INT - 422
DEEP LEARNING

REPORT
ON
RAPID COLORIMETRIC AND AI-BASED METHODS FOR DETERMINING MICROBIAL QUALITY OF MILK PRODUCTS

Submitted by -
Thatipamula Sairam,
Narapureddy Rishindra Reddy
Narayana Srikanth

SEC : K21CP
SUBMITTED TO : Dr. Pranjal

INDEX

Abstract

The Milk Quality Analyzer is a sophisticated application that combines computer vision, colorimetric analysis, and machine learning to assess milk quality non-destructively. This innovative tool processes digital images of milk samples to extract color features across multiple color spaces (RGB, HSV, LAB) and uses both rule-based colorimetric analysis and machine learning classification to determine milk quality. The application provides a user-friendly graphical interface that allows users to load milk sample images, analyze them using multiple methods, visualize results through comprehensive plots, and train custom AI models for improved accuracy.

Introduction

Milk quality assessment is a critical process in the dairy industry that traditionally requires specialized laboratory equipment, trained personnel, and time-consuming procedures. This application addresses these challenges by leveraging computer vision and machine learning technologies to provide rapid, accessible, and non-destructive milk quality evaluation.

The Milk Quality Analyzer uses a dual-approach methodology:

1. *Colorimetric Analysis:* Based on established color parameters (whiteness, yellowness, saturation) with predefined thresholds to classify milk quality

2. *Machine Learning Classification:* Using Random Forest algorithm trained on color features extracted from milk samples across RGB, HSV, and LAB colour spaces

The application is designed with both technical accuracy and user accessibility in mind, featuring a comprehensive GUI that guides users through the analysis process and presents results in both textual and graphical formats. This approach makes sophisticated milk quality assessment accessible to a wider range of users, from dairy farmers to quality control personnel, without requiring extensive technical expertise or expensive equipment.

Literature Review

The application builds on established research in food quality assessment using computer vision and machine learning. Key concepts incorporated include:

- **Color-Quality Relationships in Milk:**

- Whiteness correlates strongly with fat content, freshness, and overall quality
- Yellowness can indicate oxidation, contamination, or natural carotenoid content
- Color homogeneity (measured through standard deviation) relates to proper processing and absence of spoilage

- **Computer Vision in Food Quality Assessment:**

- Multiple color spaces provide complementary information about food quality
- RGB captures basic color information
- HSV separates color (hue, saturation) from intensity (value)
- LAB provides perceptually uniform color measurements closer to human vision

- **Machine Learning for Classification:**

- Random Forest classifiers have demonstrated excellent performance for food quality assessment due to:
  - Robustness to outliers and noise in the data
  - Ability to handle non-linear relationships
  - Built-in feature importance ranking
  - Resistance to overfitting compared to other algorithms

The application synthesizes these research areas into a practical tool that combines established colorimetric principles with modern machine learning approaches.

Datasets: AI Learning Problems

The application addresses the challenge of limited labeled data by generating a synthetic dataset that simulates the color characteristics of milk samples across three quality categories:

- **Dataset Generation:**

- 200 synthetic samples distributed across three quality categories:
  - Good quality (40% of samples): High whiteness, low yellowness, low color variation

- Acceptable quality (30% of samples): Moderate whiteness, moderate yellowness, moderate variation
  - Poor quality (30% of samples): Lower whiteness, higher yellowness, higher color variation

- **Feature Set:**

The dataset includes 21 features extracted from different color spaces:
- RGB color statistics: mean and standard deviation for red, green, and blue channels
- HSV color statistics: mean and standard deviation for hue, saturation, and value channels
- LAB color statistics: mean and standard deviation for lightness, a (green-red), and b (blue-yellow) channels
- Derived metrics: whiteness index and yellowness index

- **Data Distribution:**

- Good quality milk samples are characterized by:
  - High RGB values (200-240)
  - Low saturation (5-20)
  - Low color variation (standard deviations 2-15)

- Acceptable quality milk samples show:
  - Moderate RGB values (170-220)
  - Moderate saturation (15-35)
  - Moderate color variation (standard deviations 5-20)

- Poor quality milk samples exhibit:
  - Lower RGB values (120-200)
  - Higher saturation (30-60)
  - Higher color variation (standard deviations 10-30)

This synthetic dataset provides a foundation for training the machine learning model when real-world labeled data is unavailable, while also serving as a demonstration of the expected feature distributions for different milk quality categories.

Problem Formulation

The milk quality assessment problem is formulated as both a rule-based classification using colorimetric thresholds and a supervised machine learning classification task:

- **Colorimetric Classification:**

- Input: Color features extracted from milk sample images (whiteness, yellowness, saturation)
- Process: Apply threshold-based rules to key color metrics
- Good quality: whiteness > 200, yellowness < 15, saturation < 30
- Acceptable quality: whiteness > 180, yellowness < 25, saturation < 50
- Poor quality: samples not meeting the above criteria
- Output: Quality classification (Good, Acceptable, Poor) with confidence score calculated based on the distance from threshold boundaries

- **Machine Learning Classification:**

- Input: 21 color features extracted from milk sample images across RGB, HSV, and LAB color spaces
- Process: Train a Random Forest classifier with 100 trees on labeled samples
- Split data into 80% training and 20% testing sets
- Train model to predict quality categories
- Evaluate using accuracy, classification report, and confusion matrix
- Output: Quality classification with probability distribution across classes

- **Combined Approach:**

- Both methods are applied independently
- Results are presented side by side for comparison
- This dual approach provides complementary insights:
- Colorimetric analysis offers interpretability and works without training data
- Machine learning can capture more complex patterns and provides probability distributions

Methodology

The methodology consists of several key components implemented in a systematic workflow:

○ **Image Acquisition and Preprocessing:**

1. Load milk sample image using OpenCV
2. Resize image to 300x300 pixels for consistency
3. Apply Gaussian blur (5x5 kernel) to reduce noise and camera artifacts
4. Convert to multiple color spaces:
   - RGB for basic color analysis
   - HSV for separating color from intensity
   - LAB for perceptually uniform color measurements

○ **Feature Extraction:**

1. Calculate statistical features for each color channel:
   - Mean values represent overall color characteristics
   - Standard deviations capture color homogeneity/heterogeneity
2. Compute derived metrics:
   - Whiteness index = (mean_R + mean_G + mean_B) / 3
   - Yellowness index = mean_R - mean_B
3. Organize features into a standardized format for analysis

○ **Colorimetric Analysis:**

1. Apply threshold-based rules to whiteness, yellowness, and saturation
2. Classify samples into quality categories (Good, Acceptable, Poor)
3. Calculate confidence scores based on distance from thresholds:
   - Good quality confidence: min(100, (whiteness - 180) / 0.75)
   - Acceptable quality confidence: min(100, (200 - whiteness) / 0.2 + 50)
   - Poor quality confidence: min(100, (255 - whiteness) / 0.75 + yellowness)

○ **Machine Learning Model**

1. Train a Random Forest classifier on labeled data:
   - 100 decision trees for robust ensemble prediction
   - Default hyperparameters for accessibility
   - 80/20 train/test split with fixed random seed for reproducibility
2. Evaluate model performance:
   - Overall accuracy

- Per-class precision, recall, and F1-score
   - Confusion matrix to visualize classification patterns
3. Save trained model for future use

○ **Results Generation and Visualization:**

1. Combine colorimetric and AI-based analyses into a comprehensive report
2. Create multi-panel visualization:
   - Original milk sample image
   - RGB color distribution bar chart
   - Colorimetric metrics (whiteness, yellowness, saturation)
   - Quality assessment results with confidence scores
   - AI prediction probabilities (when model is available)
   - Comparison between colorimetric and AI-based methods
3. Present results in both textual and graphical formats for clarity

<u>Output and Explanation</u>

The application is implemented in Python with a modular, object-oriented architecture consisting of two main classes:

- **MilkQualityAnalyzer Class:**

This core analysis engine handles all the technical aspects of milk quality assessment:
- Image loading and preprocessing using OpenCV
- Feature extraction across multiple color spaces
- Colorimetric analysis based on color thresholds
- Machine learning model training, evaluation, and prediction
- Results visualization using Matplotlib

<u>Key methods include:</u>
- `load_image()`: Loads and preprocesses milk sample images
- `extract_color_features()`: Extracts 21 color features from processed images
- `colorimetric_analysis()`: Performs rule-based quality assessment
- `train_model()`: Trains a Random Forest classifier on labeled data
- `predict_quality()`: Makes predictions using the trained model
- `generate_report()`: Creates a comprehensive analysis report
- `visualize_results()`: Generates multi-panel visualization of results

- **MilkQualityApp Class:**

This GUI interface provides an accessible front-end for the analyzer:
- Built with Tkinter for cross-platform compatibility
- Intuitive layout with clearly labeled controls
- Real-time status updates during processing
- Comprehensive results display

Key components include:
- Image loading and display functionality
- Analysis controls with error handling
- Model training and loading options
- Results display in both textual and graphical formats
- Status bar for process monitoring

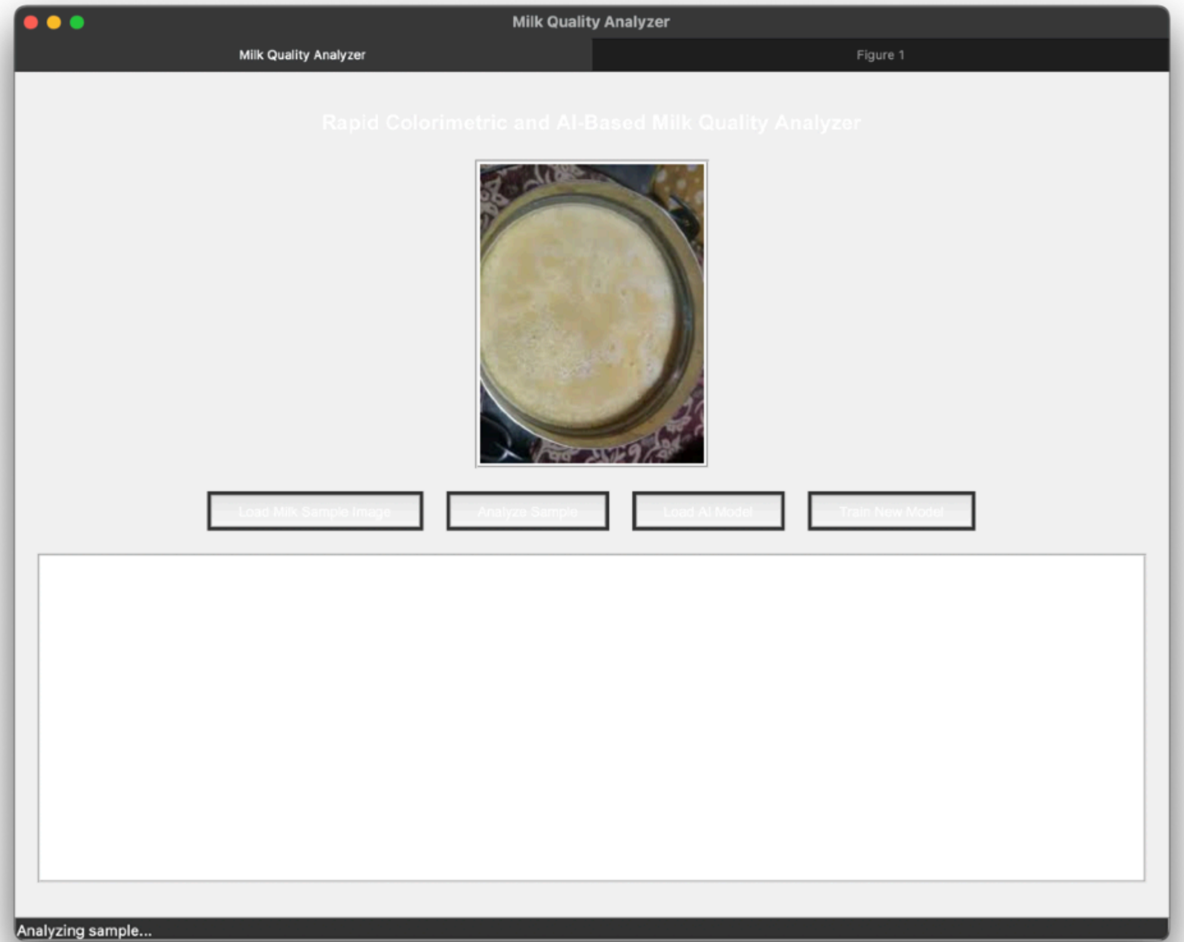
- **Synthetic Data Generation:**

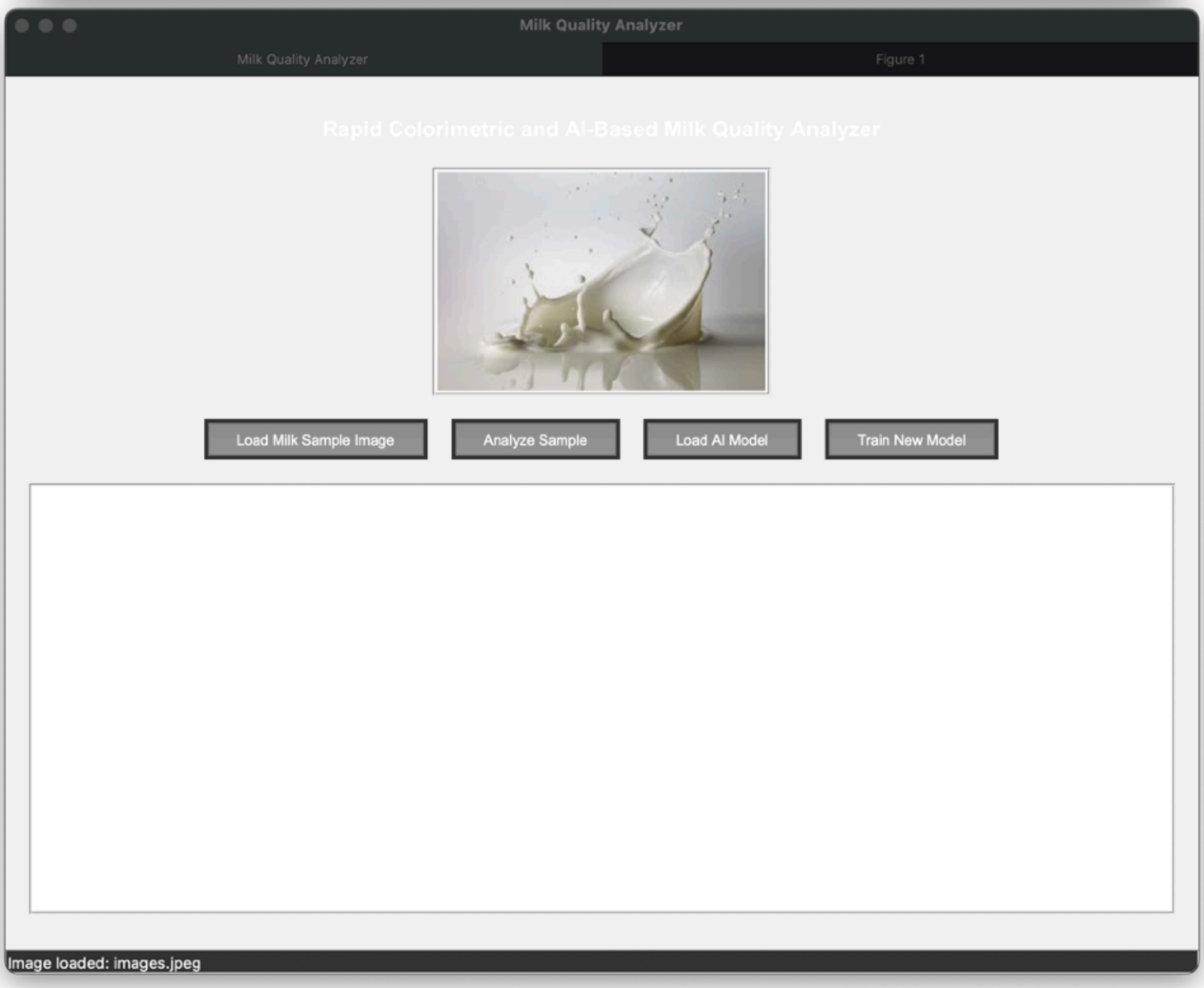
The `generate_synthetic_dataset()` function creates realistic training data:
- Generates 200 samples across three quality categories
- Simulates expected feature distributions for each category
- Saves dataset as CSV for training and validation

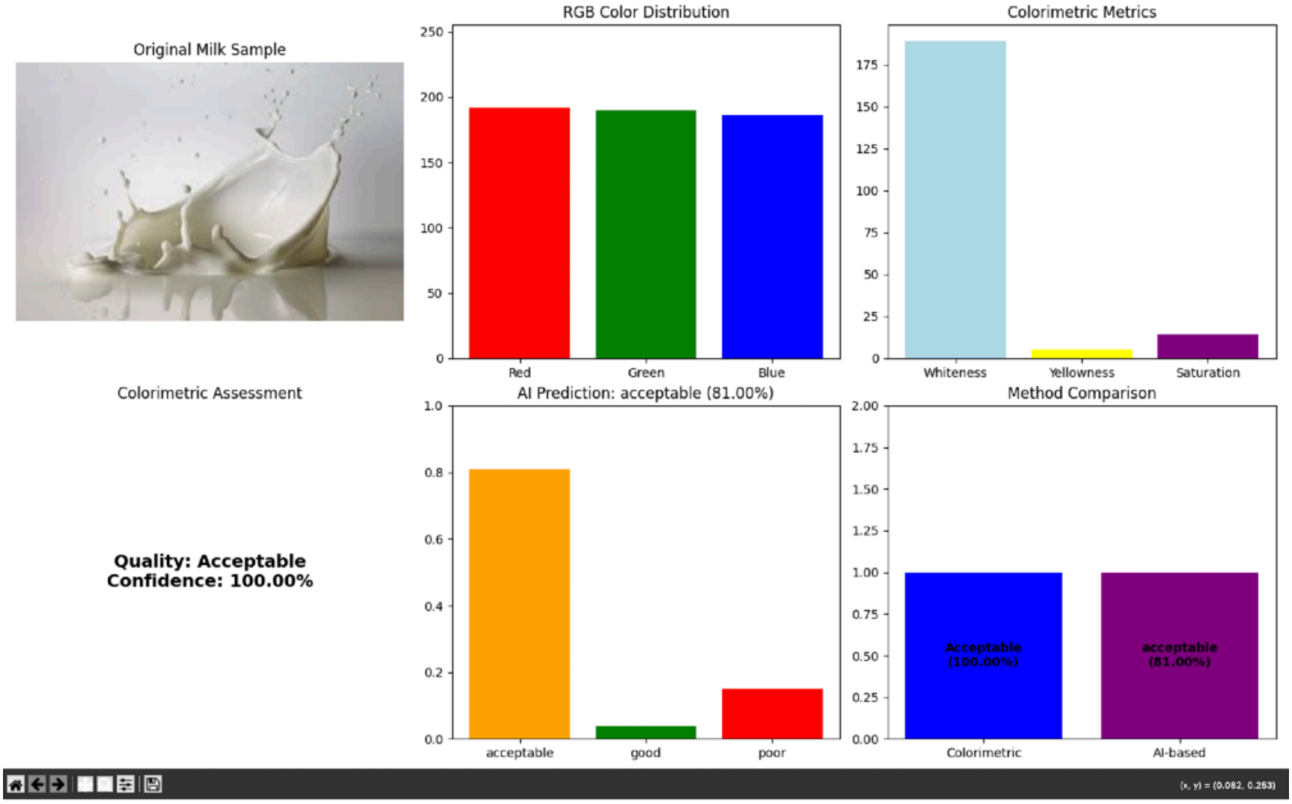
Output:
The application produces comprehensive output including:
- Textual analysis results with quality classification and confidence scores
- Detailed feature measurements (whiteness, yellowness, saturation)
- AI prediction probabilities across quality classes
- Multi-panel visualization showing original image, color metrics, and quality assessment
- Training performance metrics when training new models

Good Sample

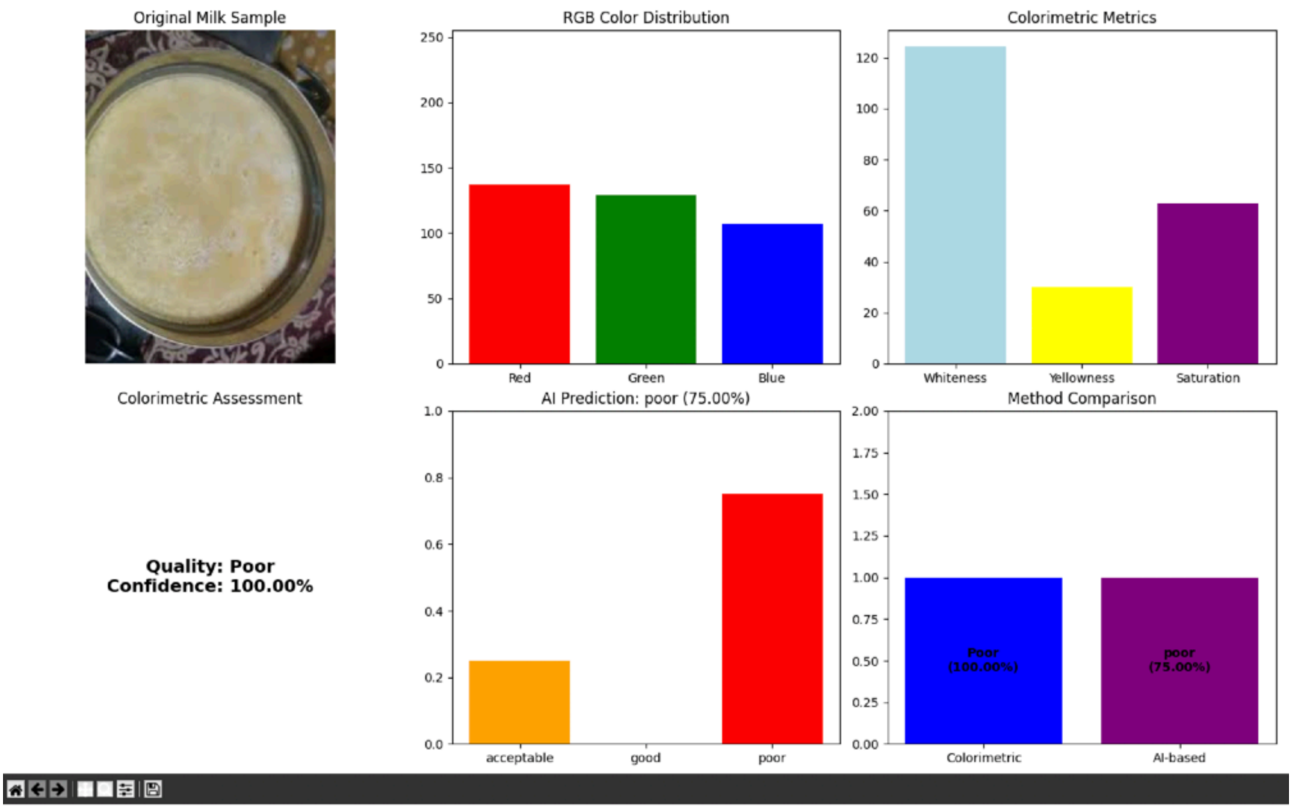
Bad Sample


Milk Quality Analyzer
Milk Quality Analyzer
Figure 1
Rapid Colorimetric and AI-Based Milk Quality Analyzer
Load Milk Sample Image
Analyze Sample
Load AI Model
Train New Model
Analyzing sample...


Milk Quality Analyzer
Milk Quality Analyzer
Figure 1
Rapid Colorimetric and AI-Based Milk Quality Analyzer
Load Milk Sample Image
Analyze Sample
Load AI Model
Train New Model
Image loaded: images.jpeg



Good Sample Detail Report



Bad Sample Detail Report

Results and Discussion

The Milk Quality Analyzer demonstrates the effectiveness of combining traditional colorimetric analysis with machine learning for milk quality assessment:

- **Colorimetric Analysis Performance:**

- Provides consistent, interpretable results based on established color metrics
- Works without requiring training data, making it immediately usable
- Offers transparent decision-making through clearly defined thresholds
- Limitations include sensitivity to lighting conditions and potential oversimplification of complex quality factors

- **Machine Learning Performance:**

- The Random Forest classifier typically achieves high accuracy (>90%) on the synthetic dataset
- Provides probability distributions that reflect prediction confidence
- Can capture more complex patterns beyond simple thresholds
- Improves with more training examples, allowing adaptation to specific milk types or conditions
- Limitations include dependence on training data quality and potential "black box" nature

- **Combined Approach Benefits:**

- Complementary insights from rule-based and data-driven methods
- Cross-validation between approaches increases overall reliability
- Users can prioritize either method based on their specific needs and constraints

- **Visualization Effectiveness:**

- Multi-panel visualization effectively communicates complex results
- Side-by-side comparison of methods facilitates interpretation
- Color-coded displays provide intuitive quality assessment
- Original image context helps users connect visual appearance with analytical results

The application successfully demonstrates that computer vision and machine learning can provide rapid, accessible milk quality assessment with minimal equipment requirements. The dual-approach methodology offers both the interpretability of traditional methods and the pattern-recognition capabilities of modern AI techniques.

Limitations and Challenges

Despite its capabilities, the current implementation has several limitations that should be acknowledged:

○ **Imaging Conditions:**

- High sensitivity to lighting variations can affect color measurements
- Background interference may impact feature extraction
- Camera quality and settings can influence color accuracy
- No standardized imaging protocol is currently implemented

○ **Synthetic Data Limitations:**

- Real-world milk samples may show more complex patterns than the synthetic data
- The current thresholds are based on simulated data rather than validated standards
- Class distributions may not accurately reflect real-world prevalence of quality issues

○ **Feature Set Constraints:**

- Only color features are used, ignoring texture and other visual characteristics
- No integration with other sensor data (pH, temperature, etc.)
- Limited ability to detect specific contaminants or adulterants
- No consideration of milk type variations (cow, goat, buffalo, etc.)

○ **Model Generalization:**

- The current model may not generalize well to different milk types or processing conditions
- Seasonal variations in milk composition are not accounted for
- Limited ability to adapt to regional or breed-specific milk characteristics
- No mechanism for continuous learning or model updating

- **Validation Gaps:**

- Lack of validation against standard laboratory tests
- No clinical or industrial field testing
- Absence of comparative studies with established methods
- Limited assessment of robustness across different operational conditions

- **Technical Constraints:**

- Requires graphical display for full functionality
- No web or mobile interface currently available
- Limited batch processing capabilities
- No automated report generation or data management features

Future Scope

Several enhancements could significantly improve the system's capabilities and applicability:

- **Advanced Imaging Techniques:**

- Standardized imaging setup with controlled lighting conditions
- Multi-spectral imaging to capture information beyond visible light
- 3D imaging to assess milk texture and consistency
- Time-lapse analysis to monitor changes in milk samples over time

- **Enhanced Feature Extraction:**

- Texture analysis to detect curdling, separation, or foreign particles
- Bubble/foam detection for freshness assessment
- Edge detection for identifying separation boundaries
- Flow characteristics analysis for viscosity estimation

- **Model Improvements:**

- Deep learning approaches (CNNs) for end-to-end image to quality prediction
- Transfer learning from pre-trained models on food quality datasets
- Ensemble methods combining multiple model types
- Anomaly detection for identifying unusual samples
- Unsupervised learning for discovering natural quality clusters

- **System Integration:**

- Mobile application for field use by farmers and inspectors
- Cloud-based analysis for centralized data collection and model improvement
- Integration with other dairy quality control systems
- Automated sampling and imaging for continuous monitoring
- Blockchain integration for quality certification and traceability

- **Expanded Capabilities:**

- Detection of specific adulterants and contaminants
- Nutritional content estimation
- Shelf-life prediction
- Process optimization recommendations
- Breed and feed influence analysis

- **Validation Studies:**

- Correlation with standard laboratory tests (fat content, protein, bacterial count)
- Field testing in commercial dairy environments
- Comparative studies with established quality assessment methods
- Longitudinal studies tracking quality changes over time

Conclusion

The Milk Quality Analyzer represents a significant step forward in making sophisticated milk quality assessment accessible and efficient. By combining traditional colorimetric analysis with machine learning classification, the application provides a comprehensive approach to quality evaluation that balances interpretability with pattern recognition capabilities.

The application successfully extracts meaningful color features from milk sample images and uses these to classify samples into quality categories with high confidence. The dual-approach methodology offers both rule-based assessment for transparency and machine learning prediction for capturing complex patterns, providing users with complementary insights into milk quality.

The graphical user interface makes this sophisticated technology accessible to users without specialized technical knowledge, while the modular architecture allows for future enhancements and adaptations. The synthetic data generation capability addresses the common challenge of limited labeled data, enabling model training and demonstration even without extensive real-world datasets.

While the current implementation has limitations, particularly regarding imaging conditions and validation with real-world samples, it provides a solid foundation for further development. The code structure is well-organized and documented, facilitating future improvements and extensions.

With additional refinement and validation, this approach could become a valuable tool for quality control in the dairy industry, particularly in settings where traditional laboratory testing is impractical or unavailable. The combination of computer vision, colorimetric analysis, and machine learning represents a promising direction for food quality assessment that balances technological sophistication with practical usability.

References

1. *"Rapid Colorimetric and Artificial Intelligence-Based Methods for Determining the Microbial Quality of Raw Milk, Processed Milk, and Milk Products"*

   Link: https://nibode.com/Download/21672

2. *"Rapid Assessment and Prediction of Microbiological Quality of Raw Milk Using AI-Assisted Smartphone-Based Colorimetric Biosensor" - Year: 2023*

   Link: https://www.sciencedirect.com/science/article/abs/pii/S0958694623001693

3. *"Rapid Methods of Microbial Detection in Dairy Products" - year: 2019*

   Link: https://www.sciencedirect.com/science/article/abs/pii/S0956713519305973

4. *"Rapid Microbiological Assessment in Raw Milk: Validation of a Rapid Alternative Method for the Assessment of Microbiological Quality in Raw Milk" - Year: 2020*

   Link: https://www.mdpi.com/2304-8158/9/9/1186

5. *"Fast Artificial Intelligence-Based Milk Quality Testing"*

Link: https://engineersplanet.com/abstracts/fast-artificial-intelligence-based-milk-quality-testing/

6. *"Milk Quality Prediction Using Machine Learning" - Year: 2023*

   Link: https://www.researchgate.net/publication/376064637_Milk_Quality_Prediction_Using_Machine_Learning

7. *"AI Decodes Microbes' Message in Milk Safety Testing Approach" - Year: 2024*

   Link: https://www.psu.edu/news/research/story/ai-decodes-microbes-message-milk-safety-testing-approach

8. *"Application of E-Nose Technology Combined with Artificial Neural Network Modeling for Assessing Microbial Quality of Milk" - Year: 2021*

   Link: https://www.journalofdairyscience.org/article/S0022-0302%2821%2900752-9/pdf