Date: 5-4-2025

The Editor Chief of the journal ‘*Theriogenology*’

Subject: Submission of manuscript.

Dear Editor,

We hereby submit the manuscript titled "RLST-KNN: An Efficient Machine Learning Method for Prediction of Subclinical Ketosis of Dairy Cows Based on Imbalanced Data Processing Algorithm" for review and potential publication in the journal ‘*Theriogenology*’.

The reasons for submitting this manuscript are as follows: This paper constructs an effective method for predicting subclinical ketosis in dairy cows based on imbalance data processing algorithms and improved missing value imputation methods.

Subclinical ketosis in dairy cows is a major metabolic disease that reduces milk production and causes economic losses. Developing an early prediction method using machine learning is critical to address these issues. However, challenges like missing value imputation and data imbalance remain. This paper proposes the RLST-KNN method to predict subclinical ketosis in dairy cows.

The RLST-KNN method utilizes the Random Forest and Local Outlier Factor algorithm to impute in missing values, applies the Synthetic Minority Over-sampling Technique with Tomek Links algorithm to achieve data balance, and finally uses the K-Nearest Neighbors algorithm to classify diseased and healthy cows. The experimental results show Accuracy, F1-score, sensitivity, positive predictive value, negative predictive value, and AUC scores of 0.7501, 0.7486, 0.8946, 0.6436, 0.8961, and 0.8727, respectively. This study advances subclinical ketosis prediction, offering practical value for dairy farms.

With best regards,

Sincerely Yours,

Corresponding author:

Qiang Dong([ardour@126.com](mailto:ardour@126.com))

Yan Feng([yanfeng52@126.com](mailto:yanfeng52@126.com))