Kocatepe Veterinary Journal

Kocatepe Vet J (2025) 18(3):227-235
DOI: 10.30607/kvj.1594704

RESEARCH ARTICLE

Explainable Machine Learning Framework for Milk Quality Grading

Bekir ÇETİNTAV[1]*, Ahmet YALÇIN[2]

[1]*Department of Biostatistics, Faculty of Veterinary Medicine, Burdur Mehmet Akif Ersoy University, Burdur, Türkiye*
[2]*Institute of Science, Burdur Mehmet Akif Ersoy University, Burdur, Türkiye*

ABSTRACT

This study introduces an explainable machine learning framework for milk quality grading, combining high predictive performance with transparency and practicality. Utilizing Random Forest and HistGradientBoost models, alongside interpretability techniques like Permutation Feature Importance and LIME, the framework achieves robust classification while providing actionable insights. Global explanations identify pH and Temperature as critical factors, highlighting their significance in real-time monitoring and microbial control. Local explanations, based on the two presented examples, demonstrate the practical utility of individual predictions, offering targeted interventions such as optimizing storage conditions or addressing contamination risks. By bridging the gap between predictive accuracy and interpretability, this framework not only enhances trust and usability for stakeholders but also establishes a new perspective for integrating AI-driven quality control systems into the dairy industry.
Keywords: Dairy Industry, Explainable AI (XAI), Machine learning, Milk quality, Veterinary Food Safety

***

Süt Kalitesi Derecelendirmesi için Açıklanabilir Makine Öğrenimi Çerçevesi

ÖZ

Bu çalışma, süt kalitesinin değerlendirilmesinde yüksek tahmin doğruluğunu şeffaflık ve kullanışlılık ile birleştiren açıklanabilir bir makine öğrenimi yaklaşımı sunmaktadır. Random Forest ve HistGradientBoost modellerinin yanı sıra Permutasyon Feature Importanee ve LIME gibi yorumlanabilirlik tekniklerini kullanan bu yaklaşım, güçlü bir sınıflandırma performansı sağlarken uygulanabilir içgörüler de sunmaktadır. Global yorumlanabilirlik sonuçları, pH ve Sıcaklık gibi kritik faktörleri belirleyerek gerçek zamanlı izleme ve mikrobiyal kontroldeki önemlerini vurgulamaktadır. Yerel yorumlanabilirlik sonuçları ise, sunulan 2 örnek üzerinden, bireysel tahminlerin pratik faydasını göstererek depolama koşullarının optimize edilmesi veya kontaminasyon risklerinin ele alınması gibi hedefe yönelik müdahalelere olanak tanımaktadır. Tahmin doğruluğu ile yorumlanabilirlik arasındaki boşluğu kapatan bu yaklaşım, yalnızca paydaşlar için güven ve kullanılabilirliği artırmakla kalmayıp, aynı zamanda AI destekli kalite kontrol sistemlerinin süt endüstrisine entegrasyonu için yeni bir perspektif sunmaktadır.
Anahtar kelimeler: Açıklanabilir Yapay Zeka (XAI), Makine öğrenimi, Süt endüstrisi, Süt kalitesi, Veteriner gıda güvenliği

To cite this article: Çetintav B. Yalçın A. Explainable Machine Learning Framework for Milk Quality Grading. Kocatepe Vet J (2025) 18(3):227-235

Submission: 02.12.2024 **Accepted**: 04.08.2025 **Published Online**: 05.09.2025
ORCID; *BÇ: 0000-0001-7251-1211, AY: 0009-0001-6093-8282*
*Corresponding author e-mail: bekircetintav@mehmetakif.edu.tr

INTRODUCTION

Milk, a cornerstone of human nutrition, is valued for its high nutritional content, including proteins, fats, lactose, vitamins, and minerals. Despite its importance, milk's perishable nature and susceptibility to adulteration present significant challenges in maintaining its quality throughout the supply chain (Frizzarin et al., 2021; Polat et al., 2021). These challenges necessitate the development of robust and reliable quality assessment methodologies to ensure consumer safety and product consistency.

Fourier-transform mid-infrared spectroscopy (MIRS) has emerged as a promising tool for analyzing milk composition, offering a non-destructive, rapid, and cost-effective method for assessing key parameters like fat, protein, and lactose (Frizzarin et al., 2021). Combined with machine learning techniques, MIRS enables high-precision predictions of technological properties such as coagulation time, paving the way for more efficient dairy processing (Polat et al., 2021). Other advanced methods, such as Raman and near-infrared spectroscopy, coupled with machine learning, have proven effective in detecting anomalies and contaminants in milk, ensuring compliance with safety standards.

Machine learning models, including random forests, support vector machines, and deep learning networks, have demonstrated exceptional capabilities in milk quality classification. These algorithms are employed to analyze parameters like pH, temperature, turbidity, and sensory attributes, achieving high classification accuracy (Bhavsar et al., 2023; Sheng et al., 2022). However, while predictive accuracy is essential, understanding the rationale behind predictions is equally critical.

Explainable artificial intelligence (XAI) addresses the need for transparency in machine learning models, particularly in safety-critical applications like food quality assessment. XAI techniques, such as Local Interpretable Model-agnostic Explanations (LIME) values, provide insights into model behavior, enabling stakeholders to trust and effectively utilize AI-driven decisions (Islam et al., 2022). For instance, XAI has been instrumental in food fraud detection, allowing users to interpret predictions and identify key factors influencing decisions (Buyuktepe et al., 2023). By improving model interpretability, XAI not only enhances trust but also supports error identification, bias reduction, and feature importance analysis (Przybył, 2024).

While machine learning models excel at predictive tasks, their black-box nature often obscures the underlying decision-making processes, which can lead to mistrust among stakeholders (Dang et al., 2022). XAI bridges this gap by offering local and global interpretability, making it possible to understand individual predictions and overall model behavior. This dual perspective ensures that models are not only accurate but also aligned with domain-specific requirements and ethical considerations (Islam et al., 2022; Przybył, 2024).

In this study, we propose a novel framework for milk grading that integrates machine learning and XAI techniques. Our approach classifies milk into three quality categories (Bad, Moderate, and Good) while employing interpretability tools to provide transparent insights into the model's decision-making process. By leveraging features such as pH, temperature, and sensory attributes, this framework aims to enhance the reliability and effectiveness of milk quality assessments.

MATERIAL and METHODS

Study Design and Ethical Statement

This study was designed to assess and interpret milk grading using machine learning and explainable artificial intelligence (XAI) methodologies. The dataset, obtained from Kaggle (https://www.kaggle.com/datasets/prudhvignv/milk-grading), includes key milk quality parameters such as pH, temperature, taste, odor, fat, turbidity, and color (GNV, 2020). The dataset is licensed under the "EU ODP Legal Notice," which allows for its use and redistribution under open data standards, ensuring compliance with ethical data practices. Preprocessing steps, including normalization and feature encoding, were applied to prepare the data for machine learning analysis. Models such as Random Forest and Gradient-Boosted Decision Trees were employed to classify milk quality into three grades: Bad, Moderate, and Good. Explainability was ensured using XAI techniques, including LIME for local interpretations and global feature importance, providing transparency into the decision-making processes of the models (Figure 1). As the dataset is publicly available, ethically sourced, and used in accordance with its licensing terms, no additional ethical approval was required for this study.

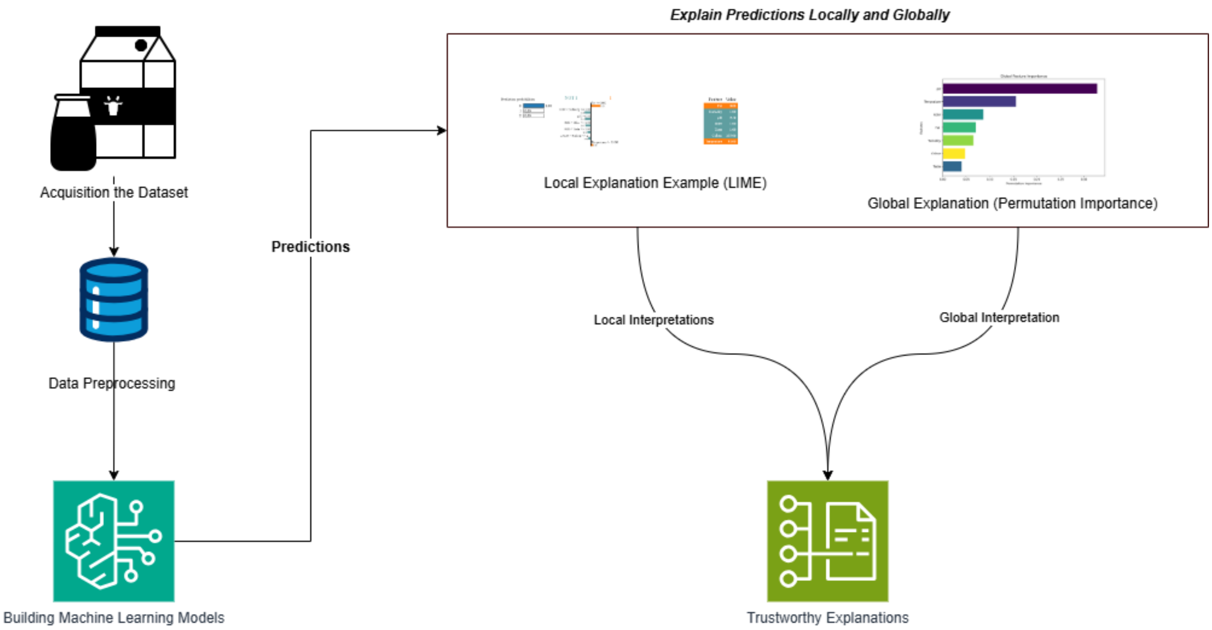

Figure 1: Framework for Generating Predictions and Interpreting Model Explanations

Dataset and Pre-processing

Several important features that are likely to affect the likelihood of a milk quality are included in this dataset (Table 1). Every milk has a distinct set of qualities that enable the development of a prediction model that pinpoints the variables affecting quality results. The dataset includes 1059 data and the following features:

Table 1. Milk Grade Dataset Feature Descriptions

Feature	Description
Taste	Taste of milk (satisfies optimal conditions assign 0 otherwise 1)
Odor	Smell of milk (satisfies optimal conditions assign 0 otherwise 1)
Fat	Fat of milk (satisfies optimal conditions assign 0 otherwise 1)
Turbidity	Turbidity of milk (satisfies optimal conditions assign 0 otherwise 1)
Color	The color of the milk (GreyScale)
pH	pH value of milk
Temperature	Milk temperature Immediately After Milking
Grade	0 (Bad quality), 1 (Moderate quality), 2 (Good quality).

Feature engineering was performed to prepare the dataset for effective machine learning modeling. Since there were no missing values in the dataset, no imputation techniques were required. Continuous variables, such as pH and temperature, were normalized using Min-Max scaling to transform their values into the range [0, 1], ensuring comparability across features. The binary categorical variables (Taste, Odor, Fat, and Turbidity) were retained in their original format (0 or 1) for seamless compatibility with machine learning algorithms. As the dataset was balanced across the target classes, no data augmentation techniques were applied, allowing the analysis to proceed directly with the available data.

Model Development and Evaluation

We examined the effectiveness of various machine learning models on the dataset in order to predict the likelihood of milk grade. We made sure that each of the 6 machine learning models we chose had a distinct mathematical foundation. Additionally, we experimented with decision tree techniques based on various architectures.

Several machine learning models were employed to predict milk grading, leveraging their diverse strengths in handling structured data. Random Forest is an ensemble learning method based on decision trees, known for its robustness against overfitting and its ability to handle non-linear relationships. Sample academic studies using this method (Bovo et al. 2021; Vishnu & Kumar, 2024). Support Vector Machine

(SVM) was utilized for its capacity to find an optimal hyperplane that separates classes, making it particularly effective in high-dimensional feature spaces. Sample academic studies using this method (Mu et al. 2020; Mammadova & Keskin, 2013).K-Nearest Neighbors (KNN), a simple yet powerful algorithm, classified data points based on the majority class of their nearest neighbors, relying on distance metrics. Sample academic studies using this method (Samad et al. 2024; Neware, 2023). XGBoost and LightGBM are gradient boosting algorithms that excel in speed and efficiency, with XGBoost offering regularization techniques to reduce overfitting and LightGBM being optimized for large datasets with high-dimensional features. Sample academic studies using these methods (Mota et al., 2022; Satola & Satola, 2024). HistGradientBoost is another gradient boosting approach that employs histogram-based techniques to accelerate training and enhance performance. Sample academic studies using this method (Ebrahimiet al., 2019; Sheng et al., 2022). By incorporating these diverse models, the study ensured comprehensive exploration of predictive capabilities, selecting the best-performing model based on evaluation metrics such as accuracy, precision, and recall.

To evaluate the performance of each machine learning model, we utilized four key metrics: accuracy, recall, precision, and F1-score. Accuracy (Eq. 1) measures the proportion of correctly classified instances out of the total instances, providing an overall measure of model performance. Recall (Eq. 2) evaluates the model’s ability to identify all relevant instances by measuring the proportion of correctly predicted positive cases to all actual positive cases, which is crucial when minimizing false negatives.

$$\text{Accuracy}=\frac{\text{True Positives}+\text{True Negatives}}{\text{Total Instances}} \quad \textbf{(1)}$$

$$\text{Recall}=\frac{\text{True Positives}}{\text{True Positives}+\text{False Negatives}} \quad \textbf{(2)}$$

$$\text{Precision}=\frac{\text{True Positives}}{\text{True Positives}+\text{False Positives}} \quad \textbf{(3)}$$

$$\text{F1 Score}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad \textbf{(4)}$$

Precision (Eq. 3) focuses on the quality of positive predictions, calculating the proportion of true positives to all predicted positives, which helps reduce false positives. **F1-score** (Eq. 4) combines precision and recall into a single metric by taking their harmonic mean, offering a balanced measure when there is a trade-off between these two metrics. Together, these metrics provide a comprehensive assessment of the models’ predictive performance and their ability to grade milk quality accurately and reliably.

Explainable Machine Learning

To provide transparency and interpretability to the predictive modeling of milk grading, both local and global explainability techniques were employed. Local explanations were implemented using LIME (Local Interpretable Model-Agnostic Explanations), which interprets individual predictions by approximating the complex model locally with an interpretable surrogate (Ribeiro et al., 2016). This method identifies the contribution of each feature, such as pH or Temperature, to the classification of a specific milk sample as "good," "moderate," or "bad." These insights are particularly valuable for validating predictions and identifying specific factors influencing individual outcomes.

For a broader understanding, global explanations were provided using Permutation Feature Importance. This technique evaluates the overall significance of each feature by quantifying the drop in model performance when the values of a particular feature are permuted. Unlike traditional feature importance methods that can be biased towards features with more categories, permutation importance provides a corrected measure of feature relevance by considering randomized scenarios, thereby improving reliability and interpretability.

The integration of local and global explainability creates a comprehensive framework, addressing both instance-specific and overall model behavior. This approach corrects biases inherent in traditional feature importance metrics, ensuring that variables critical to the prediction process are accurately ranked (Altmann et al., 2010). Such a dual-layered approach is indispensable in high-stakes applications like milk grading, as it ensures not only the transparency of specific decisions but also the identification of the most influential factors shaping the model’s predictions. This enhanced interpretability makes the predictive system a trustworthy and actionable tool for stakeholders.

Global Interpretability with Permutation Importance

Permutation Feature Importance is a model-agnostic technique used to assess the global importance of features in a predictive model. The method involves randomly shuffling the values of a single feature and observing the decrease in the model's performance. The rationale is that if a feature is important, shuffling its values will disrupt the model's predictions, leading to a significant drop in performance. This approach is particularly advantageous as it accounts for non-linear interactions and does not assume any specific model structure.

The importance of a feature X_i is calculated as the difference between the baseline performance of the model (without permutation) and the performance after permuting X_i. Mathematically, the permutation feature importance $I(X_i)$ can be expressed as:

$$I(X_i)=E_{\pi\sim Perm(Xi)}\left[M\left(D_{\pi(Xi)}\right)\right]-M(D) \quad \textbf{(5)}$$

where:

- D: Original dataset.
- $D_{\pi(Xi)}$: Dataset where feature X_i is permuted (shuffled).
- $M(D)$:Model performance metric (e.g., accuracy, F1-score) on the original dataset.
- $M(D_{\pi(Xi)})$: Model performance metric on the dataset with X_i permuted.
- E: Expectation operator to average over multiple permutations π.

This formula captures how much the performance metric changes when the feature's information is disrupted, quantifying its importance in the model's predictive power.

Local Interpretability with LIME

LIME is a technique used to explain individual predictions of complex machine learning models by creating an interpretable local surrogate model around the instance being explained. The idea is to approximate the behavior of the black-box model f in the vicinity of a specific instance x by training a simpler, interpretable model g such as a linear regression or decision tree. This enables understanding of how features contribute to the prediction for x.

The core of LIME involves generating perturbed samples around x and observing the corresponding predictions from the black-box model. The surrogate model g is then fitted to this locally weighted dataset, providing a locally interpretable explanation. Mathematically, LIME optimizes the following objective to train the surrogate model g :

$$\text{argmin}_{g\in G}\ L(f,g,\pi_x)+\Omega(g) \quad \textbf{(6)}$$

where:

- f: The black-box model being explained.
- g: The interpretable surrogate model.
- G: The set of all possible interpretable models.
- $L(f,g,\pi_x)$: The loss function measuring how well g approximates f in the locality of x
- π_x: A locality kernel defining the weight of perturbed samples based on their distance to x.
- $\Omega(g)$: A complexity penalty to ensure g remains interpretable.

LIME's power lies in providing interpretable insights (e.g., feature importance weights) for a single prediction while maintaining flexibility across various model types and datasets.

(Equation 5 is taken from Altmann et al., (2010) and Equation 6 is taken from the main study of Ribeiro et al., (2016).)

RESULTS

Model Performance

The data set was first preprocessed to be suitable for machine learning models and divided 80 by 20 for train/test procedures. The performance of the models is summarised in Table 2. Based on the key evaluation metrics, several models showed strong predictive capabilities for the prediction task.

Table 2. Prediction performance of the models

Models	Accuracy	Recall	Precision	F1
Random Forest	0.995	0.993	0.994	0.994
HistGradientBoost	0.993	0.991	0.992	0.991
KNN	0.985	0.986	0.986	0.986
XGBoost	0.973	0.971	0.972	0.972
LGBM	0.970	0.972	0.973	0.972
SVM	0.566	0.564	0.468	0.489

The Random Forest model demonstrated the highest performance across all metrics, achieving an Accuracy of 99.5%, a Recall of 99.3%, a Precision of 99.4%, and an F1-score of 99.4%. This indicates that Random Forest not only accurately predicts the quality of milk but also maintains an excellent balance between correctly identifying positive cases (recall) and minimizing false positives (precision). Similarly, the HistGradientBoost model showed exceptional results with an Accuracy of 99.3% and an F1-score of 99.1%, proving to be a robust alternative with slightly lower metrics than Random Forest. While KNN, XGBoost and LGBM provide reasonable performance, SVM failed to meet the required predictive quality.

Global Interpretability

The global feature importance analysis, as depicted in the permutation importance graph, provides a comprehensive understanding of the factors influencing the milk grading predictions. The Permutation Feature Importance Graph was used to interpret the influence of the features on the prediction results over the entire dataset (Figure 2). This global interpretation provides insight into how each feature influences the model's milk grade probability predictions. The values on the x-axis are permutation importance scores. The y-axis of the graph contains the features. They are ordered from highest influence to lowest influence from top to bottom.

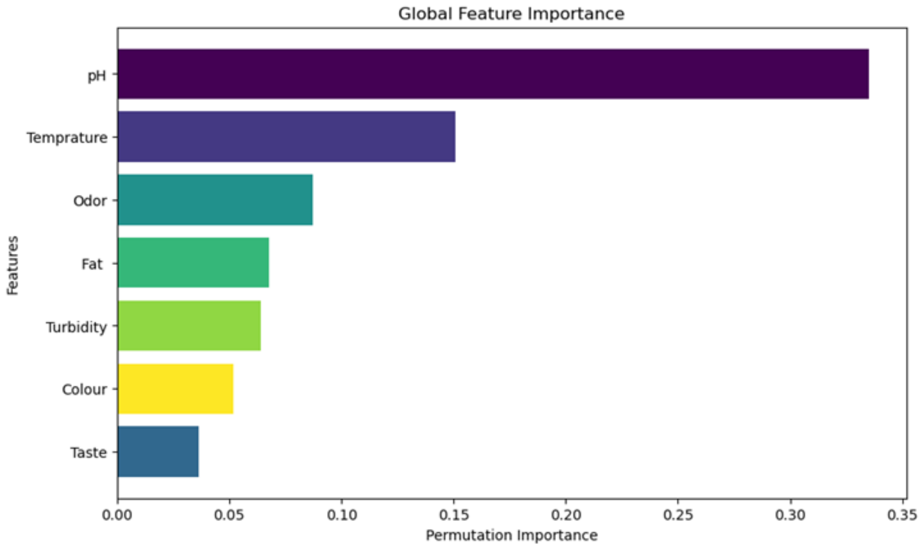

Figure 2: Permutation Importance Graph for Global Explanations

Among the seven independent variables, pH was identified as the most significant feature, showcasing the highest contribution to the model's performance. This result underscores the critical role of pH in determining milk quality, as it directly impacts freshness and spoilage dynamics.

Temperature ranked as the second most important feature, highlighting its pivotal role in maintaining milk quality. The significance of temperature aligns with its known effects on microbial activity and shelf life, making it a key parameter in milk grading.

Sensory attributes such as Odor, Fat, and Turbidity demonstrated moderate importance, reflecting their contributions to both the physical and sensory characteristics of milk. These features play an essential role in assessing milk quality from a consumer and nutritional perspective. Meanwhile, Color and Taste, while contributing to the predictions, were found to have relatively lower importance compared to pH and Temperature, suggesting their secondary role in the milk grading process.

Local Interpretability

Two samples were selected from the data set for local interpretation. Care was taken to ensure that the target variables of the samples had a 'Bad' rating. It will be seen which features of the samples with a 'Bad' rating affect how. In this context, LIME Explanation Graph was used. With this graph, it is possible to obtain information such as Prediction Probabilities, Feature Values, Feature Importance.

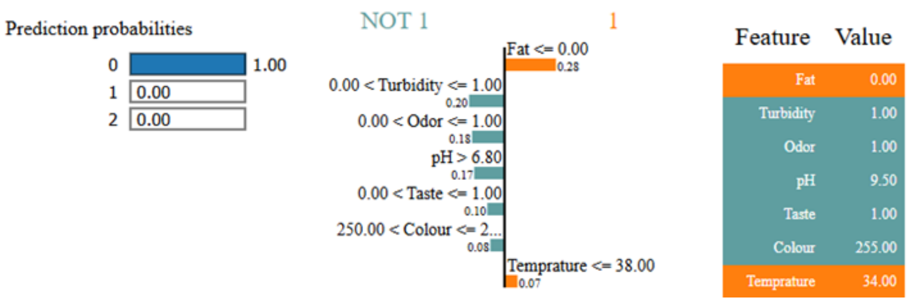

Figure 3: LIME Explanation Graph of Sample 1 (Predicted as "Bad")

The sample in Figure 3 classified as "Bad quality" (Grade 0) was primarily influenced by high turbidity (1.00) and elevated pH (9.50), which deviated significantly from optimal conditions, contributing 20% and 17% to the classification, respectively. Odor (1.00) and taste (1.00) also played significant roles, with contributions of 18% and 11%, indicating suboptimal sensory characteristics. While the fat content (0.00) met optimal conditions, contributing 29% toward the "Not Bad" or "Good" classification, it was insufficient to offset the negative impacts of the other features. The temperature (34.00), though within the normal range immediately after milking (35°C–37°C). Overall, the classification was driven by multiple factors, with turbidity, pH, and sensory attributes playing dominant roles in determining the milk's poor quality.

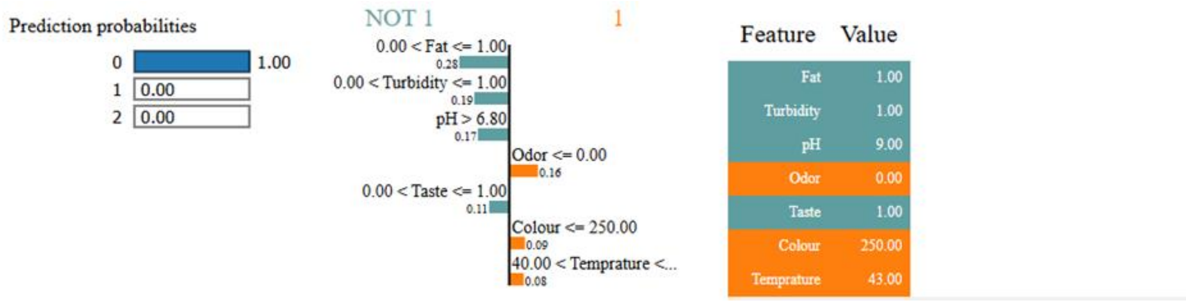

Figure 4: LIME Explanation Graph of Sample 2 (Predicted as "Bad quality")

The second sample (in Figure 4), also classified as "Bad quality" (Grade 0), was primarily influenced by high turbidity (1.00) and elevated pH (9.00), contributing 19% and 17% to the classification, respectively. While the odor (0.00) met optimal conditions, it contributed only 16% toward opposing the "Bad" classification, outweighed by other suboptimal features. Taste (1.00) and fat content (1.00) were suboptimal, with contributions of 11% and 28%, reinforcing the "Bad quality" classification. The temperature (43.00), exceeding but close to the ideal range for milk storage, further contributed 8%, emphasizing the need for rapid cooling after milking. Despite the color (250.00) being close to an acceptable range, it added 9% to the "not bad" classifications. Overall, the combination of turbidity, pH, and temperature, along with sensory attributes, strongly influenced the sample's poor quality classification.

DISCUSSION

This study presents a comprehensive framework for milk quality grading by integrating machine learning models with explainable artificial intelligence (XAI) techniques. The results demonstrate that the Random Forest and HistGradientBoost models outperformed other machine learning algorithms in accuracy, precision, recall, and F1-score. This finding aligns with prior studies, where ensemble learning methods have shown superior predictive capabilities in milk quality assessment (Bhavsar et al., 2023; Sheng et al., 2022). However, the poor performance of SVM highlights the need for algorithm selection tailored to dataset characteristics, such as feature scale and dimensionality (Mu et al., 2020).

The high predictive accuracy ensures reliable milk grading, but accuracy alone is insufficient in safety-critical applications like dairy production. This study addresses the "black-box" nature of machine learning models by integrating XAI techniques, bridging the gap between model performance and interpretability (Islam et al., 2022; Przybył, 2024).

Global explanations using Permutation Feature Importance identified pH and Temperature as the most critical features influencing milk quality. This aligns with scientific evidence highlighting the importance of chemical stability and microbial control in milk (Frizzarin et al., 2021; Polat et al., 2021). For instance, pH values outside the optimal range can indicate spoilage or adulteration, while temperature management is critical for maintaining freshness. These findings underscore the need for real-time monitoring systems that prioritize these parameters.

The moderate contributions of sensory attributes such as Odor, Fat, and Turbidity suggest their complementary role in grading. In practical applications, these features could be integrated into automated quality control systems to supplement chemical and physical measurements. The lower importance of Color and Taste may reflect their subjective nature and the challenge of quantifying these parameters reliably.

The local interpretability results, derived from LIME, provide granular insights into individual predictions, which are invaluable in field settings. For example:

In Case 1, the sample was classified as "Bad" primarily due to high Turbidity and elevated pH. These factors can indicate microbial activity or contamination during storage. For a dairy farm, this insight could prompt a review of milk handling processes or sanitation protocols.

In Case 2, the "Bad" classification was influenced by high temperature and suboptimal Fat content. In this scenario, the results suggest a failure in cooling systems or dietary imbalances in the cattle. These actionable insights enable targeted interventions, such as adjusting cooling equipment or modifying feed compositions.

The integration of global and local interpretability techniques makes this framework highly applicable to field operations. Dairy producers and quality control managers can leverage these insights to address specific issues, such as optimizing storage conditions, detecting contamination, and ensuring compliance with quality standards. Moreover, the transparent nature of these methods fosters trust among consumers and regulators, who require clarity in quality assessment processes (Buyuktepe et al., 2023).

Future studies should explore the scalability of this framework for larger datasets and diverse milk sources. Additionally, combining machine learning with real-time sensors, such as Fourier-transform mid-infrared spectroscopy (Frizzarin et al., 2021) or Raman spectroscopy, could enhance the precision and speed of quality assessments. Expanding the feature set to include advanced biochemical markers may further improve model performance and interpretability.

CONCLUSION

In conclusion, this study highlights the transformative potential of explainable machine learning in milk quality grading. By combining high predictive performance with interpretability, the proposed framework addresses the practical needs of the dairy industry, offering a robust and transparent tool for quality assessment. The insights gained from both global and local explanations not only improve operational efficiency but also pave the way for more trustworthy AI applications in food safety and quality monitoring.

Conflict of interest: The authors have no conflicts of interest to report.

Authors' Contributions: Conceptualization, BC., methodology, BC and AY.; software, AY; validation, BC and AY.; formal analysis, BC.; investigation, BC and AY.; resources, BC.; data curation, BC.; writing—original draft preparation, BC and AY.; writing—review and editing, BC.; visualization, AY.; supervision, BC. All authors have read and agreed to the published version of the manuscript.

Ethical approval: This study is not subject to the permission of HADYEK in accordance with the "Regulation on Working Procedures and Principles of Animal Experiments Ethics Committees" 8 (k). The data, information and documents presented in this article were obtained within the framework of academic and ethical rules. The dataset, obtained from open source database (Kaggle), available at: (https://www.kaggle.com/datasets/prudhvignv/milk-grading). The dataset is licensed under the "EU ODP Legal Notice," which allows for its use and redistribution under open data standards, ensuring compliance with ethical data practices. As the dataset is publicly available, ethically sourced, and used in accordance with its licensing terms, no additional ethical approval was required for this study.

Acknowledgement: We extend our gratitude to the esteemed researchers who made this dataset openly available.

REFERENCES

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. Bioinformatics, 26(10), 1340-1347. https://doi.org/10.1093/bioinformatics/btq134

Bhavsar, D., Jobanputra, Y., & Swain, N. K. (2023). Milk quality prediction using machine learning. EAI Endorsed Transactions on Internet of Things. https://doi.org/10.4108/eetiot.4501

Bovo, M., Agrusti, M., Benni, S., Torreggiani, D., & Tassinari, P. (2021). Random forest modelling of milk yield of dairy cows under heat stress conditions. Animals, 11(5), 1305. https://doi.org/10.3390/ani11051305

Buyuktepe, O., Catal, C., Kar, G., Bouzembrak, Y., Marvin, H., & Gavai, A. (2023). Food fraud detection using explainable artificial intelligence. Expert Systems. https://doi.org/10.1111/exsy.13387

Dang, M., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: A comprehensive review. Artificial Intelligence Review, 55, 3503–3568. https://doi.org/10.1007/s10462-021-10088-y

Ebrahimi, M., Mohammadi-Dehcheshmeh, M., Ebrahimie, E., & Petrovski, K. R. (2019). Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. Computers in Biology and Medicine, 114, 103456. https://doi.org/10.1016/j.compbiomed.2019.103456

Frizzarin, M., Gormley, I. C., Berry, D. P., & Murphy, T. B. (2021). Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. Journal of Dairy Science, 104(7), 7438–7447. https://doi.org/10.3168/jds.2020-19576

GNV, P. (2020). Milk Grading (Classification) (Version 1) [Dataset; Kaggle]. Kaggle. https://www.kaggle.com/datasets/prudhvignv/milk-grading/data

Islam, M. R., Ahmed, M. U., Barua, S., & Begum, S. (2022). A systematic review of explainable artificial intelligence in terms of different application domains and tasks. Applied Sciences, 12(1353). https://doi.org/10.3390/app12031353

Mammadova, N., & Keskin, İ. (2013). Application of the support vector machine to predict subclinical mastitis in dairy cattle. The Scientific World Journal, 2013(1), 603897. https://doi.org/10.1155/2013/603897

Mota, L. F., Giannuzzi, D., Bisutti, V., Pegolo, S., Trevisi, E., Schiavon, S., ... & Cecchinato, A. (2022). Real-time milk analysis integrated with stacking ensemble learning as a tool for the daily prediction of cheese-making traits in Holstein cattle. Journal of Dairy Science, 105(5), 4237-4255. https://doi.org/10.3168/jds.2021-21426

Mu, F., Gu, Y., Zhang, J., & Zhang, L. (2020). Milk source identification and milk quality estimation using an electronic nose and machine learning techniques. Sensors, 20(15), 4238. https://doi.org/10.3390/s20154238

Neware, S. (2023). Cow Milk Quality Grading using Machine Learning Methods. International Journal of Next-Generation Computing, 14(1).

Polat, O., Akçok, S. G., & Akbay, M. A. (2021). Classification of raw cow milk using information fusion framework. Journal of Food Measurement and Characterization, 15, 5113–5130. https://doi.org/10.1007/s11694-021-01076-5

Przybył, K. (2024). Explainable AI: Machine learning interpretation in blackcurrant powders. Sensors, 24(3198). https://doi.org/10.3390/s24103198

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

Samad, A., Taze, S., & Ucar, M. (2024). Enhancing milk quality detection with machine learning: A comparative analysis of KNN and distance-weighted KNN algorithms. International Journal of Innovative Science and Research Technology, 9(3). https://doi.org/10.38124/ijisrt/IJISRT24MAR2123

Satoła, A., & Satoła, K. (2024). Performance comparison of machine learning models used for predicting subclinical mastitis in dairy cows: Bagging, boosting, stacking, and super-learner ensembles versus single machine learning models. Journal of Dairy Science, 107(6), 3959-3972. https://doi.org/10.3168/jds.2023-24243

Sheng, T., Shi, S., Zhu, Y., Chen, D., & Liu, S. (2022). Analysis of protein and fat in milk using multiwavelength gradient-boosted regression tree. IEEE Transactions on Instrumentation and Measurement, 71, 2507810. https://doi.org/10.1109/TIM.2022.3165298

Vishnu, K., & Kumar, S. P. (2024). Enhancing the quality of milk prediction using random forest algorithm and logistic regression algorithm. AIP Conference Proceedings, 3193(1). https://doi.org/10.1063/5.0238130