Journal of Technology and System Information Vol: 3, No 1, 2026, Page: 1-16

Analysis of Different Sensor Data Using Machine Learning Methods for the Purpose of Determining Milk Quality

Sinan Sevinc[1]*, Yavuz Selim Taspinar[2]

[1]Merlab Innovation,
[2]Department of Mechatronic Engineering,Technology Faculty

DOI:
https://doi.org/10.47134/jtsi.v3i1.5367
*Correspondence: Full Name
Email: email@e-mail.com

Received: 23-11-2025
Accepted: 19-12-2025
Published: 21-01-2026

Abstract: Milk is a product with high nutritional value, but its quality may vary depending on factors from production to consumption. Milk is a food that can spoil over time and carries a disease risk due to microorganism growth. Therefore, continuous monitoring of milk quality is important. Quality loss can cause changes in milk components such as protein, fat, and lactose. In recent years, sensors have been used to evaluate milk quality by quickly measuring parameters such as chemical components, pH value, temperature, and fat content. These sensor data provide information not only about milk quality but also about the productivity and health of cows. This enables more efficient production processes and early detection of potential diseases. Sensor measurements help determine both milk quality and cow care needs. In this study, quality classification was performed using data from 1059 different milk samples. The dataset consists of 7 features and 1 class feature, and milk quality was classified into three classes: “high”, “medium”, and “low”. kNN (k-Nearest Neighbor), ANN (Artificial Neural Network), DT (Decision Tree), and RF (Random Forest) methods were used for classification. Model performance was evaluated using confusion matrix, accuracy, precision, recall, and F1 score, and detailed analysis was performed using the ROC curve. The kNN model achieved 99.8% accuracy, the ANN model 99.9%, the DT model 99.4%, and the RF model 100%. The RF model showed the highest success. Overall, the classification performances of all models were close to each other, and all can be used to determine milk quality.

Keywords: milk quality, sensor data, classification, performance analysis, machine learning

Introduction

Rapid advances in technology have also significantly transformed the fields of agriculture and animal husbandry. With the widespread adoption of sensors, data collection, and automation systems, milk production and quality control can now be monitored in a more systematic manner. This is crucial for both increasing production efficiency and maintaining the quality of the milk produced. Nowadays, herd management systems are used on many farms, and these systems provide continuous data on the milk yield, health status, and behavior of animals. While this data is analyzed for purposes such as early detection of diseases or productivity estimation, determining milk quality is also becoming increasingly important. Milk is one of the staple foods in human nutrition and has high nutritional value due to its fat, protein, and mineral content. However, these values can vary depending on the animal's diet, health status, environmental conditions, and care processes. Therefore, regular monitoring of milk quality is essential for both producer sustainability and consumer health. In this context, data analysis and machine learning-based methods enable the numerical classification of quality using physical and chemical data obtained from milk samples. Parameters such as pH, temperature, taste, odor, fat content, turbidity, and color are considered fundamental indicators in determining the overall quality of milk. In recent years, the use of machine learning-based models has significantly increased the accuracy rates in the quality control and classification processes of food products. These methods reduce human-induced evaluation errors by automatically extracting meaningful attributes from the obtained data, making the analysis process faster and more reliable. Numerous studies have been conducted on milk quality in literature. Some of these studies are listed below in order.

Saumya Kumari et al. aimed to identify the key factors affecting milk quality in this study. Within this scope, nine different parameters were analyzed, including temperature, fat content, pH, taste, odor, color, turbidity, and overall quality. According to the Principal Component Analysis (PCA) results, color and temperature were found to be the variables with the most significant effect on milk quality. The PCA-1 and PCA-2 components explained more than 95% of the total variance. Furthermore, milk samples were classified into three categories: low, medium, and high quality, and an artificial neural network (ANN)-based machine learning model was used in the classification process. The developed ANN model demonstrated high performance with an accuracy rate of 0.9988 and produced more consistent and accurate results compared to other methods. The study stated that classification success could be improved in the future by using different and more advanced machine learning or deep learning approaches (Kumari et al., 2023).

In the study by Ravi Kumar et al., the aim was to predict milk quality using prediction models that consider factors such as fat and protein content, somatic cell count, and bacterial load. In this research, a model using neural networks was proposed to predict milk quality based on characteristics such as color, fat content, and pH. Through feature selection and parameter optimization, the model was designed to achieve an accuracy rate of 0.95. However, considering that relying solely on the accuracy value does not sufficiently reflect the model's performance, metrics such as F-score, sensitivity, and recall were also evaluated. Furthermore, robustness validation studies using the bootstrap method demonstrated that the model provides reliable and robust results, avoiding overfitting and underfitting. This comprehensive approach enhances the model's reliability and performance in predicting milk quality (R. R. Kumar et al., 2024).

Ahmet Çelik aimed to classify milk quality using machine learning algorithms in this study. The "Milk Quality" dataset, obtained from the Kaggle data repository and containing 1059 samples, was used. Milk samples were classified as low, medium, and high quality using seven different features. During the classification phase, the commonly used ANN and Adaptive Boosting (AdaBoost) algorithms were implemented, and the Python-based open-source Orange platform was preferred. As a result of the application, the AdaBoost algorithm classified with 99.9% accuracy, while ANN classified with 95.4% accuracy. The findings indicate that the AdaBoost algorithm is more successful in determining milk quality (Çelik, 2022).

S. Sunithamani et al. present an innovative system in this study that integrates machine learning and sensor technology to address the adulteration problem commonly encountered in dairy products. Data was collected from eight different milk samples using sensors measuring parameters such as pH, turbidity, and temperature, along with an Arduino Uno-based hardware infrastructure. The collected data was analyzed using machine learning algorithms such as Support Vector Machine (SVM), Adaboost, and RF, successfully predicting milk quality. In contrast to the inadequacy of traditional methods in detecting adulteration at low concentrations, this system offers a more sensitive and reliable solution. Additionally, the web-based distribution of the system enables real-time monitoring and evaluation (Sunithamani et al., 2024).

In this study, Boyu Ji et al. proposed a machine learning–based framework for estimating milk yield, milk composition (fat and protein), and individual milking frequency using long-term data from Robotic Milking Systems (RMS). The framework was developed using behavioral, health, and production data collected from 80 cows over a period of five years. The models were designed to forecast the production performance of each cow up to 28 days ahead. To reflect real farm conditions, time series cross-validation was employed, and model performance was systematically assessed. The results showed that the models achieved a coefficient of determination (R^2) greater than 0.90 and an overall accuracy exceeding 80%. The findings indicate that the proposed framework can support improved management practices and animal welfare in robotic dairy farming systems (Ji et al., 2022).

In this study, Mu et al. developed a low-cost, non-destructive electronic nose system to estimate fat and protein ratios, which are important indicators of milk quality, and to determine the source of the milk (farm). This system consists of seven metal oxide semiconductor sensors. In classifying the milk source, odor data obtained from the E-nose device and composition data provided by DHI (Dairy Herd Improvement) analyses were evaluated both separately and together. The obtained data were processed using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) methods to reduce the feature dimension; then, classification models were created using Logistic Regression, Support Vector Machines (SVM), and Random Forest (RF) algorithms. The SVM model trained with the combined dataset after LDA yielded the most successful result with a 95% accuracy rate. In addition, GBDT, XGBoost, and RF algorithms were applied to estimate milk fat and protein ratios; the highest accuracy in terms of prediction performance was achieved with the RF method. The results of the study demonstrate that the developed approach provides a robust technical infrastructure for the rapid and reliable prediction of milk quality (Mu et al., 2020).

In the study by Narendra Kumar S. et al., a machine learning-based framework was proposed for evaluating milk quality. In this study, the RF algorithm was used to analyze six key parameters including pH, temperature, color, taste, odor, and turbidity. The obtained parameters were used to classify milk into three classes: low, medium, and high quality, and real-time predictions were provided through a user-friendly web application. The proposed system offers a high accuracy rate and provides a fast and cost-effective alternative to traditional laboratory methods; in this respect, it stands out as a valuable tool for milk producers and quality control professionals. The system guarantees that only high-quality milk reaches the consumer by making advanced quality assessments widely accessible. Future work plans include expanding the dataset, using more advanced machine learning techniques, and addressing issues such as IoT integration for continuous real-time monitoring, demonstrating the potential of machine learning to transform applications in the dairy industry (N. Kumar et al., 2024).

In their study, Bekir Çetinav and Ahmet Yalçın proposed an explainable machine learning-based framework for classifying milk quality. In this framework, RF and Hist Gradient Boost models were used in conjunction with interpretability techniques such as Permutation Feature Importance and LIME (Local Interpretable Model-agnostic Explanations) to achieve high accuracy and explainability. Global explanations revealed that pH and temperature values are critical factors in terms of milk quality and microbial control, while local explanations demonstrated the practical benefit of individual predictions and supported targeted interventions such as optimizing storage conditions or reducing contamination risks. This approach bridges the gap between prediction accuracy and interpretability, increasing trust and usability for stakeholders and offering a new perspective for integrating AI-supported quality control systems into the dairy industry (Çetintav & Yalçın, 2025).

Radu Neamt et al., aimed to investigate how milk manufacture and chemical formulation during the first 100 days of breastfeeding changed under the influence of electrical conductivity, number of births (parities) and days in milk (DIM). In the study conducted on 66 Romanian Alaca cows, electrical conductivity values were found to be significantly higher in primiparous (first calf) cows. Electrical conductivity was higher during the first 30 days of lactation and decreased over time. Multiple regression analysis showed that birth month and DIM affected conductivity, while parity did not. A weak negative correlation was found between electrical conductivity and milk fat and protein, and a weak positive correlation was found with milk yield. No significant relationship was observed with lactose. Milk production and chemical composition were significantly affected by birth month, DIM, and parity factors, but no direct effect of electrical conductivity on these components was detected (Neamț et al., 2016).

Compared to studies in literature, the original contributions of this research can be listed as follows:

1. By addressing milk quality in a multi-class structure as ‘low’, ‘medium’, and ‘high’, it has provided a more sensitive classification for the different product requirements (cheese, yogurt, etc.) of the food industry.
2. The performance of RF, ANN, DT, and kNN models was systematically compared on the same dataset (1059 samples). This comparison clearly revealed the balance between accuracy (RF and ANN) and interpretability (DT) in milk quality prediction, as well as the general performance level of kNN.
3. It offers a low-cost decision support model that quickly and automatically predicts quality, as opposed to time-consuming and costly laboratory analyses. This model allows producers to intervene more quickly in factors affecting quality (e.g., feed ration).

4. The study establishes a foundation and benchmark for sensor-based, real-time quality monitoring systems that can be developed in the future with the interpretable models it proposes.

The structure of the paper is organized as follows. Section two describes the dataset employed in this study, the applied methodologies, and the evaluation metrics. Section three reports the experimental findings and provides a discussion of the results. Section four outlines the main conclusions of the study, along with its contributions, limitations, and possible application areas.

Methodology

Milk Quality Prediction by Classification

The Milk Quality dataset used in this study was obtained from the open-source Kaggle platform (Shrijayan). Milk samples have 7 features, including pH, Temperature, Taste, Odor, Fat, Turbidity, and Color. In general, milk quality is evaluated based on these features. The objective is to classify milk samples as Low, Medium, and High quality. The Taste, Odor, Fat, and Turbidity features take values of 0 or 1, while the Temperature, pH, and Color features are actual color and numerical values. Table 1 shows the values of 15 milk samples randomly selected from the dataset.

Table 1. *Sample data from the dataset*

pH	Temprature	Taste	Odor	Fat	Turbidity	Colour	Grade
6,60	35	1	0	1	0	254	high
6,60	36	0	1	0	1	253	high
6,60	37	1	1	1	1	255	high
6,50	38	1	1	1	1	255	high
6,80	40	1	1	1	1	255	high
6,60	38	0	0	0	0	255	medium
6,50	36	0	0	0	0	247	medium
6,50	37	0	0	0	0	255	medium
6,70	45	1	1	0	0	247	medium
6,70	45	1	1	1	0	245	medium
5,60	50	0	1	1	1	255	low
7,40	90	1	0	1	1	255	low
6,80	50	0	0	1	0	255	low
6,70	50	1	1	1	0	245	low
8,60	55	0	1	0	0	255	Low

k Nearest Neighbor (kNN)

The k-Nearest Neighbor algorithm is a method that classifies new data by looking at the K nearest examples. The basic logic is to calculate the distance between a new example and all examples in the training set, then determine the K nearest neighbors. The class of the new sample is determined based on the classes to which these K neighbors belong. If all neighbors belong to the same class, the new sample is directly assigned to that class. Otherwise, a score is calculated for each class, and the new sample is assigned to the class with the highest score. In other words, the kNN algorithm starts with the test example X and expands its neighborhood, continuing this expansion until it encompasses K training examples. Then, the most frequently occurring class among these K neighbors is determined, and the test example X is assigned to that class (Wang, 2019).

Artifical Neural network (ANN)

ANN, a logical programming method inspired by the functioning mechanism of the human brain, aims to mimic the brain's basic biological processes through specific software models. ANN is an algorithm that can perform operations, make decisions, produce results, and make predictions using existing information in the face of missing data, similar to the human brain (Civalek, 1998). It is also defined as a computing system that can continuously receive data, learn from this data, and store past information in its memory. ANN is modeled after the working principles of biological neural networks and represents a simplified version of this structure. The fundamental characteristics of these systems are their ability to perform parallel processing, their learning capabilities, and their adaptable operation with distributed memory structures (KELEŞOĞLU et al., 2005; Özbay, 1999). In general, ANNs consist of three main layers. The input layer, one or more hidden layers, and the output layer. Each layer contains a specific number of neurons or nodes, and these neurons are connected to each other through weights. Inputs are carried as signals transmitted between neurons through these weights. Each neuron processes the signals it receives according to the weight values and produces an output, which can be transferred to other neurons (Çınar et al., 2009; Lorenz et al., 1997; Ronco & Fernandez, 1999).

Decision Tree (DT)

Decision Tree is a method used in supervised machine learning for both classification and regression problems. This algorithm creates a hierarchical structure by sequentially splitting the dataset based on specific features. In the resulting tree structure, internal nodes represent decision points, while leaf nodes represent the final class label or predicted value. The main goal of DT is to predict the target variable through simple and understandable decision rules obtained from the training data (Chauhan, 2020). This model is created using only training data during the training process. Decision nodes define non-leaf nodes, while leaf nodes contain class information (Eesa et al., 2015a; 2015b). The Decision Tree algorithm has the ability to work with both numerical and categorical data, and non-linear relationships between variables do not negatively affect model performance (Tian et al., 2019). Furthermore, the data preprocessing requirement is quite low. However, if the tree is overgrown, the model may show a tendency to memorize and overfitting problems may arise (Josephine et al., 2021).

Random Forest (RF)

Random Forest (RF) is an ensemble-based machine learning algorithm that uses multiple decision trees together and can exhibit nearest neighbor-like prediction behavior in some cases. The fundamental principle of ensemble approaches is to combine multiple weak learners to create a stronger and more generalizable model instead of relying on a single model. The RF algorithm implements this approach through decision trees. In this model, each tree splits the input data into subsets starting from the root node using different splits, and this process is performed independently. Random Forest produces the final prediction by combining the outputs of many trees, thus yielding more stable results compared to a single decision tree. The prominent advantages of the RF algorithm include fast processing time, robustness against noisy data, and the ability to work effectively with missing data (Lakshmi et al., 2019; VijiyaKumar et al., 2019).

Confusion Matrix

The confusion matrix is a widely used analysis tool in supervised learning methods to evaluate classification accuracy. It displays the relationship between the classes predicted by the model and the actual classes in tabular form. Each column represents the model's predictions for a specific category, while each row represents the number of actual examples belonging to that category. The matrix consists of true positive (TP), false negative (FN), false positive (FP), and true negative (TN) values. These values show which classes the model predicted correctly or incorrectly, allowing for a detailed examination of classification performance (Cui et al., 2024; Taşpınar & Selek, 2020). An example confusion matrix with three classes is shown in Table 2.

Table 2. *Confusion Matrix*

ACTUAL CLASS

	C_1	C_2	C_3
C_1	T_1	F_{12}	F_{13}
C_2	F_{21}	T_2	F_{23}
C_3	F_{31}	F_{32}	T_3

PREDICTED CLASS

Performance Metrics

The most used metrics for evaluating classification problems are accuracy, precision, recall, and F1 score. Accuracy refers to the ratio of correctly classified examples to the total number of examples. This metric indicates the overall classification success of the model. Precision indicates how many of the examples predicted as positive by the model are actually positive. A high precision value indicates that the model has a low false positive rate and therefore produces more reliable results (Alipio & Bures, 2024; Ferdowsy et al., 2021; Love, 2024). Sensitivity indicates the rate at which the model correctly predicts true positive examples. A high sensitivity value indicates that the model detects the positive class more effectively. This means that the model has a high success rate in capturing true positive cases (Farabi et al., 2024).

The F1 score is a metric that evaluates classification performance by balancing precision and sensitivity. This score is calculated by taking the harmonic means of precision and sensitivity and measures the model's ability to correctly identify positive examples and avoid false positives in a balanced manner (Yogi et al., 2024).

The performance parameters evaluated for multi-class classification are defined as follows. For a class C_i, classifier performance can be evaluated using T_{pi}, F_{ni}, T_{ni} and F_{pi}, and these values can be calculated from the number of test examples belonging to class C_i . The quality of overall classification performance can be evaluated in two different ways: micro and macro averages. The macro average evaluates all classes equally, while the micro average gives more weight to classes with more data. Various performance measures suitable for multi-class classification problems can be calculated as follows; this is a generalization of the parameters presented in Table 2 for multiple classes C_i (Ji et al., 2022). The numbers T_{pi}, F_{ni}, T_{ni} and F_{pi} are given for a class C_i respectively. The micro and macro average indices are represented by μ and M, respectively (Dinesh & Dash, 2016).

$$Accuarcy = \frac{\sum_{i=1}^{m} \frac{Tpi + Fni}{Tpi + Fni + Fpi + Tni}}{m}, \quad (1)$$

Here, m is the class number.
Other important measurements can be obtained from Equation 2 to 9 (Sokolova & Lapalme, 2009).

$$Precision_{\mu} = \frac{\sum_{i=1}^{m} Tpi}{\sum_{i=1}^{m} (Tpi + Fpi)} \quad (2)$$

$$Precision_{M} = \frac{\sum_{i=1}^{m} \frac{Tpi}{Tpi + Fpi}}{m} \quad (3)$$

$$Specificity_{\mu} = \frac{\sum_{i=1}^{m} Tni}{\sum_{i=1}^{m} (Fpi + Tni)} \quad (4)$$

$$Specificity_{M} = \frac{\sum_{i=1}^{m} \frac{Tpi}{Tpi + Fni}}{m} \quad (5)$$

$$Sensitivity_{\mu}\,(recall_{\mu}) = \frac{\sum_{i=1}^{m} Tpi}{\sum_{i=1}^{m} (Tpi + Fni)} \quad (6)$$

$$Sensitivity_{M}\,(recall_{M}) = \frac{\sum_{i=1}^{m} \frac{Tpi}{Tpi + Fni}}{m} \quad (7)$$

$$F1 - score_{\mu} = \frac{(\beta^2 + 1) Precision_{\mu} Recall_{\mu}}{\beta^2 Precision_{\mu} Recall_{\mu}} \quad (8)$$

$$F1 - score_{M} = \frac{(\beta^2 + 1) Precision_{M} Recall_{M}}{\beta^2 Precision_{M} Recall_{M}} \quad (9)$$

Cross Validation

Cross-validation is a statistical evaluation approach used to measure the generalizability of classification models. In this method, the dataset is structured to allow the model to be both trained and tested on different subsets of the data. Thus, model performance is evaluated without being constrained to a single training–test split. At each evaluation step, the model is tested on a portion of the data while being trained on the remaining portions, and this process is repeated to cover all data subsets. The obtained performance metrics are combined to determine the overall accuracy level of the model (Taşpınar et al., 2021). In this study, a 10-fold structure was preferred, as it was determined through experimental analysis to provide the most balanced and highest accuracy values. The general flow of the cross-validation process is presented in Figure 1.

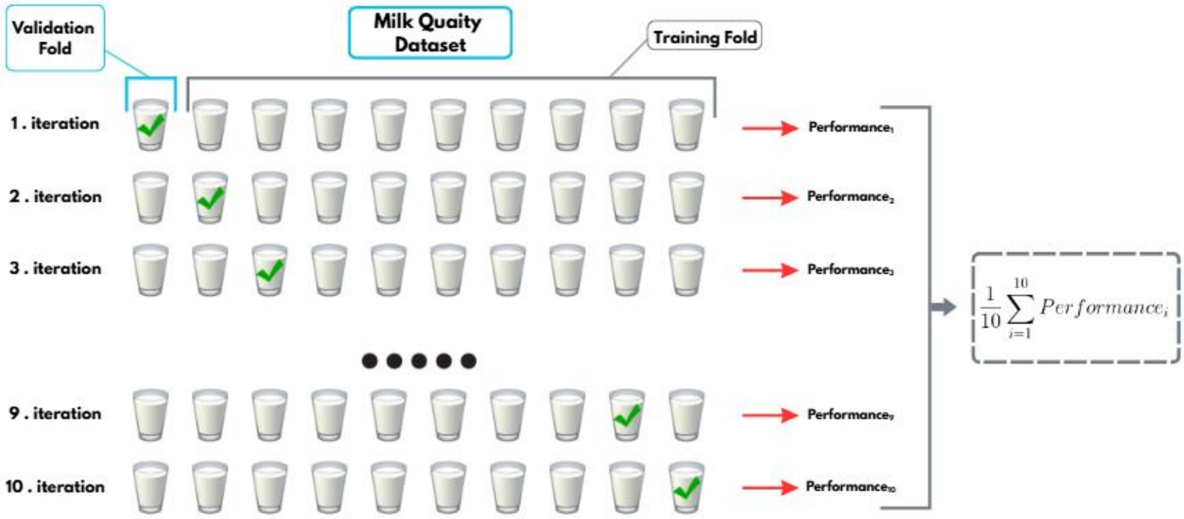


Figure 1. *Cross Validation*

Result and Discussion

In this study, the Milk Quality Prediction by Classification dataset was used to evaluate and classify milk quality. The data were divided into three classes (High, Medium, and Low) based on quality status. kNN, ANN, DT, and RF methods were applied for classification. The Python programming language was used as the development environment. For the kNN method, the number of neighbors (k) = 5, the distance metric = Euclidean, and the weight type = uniform. For the ANN model, 100 hidden layers, the ReLU activation function, a regularization coefficient of 0.0001, and 200 iterations were used. For the DT method, a binary tree structure was created, with a minimum number of samples in the leaves set to 2, a minimum number of samples for splitting subsets set to 5, and a maximum tree depth set to 100. Additionally, a stopping condition was applied when the majority ratio reached 95% during the classification process. The 10-fold cross-validation method was used to compare the performance of the models. Training and testing were performed using this validation method, and the classification accuracy, precision, recall, and F1 scores of each model were compared. The general steps applied in the study are shown in Figure 2.

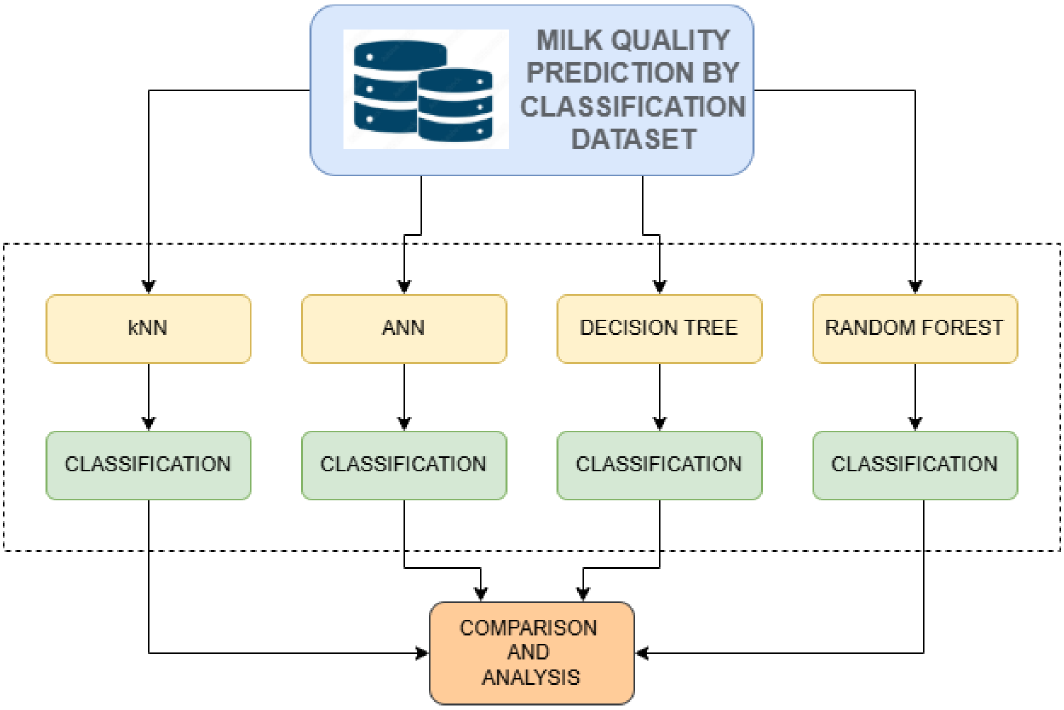


Figure 2. *Process steps*

In this study, seven features obtained from 1059 data points used to classify milk quality were provided as input to machine learning methods, and the training processes of the models were carried out. The confusion matrices obtained from the machine learning methods are presented in order. The confusion matrix for the kNN method is shown in Table 3.

Table 3. *Confusion Matrix of kNN model*

Predicted

Actual

	High	Low	Medium	∑
High	255	0	1	256
Low	0	426	3	429
Medium	1	0	373	374
∑	256	426	377	1059

According to Table 3, the KNN method correctly classified 255 data points belonging to the High class. One data point belonging to the High class was classified as Medium, and there were no misclassifications in the Low class. In total, one data point belonging to the High class was misclassified. 426 data points belonging to the Low class were correctly classified. Three data points belonging to this class were classified as Medium, and there were no misclassifications in the High class. In total, 3 data points belonging to the Low class were misclassified. 373 data points belonging to the Medium class were correctly classified, and 1 data point was classified as High. There were no misclassifications in the Low class. In total, 1 data point belonging to the Medium class was misclassified. Table 4 shows the confusion matrix for the ANN method.

Table 4.*Confusion Matrix of ANN model*

		Predicted			
		High	Low	Medium	∑
Actual	High	255	0	1	256
	Low	3	426	0	429
	Medium	22	1	351	374
	∑	280	427	352	1059

According to Table 4, the ANN method correctly classified 255 data points belonging to the High class. One data point belonging to the High class was classified as Medium, and there were no misclassifications in the Low class. In summary, one data point belonging to the High class was misclassified. A total of 426 data points belonging to the Low class were correctly classified. Three data points belonging to this class were classified as Medium, and there were no misclassifications in the High class. In total, 3 data points belonging to the Low class were misclassified. 373 data points belonging to the Medium class were correctly classified, and 1 data point was classified as High. There were no misclassifications in the Low class. In total, 1 data point belonging to the Medium class was misclassified. Table 5 shows the confusion matrix for the Decision Tree method.

Table 5 .*Confusion Matrix of Decision Tree model*

		Predicted			
		High	Low	Medium	∑
Actual	High	253	0	3	256
	Low	2	426	1	429
	Medium	2	0	372	374
	∑	257	426	376	1059

Table 5 presents the results obtained using the Decision Tree classifier. Within the High class, 253 samples were assigned correctly, while three samples were labeled as Medium. No instances from this class were predicted as Low, resulting in three misclassified High-class samples. For the Low class, 426 samples were correctly identified. Misclassifications in this group included two samples predicted as High and one predicted as Medium, giving a total of three incorrectly classified Low-class samples. The Medium class contained 372 correctly classified samples, with two samples assigned to the High class. No Medium samples were classified as Low. In total, two Medium-class samples were misclassified. The confusion matrix of the Random Forest model is shown in Table 6.

Table 6. *Confusion Matrix of Random Forest model*

Predicted

Actual

	High	Low	Medium	∑
High	255	1	0	256
Low	2	427	0	429
Medium	0	0	374	374
∑	257	428	374	1059

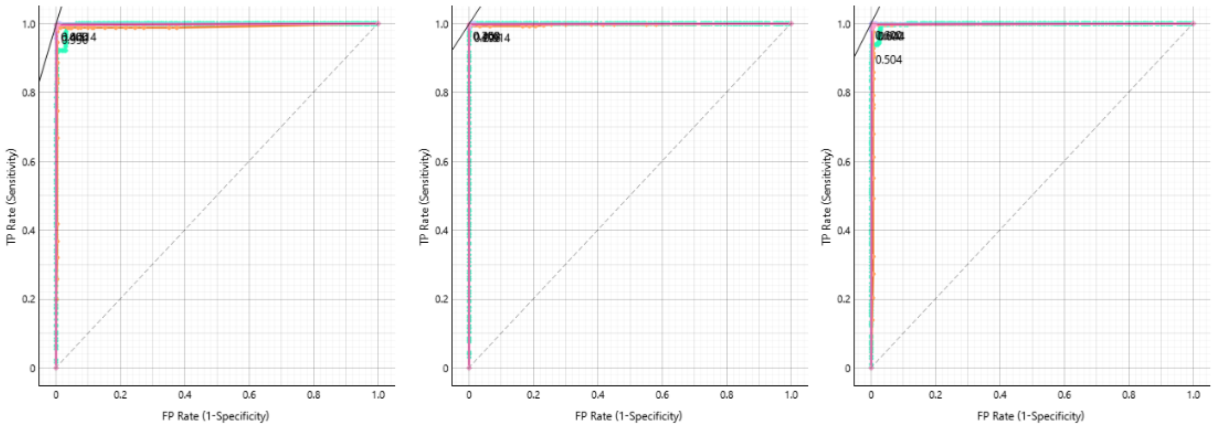
Table 6 shows the classification results of the Random Forest model. In the High class, 255 samples were correctly identified, while two samples were assigned to the Low class. No samples from this group were classified as Medium, and the number of misclassified High-class samples was two. For the Low class, 427 samples were correctly classified. Two samples in this category were labeled as High, with no instances assigned to the Medium class. In total, two Low-class samples were misclassified. All 374 samples in the Medium class were classified correctly, with no assignments to either the High or Low classes. As a result, no misclassifications were observed for the Medium class.

Performance metrics for each method were calculated using data on the complex matrices obtained from the methods. These metrics are shown in Table 3.

Table 7. *Performans Metrics of All Models*

	Accuracy	F1 Score	Precision	Recall
kNN	0,998	0,995	0,995	0,995
ANN	0,999	0,975	0,976	0,975
Decision Tree	0,994	0,992	0,992	0,992
Random Forest	1,000	0,997	0,997	0,997

According to Table 3, the highest classification success was achieved with the Random Forest model at 1.000. This is followed by ANN at 0.999, kNN at 0.998, and Decision Tree models at 0.994. When examining the F1 Score, Precision, and Recall metrics, it is seen that the classification successes of the models are parallel. Specifically, while the Random Forest model achieved the highest values across all metrics, the Decision Tree model performed slightly lower than the other models. Overall, when examining the classification accuracy and F1 Score, Precision, and Recall values, it can be said that all models achieved over 99% accuracy. Figure 3 shows the ROC curves for all models.



High Low Medium

Figure 3. *Roc of All Models*

When examining Figure 3, classification accuracies and the learning levels of the models show parallelism. The best-performing model is RF, while the least-performing model compared to others is DT. Consequently, it can be stated that the classification accuracies of the models emerge based on their learning rates.

Conclusion

Electronic sensors have recently become widely used in determining milk quality. As a result of the accurate analysis of data obtained from these sensors, milk quality can be quickly determined. Based on this, this study classified data obtained from 1059 milk samples. The highest classification success rate was achieved with the RF model at 100%.

The lowest classification success rate of 99.4% was obtained from the DT model. Upon examining the results, it can be stated that the machine learning models used in this study can be used in milk quality classification problems. The study was conducted using data obtained from 1059 types of milk samples. In real-life tests, the success rate of the models may decrease in some cases depending on data diversity and conditions. To overcome this problem, it is important to increase the amount of data and strengthen the models with broader and more diverse data. Furthermore, testing different classification methods can also help achieve higher success rates in real-world applications. Thanks to the classification models and data obtained from sensors proposed in the study, it is possible to quickly and accurately determine milk quality. These technologies will enable the monitoring of milk quality in production facilities, collection networks, packaging facilities, and businesses where milk is consumed. Thus, consumers will have the opportunity to consume the highest quality milk. Furthermore, the developed herd management systems enable the monitoring of cow-based milk quality. These systems provide important information about the health and milk productivity of cows, allowing for the early detection of potential diseases and increased productivity. Ultimately, these methods offer significant benefits in terms of both increasing the productivity of the dairy industry and ensuring food safety.

References

Alipio, M., & Bures, M. (2024). Current testing and performance evaluation methodologies of LoRa and LoRaWAN in IoT applications: Classification, issues, and future directives. *Internet of things*, *25*, 101053.

Chauhan, Y. J. (2020). Cardiovascular disease prediction using classification algorithms of machine learning. *International Journal of Science and Research*, *9*(5), 194-200.

Civalek, Ö. (1998). Nöro-Fuzzy Tekniği ile Dikdörtgen Plakların Analizi. *III. Ulusal Hesaplamalı Mekanik Konferansı*, 16-18.

Cui, H., Lin, K., Li, X., Qian, J., & Hu, H. (2024). Road Pothole Detection Model Based on AlexNet Convolutional Neural Network and Confusion Matrix. 2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE),

Çelik, A. (2022). Using machine learning algorithms to detect milk quality. *Eurasian Journal of Food Science and Technology*, *6*(2), 76-87.

Çetintav, B., & Yalçın, A. (2025). Explainable Machine Learning Framework for Milk Quality Grading. *Kocatepe Veterinary Journal*, *18*(3), 227-235.

Çınar, M., Engin, M., Engin, E. Z., & Ateşçi, Y. Z. (2009). Early prostate cancer diagnosis by using artificial neural networks and support vector machines. *Expert systems with applications*, *36*(3), 6357-6361.

Dinesh, S., & Dash, T. (2016). Reliable evaluation of neural network for multiclass classification of real-world data. *arXiv preprint arXiv:1612.00671*.

Eesa, A. S., Orman, Z., & Brifcani, A. M. A. (2015a). A new feature selection model based on ID3 and bees algorithm for intrusion detection system. *Turkish Journal of Electrical Engineering and Computer Sciences*, *23*(2), 615-622.

Eesa, A. S., Orman, Z., & Brifcani, A. M. A. (2015b). A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. *Expert systems with applications*, *42*(5), 2670-2679.

Farabi, S. F., Prabha, M., Alam, M., Hossan, M. Z., Arif, M., Islam, M. R., & Biswas, M. (2024). Enhancing credit card fraud detection: A comprehensive study of machine learning algorithms and performance evaluation. *Journal of Business and Management Studies*, *6*(3), 252-259.

Ferdowsy, F., Rahi, K. S. A., Jabiullah, M. I., & Habib, M. T. (2021). A machine learning approach for obesity risk prediction. *Current Research in Behavioral Sciences*, *2*, 100053.

Ji, B., Banhazi, T., Phillips, C. J., Wang, C., & Li, B. (2022). A machine learning framework to predict the next month's daily milk yield, milk composition and milking frequency for cows in a robotic dairy farm. *biosystems engineering*, *216*, 186-197.

Josephine, P. K., Prakash, V., & Divya, K. (2021). Supervised learning algorithms: A comparison. *Kristu Jayanti Journal of Computational Sciences (KJCS)*, 01-12.

KELEŞOĞLU, Ö., EKİNCİ, C. E., & FIRAT, A. (2005). The using of artificial neural networks in insulation computations. *Sigma*, *3*, 58-66.

Kumar, N., Patil, A., Reddy, D. R., RU, K., & Basavaraj, P. (2024). Predictive Analytics for Milk Quality Using Random Forest (RF) Algorithm. 2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS),

Kumar, R. R., Mudepalli, S., Sathish, N., Upender, T., & Mohmmad, S. (2024). An optimized neural network model to predict milk quality. 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence),

Kumari, S., Gourisaria, M. K., Das, H., & Banik, D. (2023). Deep learning based approach for milk quality prediction. 2023 11th International conference on emerging trends in engineering & technology-signal and information processing (ICETET-SIP),

Lakshmi, S. V., Meena, M., & Kiruthika, N. (2019). Diagnosis of chronic kidney disease using random forest algorithms. *International Journal of Research in Engineering, Science and Management*, *2*(3), 559-562.

Lorenz, A., Blum, M., Ermert, H., & Senge, T. (1997). Comparison of different neuro-fuzzy classification systems for the detection of prostate cancer in ultrasonic images. 1997 IEEE ultrasonics symposium proceedings. an international symposium (Cat. No. 97CH36118),

Love, P. E. (2024). A pragmatist research agenda for employing psychological heuristics in construction: Context, design, artificial intelligence, and performance evaluation. *IEEE Transactions on Engineering Management*, *71*, 11183-11197.

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.

Neamț, R., Ilie, D. E., Gavojdian, D., Acatincăi, S., Neciu, F., & Cziszter, L. (2016). Influence of electrical conductivity, days in milk and parity on milk production and chemical composition. *Scientific Papers Animal Science and Biotechnologies*, *49*(2), 128-128.

Özbay, Y. (1999). *EKG aritmilerini hızlı tanıma* Doktora Tezi].

Ronco, A. L., & Fernandez, R. (1999). Improving ultrasonographic diagnosis of prostate cancer with neural networks. *Ultrasound in medicine & biology*, *25*(5), 729-733.

Shrijayan. *Milk Quality Prediction by Classification*. https://www.kaggle.com/datasets/cpluzshrijayan/milkquality

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, *45*(4), 427-437.

Sunithamani, S., Muralidhar, D., Anne, G., & Sruthi, C. N. (2024). Milk quality prediction using Machine Learning integrated with Arduino. 2024 10th International Conference on Communication and Signal Processing (ICCSP),

Taşpınar, Y. S., Köklü, M., & Altın, M. (2021). Fire detection in images using framework based on image processing, motion detection and convolutional neural network. *International Journal of Intelligent Systems and Applications in Engineering*.

Taşpınar, Y. S., & Selek, M. (2020). Object recognition with hybrid deep learning methods and testing on embedded systems. *International Journal of Intelligent Systems and Applications in Engineering*.

Tian, X., Chong, Y., Huang, Y., Guo, P., Li, M., Zhang, W., Du, Z., Li, X., & Hao, Y. (2019). Using machine learning algorithms to predict hepatitis B surface antigen seroclearance. *Computational and mathematical methods in medicine*, *2019*(1), 6915850.

VijiyaKumar, K., Lavanya, B., Nirmala, I., & Caroline, S. S. (2019). Random forest algorithm for the prediction of diabetes. 2019 IEEE international conference on system, computation, automation and networking (ICSCAN),

Wang, L. (2019). Research and implementation of machine learning classifier based on KNN. IOP Conference Series: Materials Science and Engineering,

Yogi, K. S., Gowda, D., Sujithra, L., Prasad, K., & Midhun, P. (2024). Scalability and performance evaluation of machine learning techniques in high-volume social media data analysis. 2024 11th international conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO),