# RLST-KNN: An Efficient Machine Learning Method for Prediction of Subclinical Ketosis of Dariy Cows Based on Imbalanced Data Processing Algorithm

# Shengquan Hu a, Zhao Zhang a, Zefeng Li a, Qiang Dong b\*, Yan Feng a\*

a College of Information Engineering, Northwest A&F University, Yangling, Shaanxi, 712100, China

b College of Veterinary Medicine, Northwest A&F University, Yangling, Shaanxi, 712100, China

**Abstract**

Subclinical ketosis in dairy cows is one of the most common and prominent metabolic diseases affecting dairy production. Subclinical ketosis in dairy cows can cause loss of appetite, metabolic issues, and reduced milk production, leading to malnutrition and economic losses for producers. To reduce losses on farms, the development of a ketosis early prediction method using machine learning algorithms has become a research hotspot in recent years. However, In the process of using machine learning algorithms to establish a ketosis early prediction method, the issue of the data imbalance affecting the performance of methods needs to be addressed. To solve the problem, the paper proposed RLST-KNN method to establish a dairy cow ketosis prediction method. This method firstly utilized the Random Forest-Local Outlier Factor (RF-LOF) algorithm for imputing missing values. Then, the RLST-KNN method applied the Synthetic Minority Over-sampling Technique with the Tomek Links (SMOTETomeklinks) algorithm to enhance minority class data and achieve data balance. Finally, it used K-Nearest Neighbors (KNN) to predict subclinical ketosis. To verify the predictive performance of the RLST-KNN method this article compared the performance differences in ketosis prediction of five classifiers: logistic regression (LR), linear discriminant analysis (LDA), K-nearest neighbors (KNN), support vector machine (SVM), and naive Bayes (NB), both before and after balancing the dataset. We found that KNN had the best performance among the five classifiers. The experimental results indicated that the RLST-KNN algorithm performs excellently in predicting subclinical ketosis in dairy cows, achieving accuracy (ACC), F1-score, sensitivity (Sens), positive predictive value (PPV), negative predictive value (NPV), and AUC scores of 0.7501, 0.7486, 0.8946, 0.6471, 0.6436, 0.8961, and 0.8727, respectively. In addition, the RLST-KNN method achieved the highest performance in the early lactation period (three weeks postpartum). It demonstrates that RLST-KNN can predict ketosis in dairy cows during the peak period of subclinical ketosis occurrence.

**Keywords:**Machine Learning; Dairy Cows; Prediction; Subclinical Ketosis; Imbalanced Data

# Author:

Shengquan Hu

College of Information Engineering, Northwest A&F University

3 Tai Cheng Road, Yangling Demonstration Zone

Xianyang, Shaanxi, China

Email: [hushengquan@nwafu.edu.cn](mailto:hushengquan@nwafu.edu.cn)

Zhao Zhang

College of Information Engineering, Northwest A&F University

3 Tai Cheng Road, Yangling Demonstration Zone

Xianyang, Shaanxi, China

Email: [2360036849@qq.com](mailto:2360036849@qq.com)

Zefeng Li

College of Information Engineering, Northwest A&F University

3 Tai Cheng Road, Yangling Demonstration Zone

Xianyang, Shaanxi, China

Email: [15648169267@163.com](mailto:15648169267@163.com)

# Corresponding Author:

Qiang Dong

College of Veterinary Medicine, Northwest A&F University

3 Tai Cheng Road, Yangling Demonstration Zone

Xianyang, Shaanxi, China

Email: [ardour@126.com](mailto:ardour@126.com)

Yan Feng

College of Information Engineering, Northwest A&F University

3 Tai Cheng Road, Yangling Demonstration Zone

Xianyang, Shaanxi, China

Email: [yanfeng52@126.com](mailto:yanfeng52@126.com)

**Running Title:** Dairy cow ketosis prediction method by machine learning algorithms

**Word Count:** 5498

**Table Count:** 11

**Figure Count:** 4

**Data Availability:** The data and code that support the findings of this study is available on the request from the corresponding author.

**Funding:** This study was funded by the Agricultural Science and Technology Innovation-driven Project in Shaanxi (NYKJ-2022-YL(XN)09) and Key Research and Development Program of Shaanxi(2019NY-082) of the People’s Republic of China.

**Acknowledgments:** We gratefully acknowledge Afimilk(China) Agricultural Technology CO., Ltd. for their assistance with the raw data.

**Conflicts of Interest:** The authors declare no conflict of interest.