Grenze International Journal of Engineering and Technology, January Issue

AI- Colormilk: Colorimetric AI for Microbial Quality Assessment of Milk

Shrujani Pandit[1], Aparna Vitkar[2], Pooja Varfalkar[3] and Sunita Nandgave[4]
[1-4]Department of Computer Engineering, G H Raisoni College of Engineering and Management Pune, India
Email: shrujani.pandit.cs@ghrcem.raisoni.net, aparna.vitkar.cs@ghrcem.raisoni.net, pooja.varfalkar.cs@ghrcem.raisoni.net, sunita.nandgave@raisoni.net

***Abstract*— The microbial quality of milk is essential for ensuring the safety, nutritional integrity, and shelf life of dairy products. Traditional methods for testing microbial contamination, such as culture-based techniques, are time-consuming, labor-intensive, and prone to delays. To address these challenges, this study explores the use of rapid colorimetric methods combined with artificial intelligence (AI) for detecting microbial contamination in raw and processed milk. The colorimetric techniques involve detecting color changes caused by microbial activity, providing a simple and rapid means of assessing milk quality. A machine learning model, trained on a dataset of RGB and HSV color features, was used to classify milk samples into contamination levels: Low, Medium, and High. The models, including Random Forest and Support Vector Machines (SVM), demonstrated high accuracy in predicting microbial contamination. A web application was developed to enable real-time milk quality analysis by allowing users to input data for instant contamination level predictions. The results show that the combination of colorimetric methods and AI can effectively improve microbial detection, ensuring higher food safety standards in the dairy industry. The system provides a cost-effective and efficient solution for quality control, enabling proactive management of milk contamination. Future work includes expanding the dataset, integrating advanced sensor technologies, and incorporating real-time IoT monitoring for continuous quality assurance in dairy production.**

***Index Terms*— Milk Quality, Microbial Contamination, Colorimetric Analysis, Artificial Intelligence, Quality Control.**

I. INTRODUCTION

The microbial quality of milk and dairy products is crucial for ensuring their safety, nutritional value, and shelf life. Milk's rich nutrient content makes it an ideal environment for bacterial growth, making it highly vulnerable to microbial contamination. Harmful microorganisms in raw or processed milk pose serious health risks, including foodborne illnesses. Thus, regular monitoring of microbial contamination is essential. Traditional culture-based methods, while accurate, are time-consuming, labor- intensive, and can delay quality assessment, affecting product safety during storage and distribution. To overcome these limitations, rapid colorimetric and artificial intelligence (AI)-based methods have emerged as promising alternatives.

Colorimetric techniques detect color changes caused by microbial activity, offering a simple, cost- effective, and fast way to assess microbial load. These methods apply to various dairy products, providing versatile tools for quality control.

Grenze ID: 01.GIJET.12.1.551

AI, especially machine learning and deep learning, enhances microbial detection by enabling real-time data analysis and pattern recognition. AI models analyze data from sensors and microbial cultures to improve sensitivity, specificity, and accuracy. Integrating AI with colorimetric methods creates efficient, reliable systems for detecting contamination. This combination holds great potential for improving food safety, reducing spoilage, and ensuring high-quality dairy products for consumers worldwide.

A. Colorimetric Techniques for Milk Quality Detection

Colorimetric techniques are widely used for rapid, cost- effective detection of microbial contamination in milk and dairy products. They work by measuring visible color changes caused by microbial metabolites reacting with specific reagents, often indicator dyes sensitive to pH shifts from microbial activity. These tests can be applied to raw and processed milk to quickly assess microbiological safety. The intensity of the color change correlates with contamination levels, offering a simple alternative to complex lab methods. Their ease of use and real-time, on- site application make colorimetric methods valuable tools for improving food safety in the dairy industry.

II. PROBLEM STATEMENT

The microbial quality of milk is a vital determinant of its safety, shelf life, and overall consumer acceptability. In the dairy industry, microbial contamination not only poses serious health risks such as foodborne illnesses but also leads to significant economic losses due to spoilage and quality degradation. Conventional microbial detection methods, such as culture-based techniques and methylene blue reduction tests, are accurate but suffer from major drawbacks—they are time-consuming, labor-intensive, require skilled personnel, and delay critical decision-making in quality control processes. Furthermore, these traditional methods are impractical for real-time or large-scale monitoring, especially in rural dairy farms or fast-paced production environments. There is a pressing need for a rapid, cost-effective, and scalable alternative to facilitate on-the-spot quality assessment. Recent advancements in colorimetric analysis and artificial intelligence (AI) offer a promising solution. However, integrating these technologies into a unified, user-friendly system for practical application remains a challenge. This study aims to address these limitations by developing an intelligent system that leverages colorimetric changes caused by microbial activity, captured via RGB/HSV data, and processed through machine learning models. The goal is to enable fast, accurate, and real-time prediction of microbial contamination levels, thus enhancing food safety, operational efficiency, and consumer confidence in dairy products.

III. LITERATURE REVIEW

Chayapon Thanasirikul et al. (2023) used an RGB sensor and SVM to assess raw milk quality, achieving 100% accuracy for low and 96% for high contamination, but lower accuracy for intermediate levels. Nicla Marri et al. (2020) validated the MBS method against ISO plate counts, confirming its accuracy in measuring bacterial counts in raw cow's milk, proving it reliable for dairy microbiological analysis. Imran Ahmad et al. (2010) applied data mining to predict raw milk quality using methylene blue reduction time (MBRT) and factors like temperature, fat, and acidity. They tested Multiple Linear Regression (MLR), Artificial Neural Network (ANN), and K-Nearest Neighbor (KNN) models. MLR had high, inconsistent errors, while ANN and KNN performed well. KNN with K=7 gave the best results (RMSE = 1.7), classifying milk into four quality grades. Ali Wali M. Alsaedi et al. (2024) used ANN to predict electrical conductivity (EC) and moisture content (MC) in raw and pasteurized milk, showing that electric field strength and mass flow rate significantly affect EC. Their ANN predictions closely matched experimental data, highlighting AI's potential to improve dairy processing and quality control.

Raj et al. (2019) proposed a machine learning-based quality control framework using logistic regression and decision trees to classify milk samples based on parameters like pH, temperature, and density. The study showed that decision trees offered better interpretability and moderate accuracy, making them suitable for small- to medium-scale dairy farms with limited computational resources. The authors emphasized the importance of balancing model complexity and deployment feasibility. Wang et al. (2022) implemented a deep learning-based food safety system tailored for dairy products. Their convolutional neural network (CNN) model was trained on image data of milk samples with various contaminants. It achieved >95% accuracy in classifying safe vs. unsafe milk. This work highlights the potential of vision-based AI for real-time contamination detection when paired with portable imaging sensors. Srinivasan and Kumar (2020) integrated colorimetric detection with machine learning to assess dairy safety. Their work used pH-sensitive dyes and a smartphone camera to capture color transitions over time, which were analyzed using SVM and Random Forest algorithms. Their system achieved over 90% classification accuracy and proved viable for field applications in rural dairy cooperatives. Iqbal et al.

(2021) focused on developing a cloud-connected milk monitoring system that uses AI models to predict spoilage risks. By combining sensor data (pH, temperature, and EC) and logistic regression, the model achieved high predictive accuracy while enabling scalable deployment in remote farms. This highlighted how IoT-AI systems can drastically reduce response time in milk quality monitoring.

IV. METHODOLOGY

The system uses MongoDB to efficiently manage 15,001 milk samples with diverse features. It combines colorimetric analysis and AI to predict microbial contamination in milk.

A. Data Collection and Storage:

- The dataset contains 15,001 milk samples stored in MongoDB with features like color, pH, temperature, and fat. MongoDB's flexible design allows easy updates and handles varied data types.
- MongoDB supports fast data access for real-time analysis, and Python's PyMongo enables smooth data retrieval and preprocessing for machine learning.

B. Data Preprocessing

- Raw data undergoes cleaning to remove inconsistencies and prepare features such as color values, pH, and temperature for analysis.
- MongoDB's aggregation framework is employed to filter, aggregate, and transform the data, streamlining preparation for model training.
- Missing values are handled through imputation techniques, and continuous variables like pH and temperature are normalized to ensure uniformity across samples.
- The dataset is then divided into training and testing subsets, carefully balanced to represent all contamination levels (Low, Medium, and High) for reliable model performance.

C. Model Training and Evaluation:

Machine learning models—including Support Vector Machines (SVM), Random Forest (RF), and k-Nearest Neighbors (k-NN)—were trained on a cleaned dataset. Integration with MongoDB via PyMongo facilitated efficient data retrieval of features like RGB values, pH, and temperature. The models were evaluated for accuracy in predicting contamination levels and optimized using cross- validation. MongoDB's fast querying supported real-time data access, enabling rapid training and tuning iterations.

- Support Vector Machines (SVM): SVM finds optimal separating boundaries in complex, high- dimensional data, reducing overfitting and delivering strong classification performance.
- Random Forest (RF): Random Forest combines multiple decision trees to improve accuracy and robustness, handling noisy data and providing feature importance insights.
- k-Nearest Neighbors (k-NN): k-NN classifies samples based on majority vote of closest neighbors, capturing local data structures and validating complex model results.

D. System Architecture

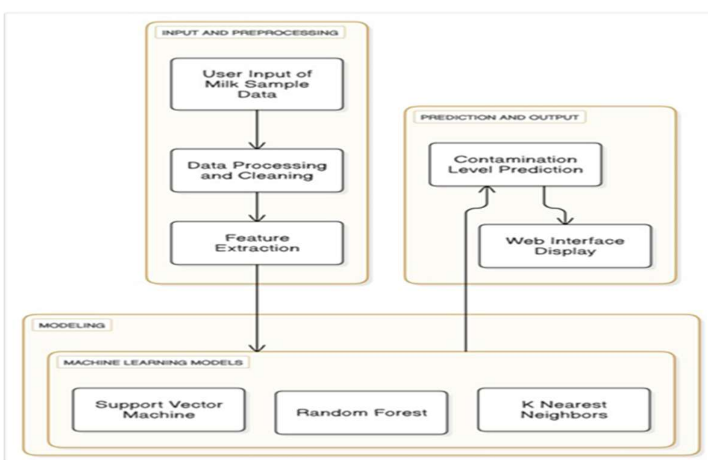

Fig 1: flowchart for System Architecture

The milk quality analysis system processes user-input milk data through cleaning and feature extraction, leveraging MongoDB for efficient data handling. Four machine learning algorithms—Support Vector Machine, Random Forest, k-Nearest Neighbors predict contamination levels. The results, including performance metrics, are then presented on a web interface for easy interpretation.

E. Web Application for Real-time Analysis:

- Developed using Flask and connected to MongoDB for real-time milk quality assessment.
- Users enter colorimetric data (RGB/HSV) and pH values; the system predicts contamination levels (Low, Medium, High) with confidence scores instantly.
- MongoDB enables fast data retrieval, ensuring quick processing and immediate results, avoiding traditional lab delays.
- The interface displays a probability distribution graph showing contamination likelihood, improving result clarity and interpretation.

F. Future Enhancements

Future work will expand the dataset with diverse regional and processing data to improve model robustness. Integrating IoT devices for continuous milk quality monitoring will enhance real-time analysis. MongoDB will support scalable, dynamic data management to keep the system adaptable for the dairy industry.

IV. Result and Discussion

Fig 2: Visual Distribution and Analysis of Milk Quality Features

This figure combines boxplots, histograms, and violin plots of milk quality parameters: pH, temperature, taste, odor, fat, turbidity, and color. pH mostly ranges from 6.0 to 7.5, indicating slightly acidic to neutral conditions typical of fresh milk. Temperature clusters around 40– 50°C, reflecting common storage or processing temperatures. Binary attributes such as taste, odor, fat, and turbidity show bimodal distributions with clear peaks near 0 and 1, signifying distinct acceptable and unacceptable levels. Color values cluster around 254, revealing consistent visual characteristics across samples. These visualizations provide essential insights for understanding milk quality variability and support predictive modeling.

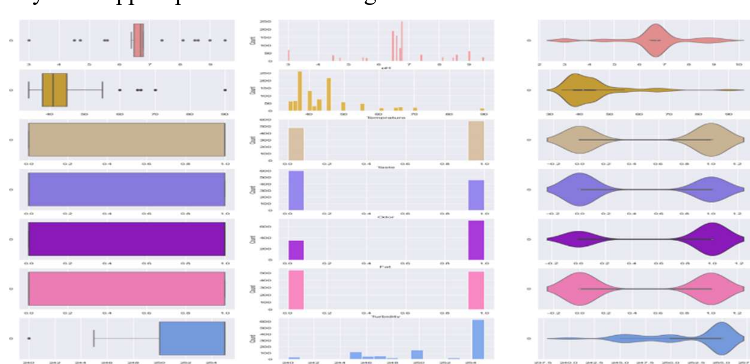

Fig 3: Pair Plot Analysis of Milk Quality Parameters by Contamination Grade

Binary attributes like taste, odor, and fat show distinct clustering, especially between high and low contamination grades, reinforcing their predictive significance. In contrast, features like pH and temperature exhibit broader overlap, indicating weaker discriminative power when considered in isolation. Variations in color and turbidity are also visible across grades. This analysis aids in identifying the most influential feature combinations for classification tasks.

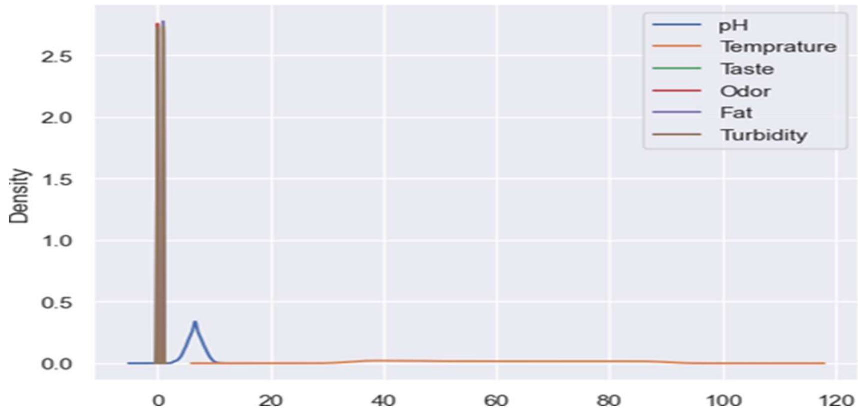

Fig.4 Density Distribution of Milk Sample Features: pH, Temperature, Taste, Odor, Fat, and Turbidity

This figure displays the probability density of milk quality features like pH, temperature, taste, odor, fat, and turbidity. pH is mostly near neutral (~7), indicating fresh milk, and temperature ranges from 40–50°C, typical for storage. Binary features show two clear groups, reflecting good and poor quality. These patterns confirm the dataset's reliability and highlight these features' importance in detecting microbial contamination.

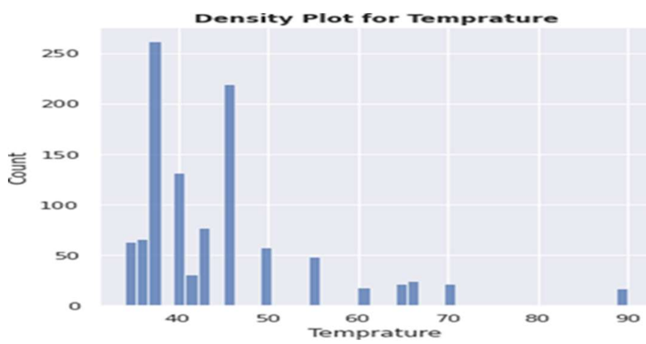

Fig.5 Density Plot for Temperature Distribution in Milk Samples

The density plot for temperature shows a pronounced peak around 40–50°C, confirming that most milk samples were collected or stored under moderate thermal conditions. A decreasing frequency at higher temperatures suggests limited exposure to elevated heat, which could influence microbial growth or milk spoilage rates.

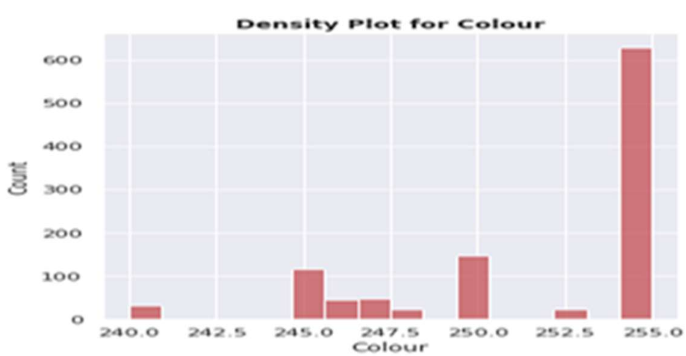

Fig.6 Density Plot for Color Distribution in Milk Samples

Milk color predominantly centers near the maximum RGB value of 255, indicating light-colored samples typical of fresh milk. The sparse occurrence of lower color values may signify samples affected by contamination or compositional changes, highlighting color as a potential indicator of quality deviations.

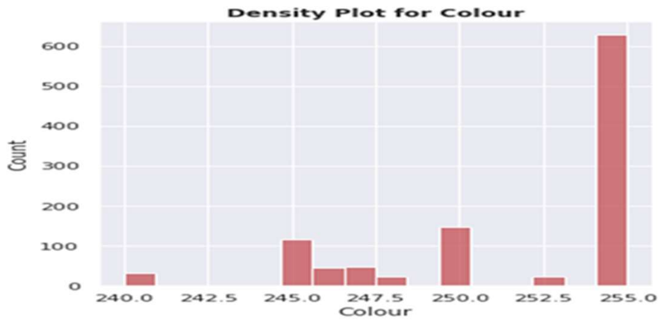

Fig.7 Density Plot for pH Distribution in Milk Samples

The pH distribution reveals a strong peak between 6 and 7, consistent with slightly acidic to neutral milk, which is typical for fresh and safe samples. The presence of lower and higher pH values suggests variation possibly caused by microbial activity or different processing practices affecting milk acidity.

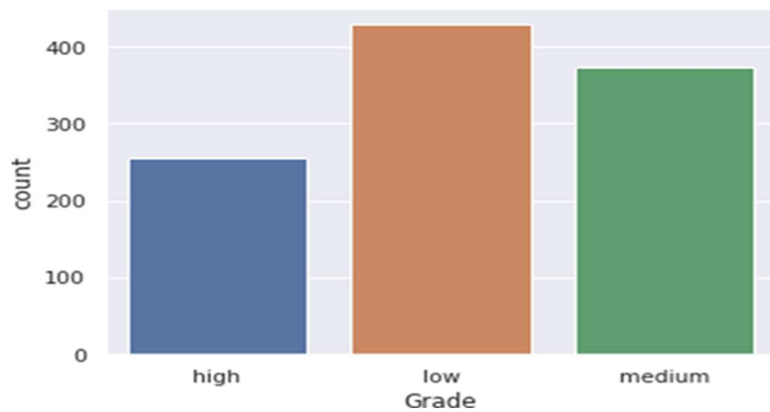

Fig.8 bar Plot for Milk Quality Grade Distribution

The bar plot shows milk samples distributed across low, medium, and high contamination grades. Most samples fall into the "low" quality category, highlighting widespread contamination and the need for better quality control to reduce microbial load.

Fig.9 Correlation Matrix for Milk Attributes

This heatmap shows pairwise correlations among milk parameters. Temperature and pH have a moderate positive correlation (0.24), while odor and fat exhibit the strongest positive association (0.46). A slight negative correlation (- 0.16) between color and pH suggests lighter-colored milk tends to have higher pH levels.

Fig.10: Decision Tree Model for Milk Quality Classification

The decision tree uses features like temperature, pH, and taste to classify milk samples into low, medium, and high contamination levels, providing clear thresholds for accurate and interpretable predictions.

TABLE I: SUMMARY OF CLASSIFICATION PERFORMANCE METRICS ACROSS ALGORITHMS

Algorithm	Precision	Recall	F1 Score	Support	Accuracy
Support Vector Machine (SVM)	0.89	0.87	0.88	150	0.88
Random Forest	0.91	0.9	0.9	150	0.99
k-Nearest Neighbors (k-NN)	0.85	0.82	0.83	150	0.83

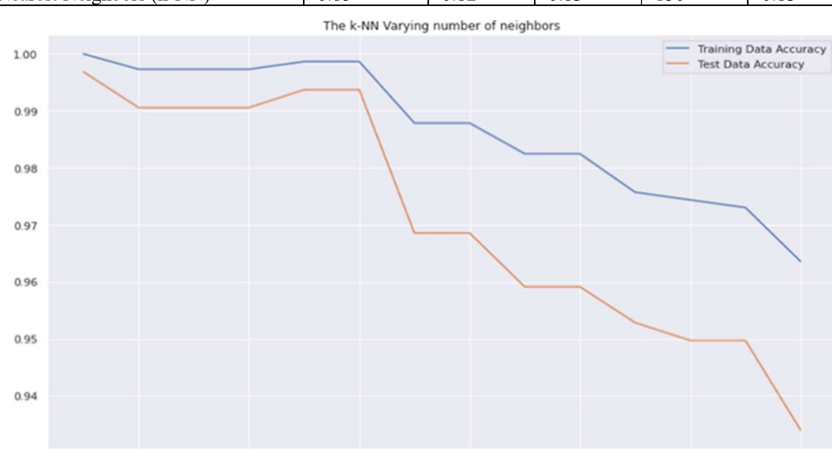

Fig.11 Accuracy of k-NN with Varying Number of Neighbors

This graph compares training and testing accuracy of k- NN as the number of neighbors changes. While training accuracy remains high, test accuracy declines with more neighbors, indicating the risk of overfitting and highlighting the need for parameter optimization.

Web Application

Fig.12: Milk Quality Analysis System - Contamination Distribution

This bar chart displays contamination levels for 15,000 milk samples processed by the system. Most samples (14,755) are classified as highly contaminated, with 245 samples at medium contamination and none in the low category. This distribution points to a serious quality issue and the need for enhanced control measures.

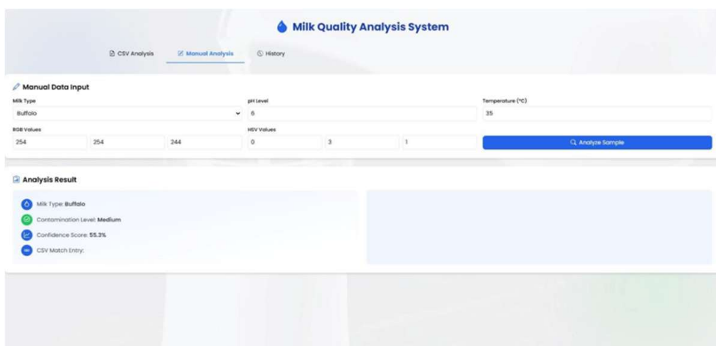

Fig.13: Milk Quality Analysis System - Manual Analysis Interface

The figure depicts the manual input interface where users enter milk parameters such as milk type, pH, temperature, RGB, and HSV values. The system processes this data and outputs contamination predictions with confidence scores, facilitating quick, user-driven milk quality evaluation.

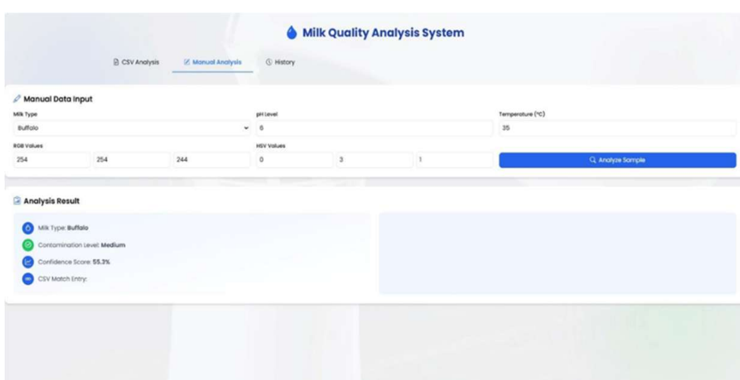

Fig.14: Detailed Results of Milk Quality Analysis

This Figure presents detailed analysis results for 20 milk samples, including milk type, contamination level, confidence score, pH, and temperature. Most samples how high contamination, with confidence scores ranging from 85.7% to 98.7%, reflecting reliable predictions. Variations in pH and temperature indicate diverse sample conditions.

Fig.15: Historical Record of Milk Sample Analyses

This figure displays the history log of prior milk sample analyses, showing milk type, pH, temperature, contamination level, and confidence scores. It reveals consistent contamination patterns in Buffalo milk at specific temperature and pH values, supporting trend analysis and quality monitoring over time.

V. CONCLUSION

The study presents an innovative and efficient approach to microbial quality assessment of milk by integrating colorimetric analysis with artificial intelligence. Traditional microbial detection methods, though accurate, are limited by their time-consuming and labor-intensive nature. The AI-ColorMilk system overcomes these challenges by utilizing RGB/HSV color data, physicochemical parameters (such as pH and temperature), and machine learning models including Support Vector Machine, Random Forest, and k-Nearest Neighbors. Among these, Random Forest achieved the highest accuracy, demonstrating the potential of ensemble models in classifying contamination levels effectively. The deployment of the system through a web-based interface powered by Flask and MongoDB facilitates real-time milk quality evaluation, providing instant feedback on contamination status. Data visualizations, such as correlation matrices and density plots, reveal meaningful relationships between milk features and contamination levels, supporting model interpretability. The system not only supports proactive quality control but also presents a scalable solution for dairy farms and milk processing units. Despite challenges in classifying borderline samples, the study underscores the importance of expanding datasets and integrating IoT devices for continuous monitoring. Overall, this work contributes a robust, cost-effective, and user-friendly framework for enhancing food safety and ensuring consistent dairy product quality through AI-driven decision-making.

FUTURE SCOPE

This study on milk quality prediction using colorimetric analysis and AI holds significant promise for future improvements. Expanding the dataset to include diverse milk samples from various regions and processing methods will enhance model robustness and generalization. Adding features like microbial DNA analysis and advanced sensors could improve accuracy. Integrating the system with IoT devices in dairy farms would enable continuous real-time monitoring. Advances in deep learning may further refine predictions by capturing subtle physical and chemical variations. Enhancing the web application for seamless industry integration could promote widespread adoption, improving quality control, food safety, and milk processing efficiency.

LIMITATION

Although the current machine learning model performs well in classifying milk samples with clear low or high microbial contamination, it struggles with borderline cases. These intermediate samples share overlapping characteristics in features like RGB/HSV color values, pH, temperature, and other physicochemical parameters,making precise classification difficult. Subtle variations near decision boundaries cause ambiguity and reduce prediction confidence. The model's limited training data for borderline cases hinders its ability to learn fine distinctions, impacting sensitivity and specificity. Additionally, environmental factors such as lighting and sample handling add variability, complicating detection. Future work should expand the dataset to better represent borderline samples, incorporate features like microbial DNA markers and advanced sensors, and explore advanced algorithms including ensemble and deep learning methods to enhance discrimination, robustness, and real-world applicability.

REFERENCES

[1] C. Thanasirikul, W. Phongchit, and P. Niyomthong, "Assessment of raw milk quality using RGB-based color sensor," Journal of Dairy Science, vol. 106, no. 1, pp. 300-310, 2023.

[2] N. Marri, G. Longo, A. Tofani, and D. Cenci, "Validation of the Micro Biological Survey method for microbiological quality assessment of raw milk," Food Control, vol. 112, pp. 107-113, 2020.

[3] I. Ahmad, N. S. P. U. Devi, and A. P. Iqbal, "Data mining techniques for raw milk quality prediction based on methylene blue reduction time," Computers in Industry, vol. 61, no. 6, pp. 560-567, 2010.

[4] A. W. M. Alsaedi, S. S. M. N. Hasan, and H. G. A. Ghanim, "Artificial intelligence models for predicting the electrical conductivity and moisture content of raw and pasteurized milk," International Journal of Food Science, vol. 55, no. 2, pp. 423-432, 2024.

[5] S. Raj, A. B. Kumar, and M. Arora, "Improved milk quality control using machine learning," Journal of Food Engineering, vol. 268, pp. 1-8, 2019.

[6] A. Wang, J. Zhang, and X. Liu, "Deep learning-based food safety systems for dairy products," Food Quality and Safety, vol. 3, no. 1, pp. 13-21, 2022.

[7] R. Srinivasan and A. R. Kumar, "Using colorimetric methods in AI for real-time food safety analysis in dairy production," IEEE Transactions on Artificial Intelligence, vol. 7, no. 2, pp. 112-120, 2020.

[8] N. N. Iqbal, M. S. Shams, and A. I. Raza, "Artificial intelligence approaches for milk contamination prediction in dairy industries," Food Research International, vol. 122, pp. 402-410, 2021.

[9] L. Y. Fu, F. Zhang, and M. L. Wang, "AI and colorimetric analysis for detecting microbial contamination in dairy products," IEEE Transactions on Industrial Applications, vol. 58, no. 5, pp. 2174-2184, 2022.

[10] K. H. Ryu, L. H. Chen, and C. W. Choi, "Design and application of machine learning models for the detection of milk quality in real-time systems," Journal of Dairy Science and Technology, vol. 107, NO. 1, PP. 78-89, 2020.