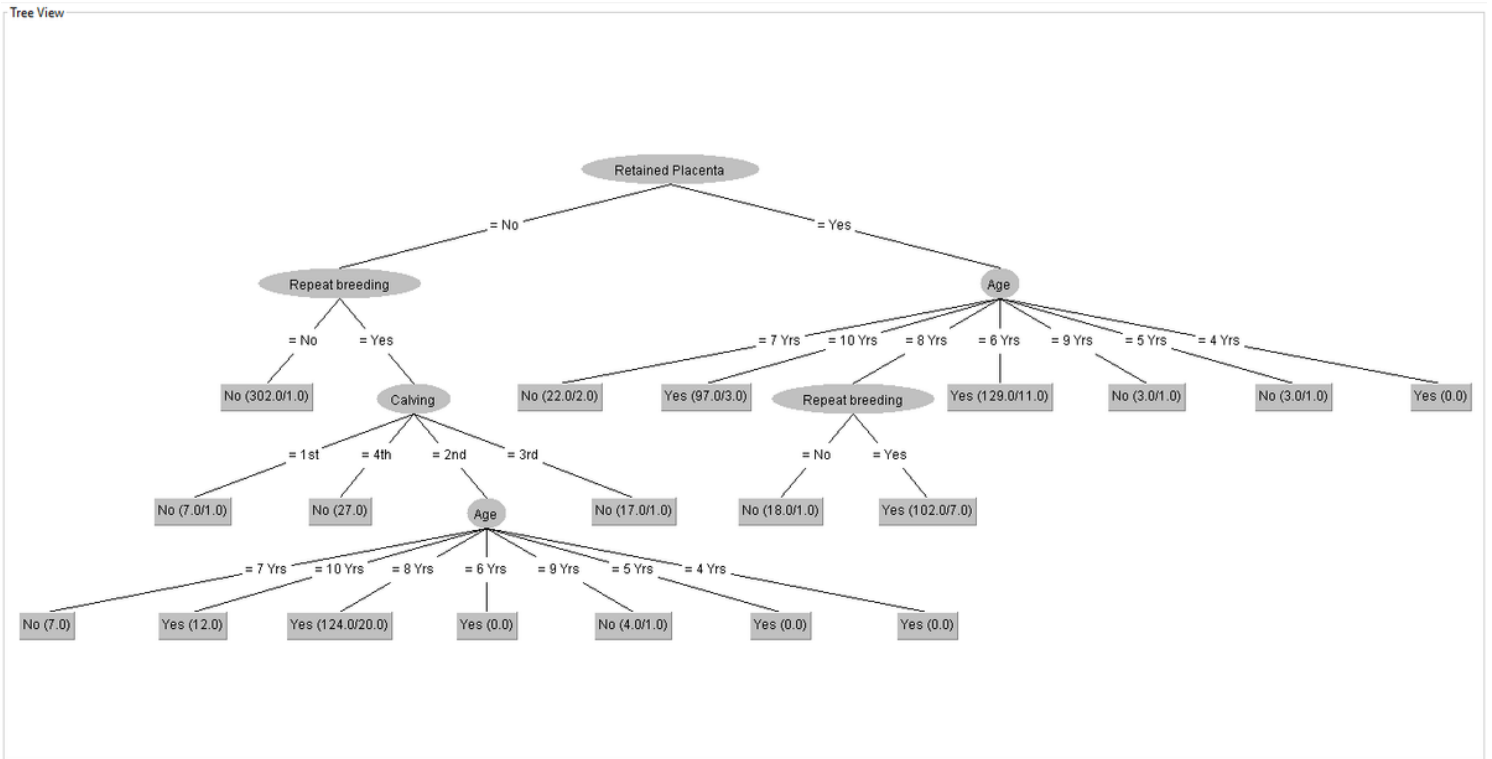


Figure 4

Decision Tree Diagram Generated by the J48 Algorithm

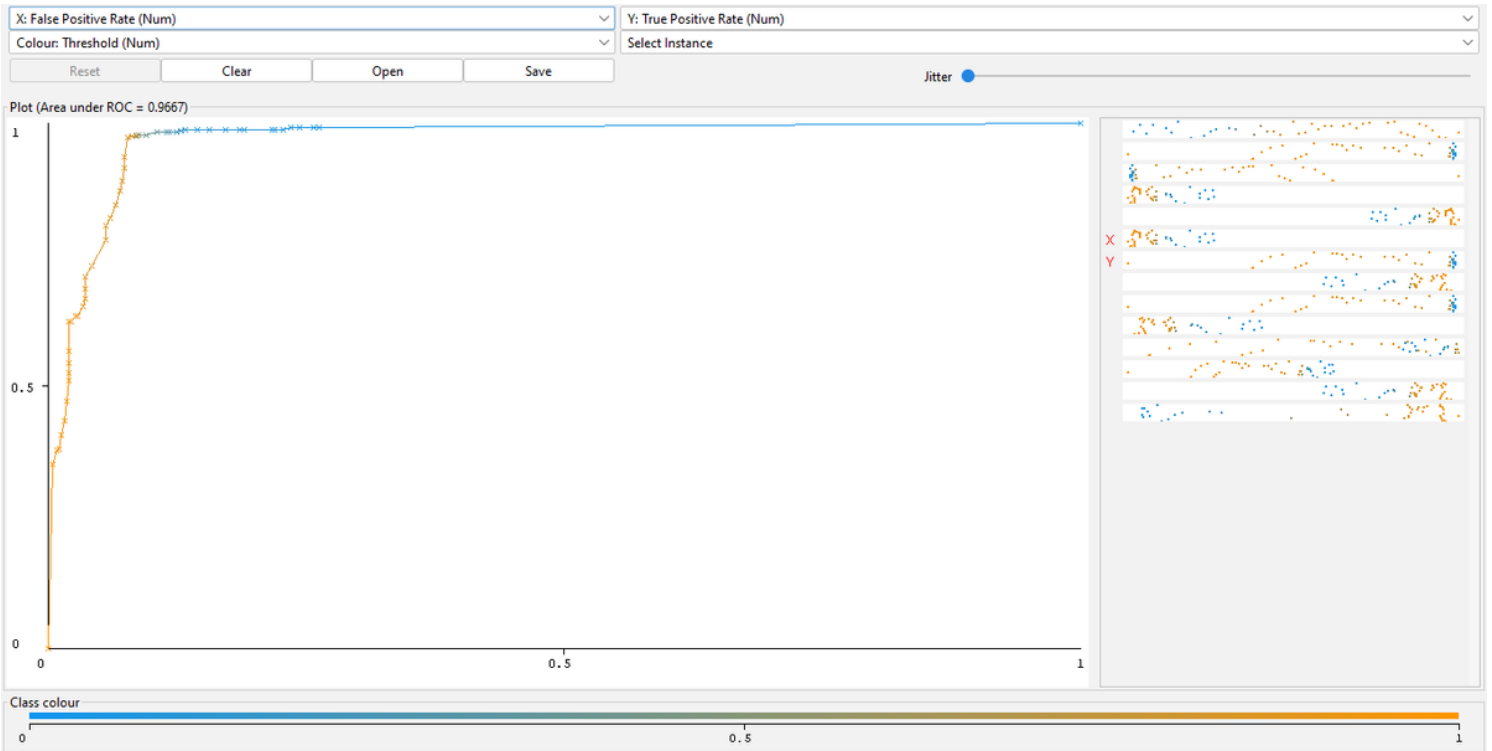


Figure 5

Receiver Operating Characteristic (ROC) Curve Performance Analysis (Yes)

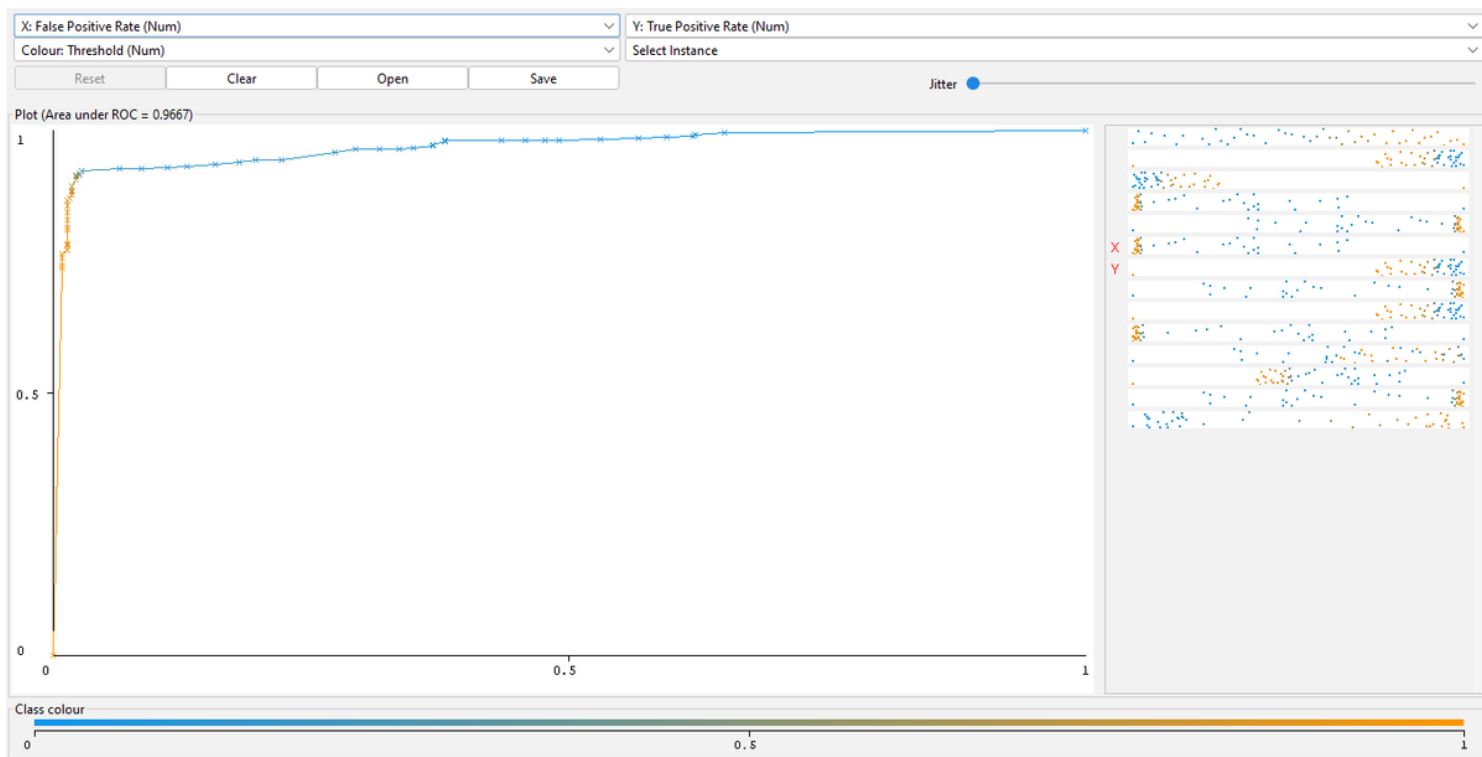


Figure 6

Receiver Operating Characteristic (ROC) Curve Performance Analysis (No)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Dataset.csv](#)