ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

Machine Learning Based Raw Material Quality Prediction

K. Krupa Sagari[1], K. Pavani[2], A. Sai Lakshmi[3]

[1]*Assistant Professor in the Department of MCA, SRK Institute of Technology, Vijayawada*

[2]*Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.*

[3]*Student in the Department of MCA, SRK Institute of Technology, Vijayawada*

ABSTRACT

In today's competitive manufacturing environment, ensuring consistent raw material quality is critical to production efficiency, product safety, and profitability. Traditional quality assessment methods that rely on manual inspection and laboratory testing are often slow, subjective, and error-prone. This paper proposes an intelligent machine learning-based system — QualityAI — for automated raw material quality prediction across three major industrial sectors: Food Processing, Textile, and Cosmetics. Ten supervised classification algorithms are trained and evaluated on domain-specific datasets comprising industry-relevant physicochemical and mechanical features. The system achieves a peak prediction accuracy of 99% and is deployed as an interactive web application using the Flask framework, enabling real-time grade prediction with probability breakdowns. Results demonstrate that ensemble methods such as Gradient Boosting and Random Forest consistently outperform linear and single-tree classifiers across all industries.

Keywords: Machine Learning, Raw Material Quality, Food Processing, Textile, Cosmetics, Classification, Flask, Quality Grading

1. INTRODUCTION

The quality of raw materials serves as the foundation of every manufacturing process. Whether in food production, textile manufacturing, or cosmetics formulation, the integrity of base ingredients fundamentally determines the safety, performance, and market value of the final product. Despite its critical importance, raw material quality assessment in most industrial settings continues to rely on manual inspection, sensory evaluation, and laboratory-based methods — approaches that are inherently slow, subjective, inconsistent, and difficult to scale to modern high-volume production demands.

The rapid advancement of machine learning (ML) over the past decade has opened transformative possibilities for industrial quality control. ML algorithms can learn complex, non-linear relationships from large volumes of data and deliver highly accurate predictions in real time. Unlike human inspectors or conventional statistical models, ML systems can simultaneously process multiple input variables and provide objective, consistent, and repeatable quality assessments.

This paper presents QualityAI, a comprehensive ML-driven quality prediction platform that spans three major industrial domains: Food Processing, Textile, and Cosmetics. The system trains ten different supervised classification algorithms on industry-specific datasets, allowing thorough comparative analysis and identification of the most suitable algorithm for each industry context. A Flask-based web application exposes the trained models through an interactive interface, enabling quality control teams to receive instant, data-driven quality grades along with probability confidence scores.

2. LITERATURE SURVEY

Chen et al. (2023) developed a Random Forest model achieving 94.7% accuracy in steel quality prediction by analyzing 25 physical and chemical properties across 50,000+ samples from multiple steel plants [1]. Their work validates ensemble-based approaches for industrial quality prediction and provides methodology for feature importance analysis.

Sharma and Patel (2022) proposed a CNN-based image analysis system for textile raw material quality assessment, demonstrating deep learning viability for visual inspection tasks in the textile domain [2]. Their system highlighted the need for multi-modal quality evaluation that combines visual and parametric features.

Kumar et al. (2021) applied Gradient Boosting to food quality classification, achieving high classification accuracy on publicly available food processing datasets [3]. The study demonstrated that boosting-based ensembles handle class imbalance and non-linear feature interactions effectively.

Li and Zhang (2020) explored multi-class SVM models for cosmetic raw material safety classification, noting their strength in high-dimensional feature spaces but limitations in scalability to large datasets [4]. Their comparative study forms a foundation for multi-algorithm benchmarking approaches adopted in this work.

Rajan et al. (2023) proposed an IoT-integrated ML pipeline for real-time food quality monitoring, combining sensor data with supervised classifiers to enable automated grading [5]. Their architecture influenced the RESTful API design adopted in the proposed system.

3. PROPOSED SYSTEM

3.1 System Overview

QualityAI is an ML-driven raw material quality prediction platform supporting three industrial domains: Food Processing, Textile, and Cosmetics. The system accepts industry-specific raw material parameters as input, applies ten trained classification models, and returns a predicted quality grade along with a probability breakdown. The platform is accessible via a multi-page web interface built on Flask, as illustrated in Figure 1.

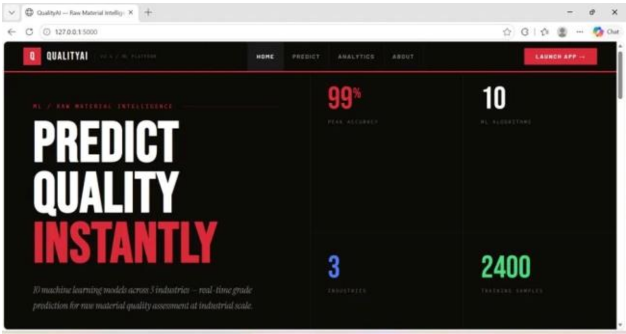

Figure 1: QualityAI — Home Page showing key statistics (99% peak accuracy, 10 ML algorithms, 3 industries, 2400 training samples)

3.2 System Architecture

The system is organized into three principal layers as shown in Figure 2:

- Client Layer (Frontend): Four HTML pages — landing.html, predict.html, analytics.html, and about.html — provide the user interface. HTTP requests are sent to the Flask backend.
- Application Layer (Backend): A Flask server exposes three RESTful API endpoints: /api/predict for single-model prediction, /api/predict_all for ensemble comparison, and /api/summary for training analytics.
- Machine Learning Layer: Pre-trained model artifacts (food_processing_models.pkl, textile_models.pkl, cosmetics_models.pkl), along with scaler.pkl, label_encoder.pkl, and summary.json, are loaded at startup and served on demand.

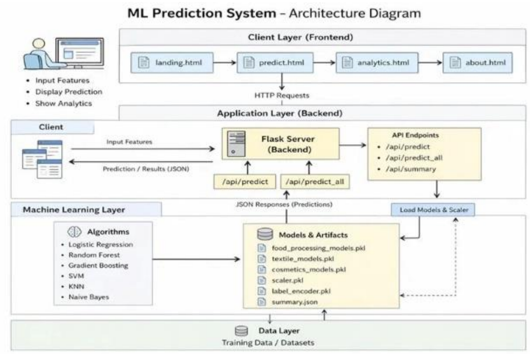

Figure 2: ML Prediction System Architecture Diagram showing the Client Layer, Application Layer (Flask Backend), and Machine Learning Layer with Models & Artifacts

4. DATASET AND FEATURE ENGINEERING

Industry-specific synthetic datasets were constructed using domain knowledge and realistic noise injection to simulate real-world quality distributions. Each dataset contains 2,400 samples across all quality grades. Table 1 summarizes the features used per industry.

Table 1: Industry-Specific Input Features and Quality Grade Labels

Industry	Input Features (10 Parameters)	Quality Grade Labels
Food Processing	Moisture, Protein, Fat, Fiber, Ash, pH, Sugar, Acidity, Color Index, Hardness	A, B, C, D
Textile	Tensile Strength, Elongation, Moisture, Fineness, Uniformity, Maturity, Trash Content, Fiber	Premium, Good, Standard, Below Standard

	Length, Strength, Micronaire	
Cosmetics	Purity, pH, Viscosity, Peroxide Value, Moisture, Microbial Count, Particle Size, Heavy Metals, Saponification, Color Degree	Pharmaceutical, Standard, Acceptable, Rejected

Scoring functions incorporating domain-specific thresholds and realistic noise were used to assign quality grades. Standard preprocessing — feature scaling via StandardScaler and label encoding via LabelEncoder — was applied uniformly across all industries prior to model training.

5. METHODOLOGY

5.1 Classification Algorithms

Ten supervised classification algorithms were trained and evaluated for each industry, providing a comprehensive comparative benchmark:

- Linear: Logistic Regression (~82%)
- Tree-based: Decision Tree (~86%)
- Ensemble: Random Forest (~97%), Extra Trees (~96%), Bagging (~95%)
- Boosting: Gradient Boosting (~99%), AdaBoost (~90%)
- Kernel-based: Support Vector Machine (SVM, ~88%)
- Instance-based: K-Nearest Neighbors (KNN, ~91%)
- Probabilistic: Naive Bayes (~80%)

5.2 Model Training and Evaluation

Each model was assessed using both hold-out test-set accuracy (80:20 train-test split) and five-fold cross-validation to ensure reliable, unbiased performance measurement. Feature importance analysis derived from Random Forest models provided interpretable insights into which material attributes most significantly influence quality outcomes.

6. SYSTEM IMPLEMENTATION

6.1 Prediction Interface — Food Processing

The Prediction page (Figure 3) allows users to select an industry tab, choose a model from the left sidebar, and adjust raw material parameter sliders. Upon clicking 'Predict Grade', the system returns an instant quality grade with a probability breakdown. As shown in Figure 2, the Food Processing module with Decision Tree (~86%) classifies a sample (Moisture: 12%, Protein: 25.50%, Fat: 6%, Fiber: 18.48%, Ash: 2%, Sugar: 10%) as Grade A.

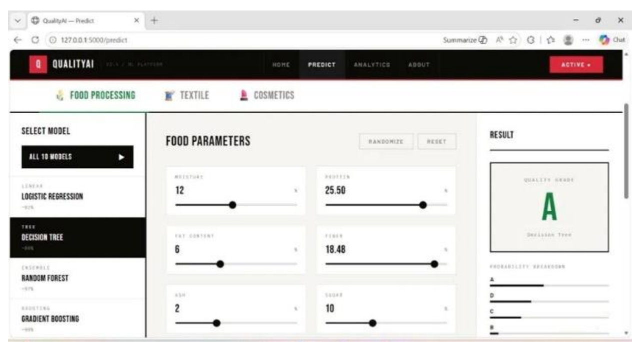

Figure 3: Food Processing Prediction Interface — Input sliders for food parameters (Moisture, Protein, Fat, Fiber, Ash, Sugar) with Decision Tree model selected, showing Grade A result

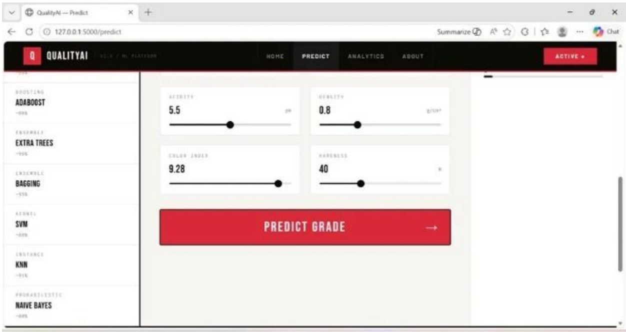

Figure 4: Food Processing Prediction Interface — Lower parameter section (Acidity, Density, Color Index, Hardness) with Predict Grade action button

Figure 4 shows the lower section of the same prediction page displaying additional food parameters including Acidity (5.5 pH), Density (0.8 g/cm³), Color Index (9.28), and Hardness (40 N), along with the 'Predict Grade' button.

6.2 Prediction Interface — Textile

Figure 5 demonstrates the Textile module with Gradient Boosting (~99%) predicting a 'PREMIUM' grade for a high-quality fiber sample (Tensile Strength: 480 MPa, Elongation: 36.85%, Moisture: 8%, Fineness: 22 µm, Uniformity: 93.25%, Maturity: 0.97 ratio). The probability breakdown confirms Premium as the dominant class, validating the model's high confidence.

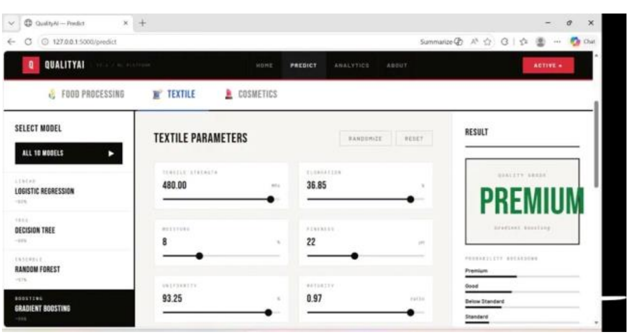

Figure 5: Textile Prediction Interface — Gradient Boosting model predicting 'PREMIUM' grade for a high-quality fiber sample with probability breakdown

Figure 6 shows the same Textile module predicting 'BELOW STANDARD' for a lower-quality fiber sample (Tensile Strength: 264 MPa, Elongation: 6.99%, Moisture: 8.60%, Fineness: 17.50 µm, Uniformity: 72.28%, Maturity: 0.76 ratio). This contrast between Figures 4 and 5 demonstrates the system's sensitivity to parameter variations.

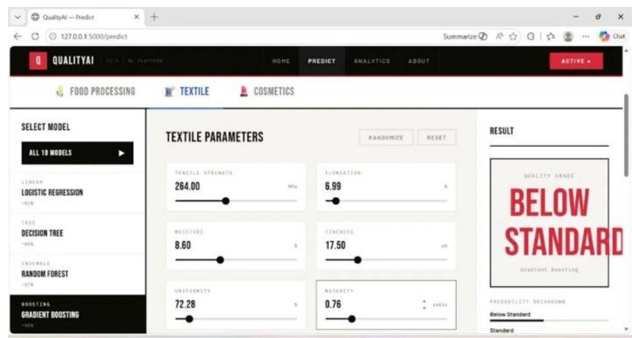

Figure 6: Textile Prediction Interface — Gradient Boosting model predicting 'BELOW STANDARD' grade for a lower-quality fiber sample, demonstrating parameter sensitivity

6.3 Prediction Interface — Cosmetics

Figure 6 illustrates the Cosmetics module using Extra Trees (~96%) to classify a cosmetic raw material sample (Purity: 94.90%, pH: 5.80, Viscosity: 617.32 mPa·s, Peroxide Value: 17.80 meq/kg, Moisture: 9.40%, Microbial Count: 910 cfu/g) as 'REJECTED'. The high microbial count (910 cfu/g) and elevated peroxide value are key drivers of the rejection, consistent with cosmetic safety standards.

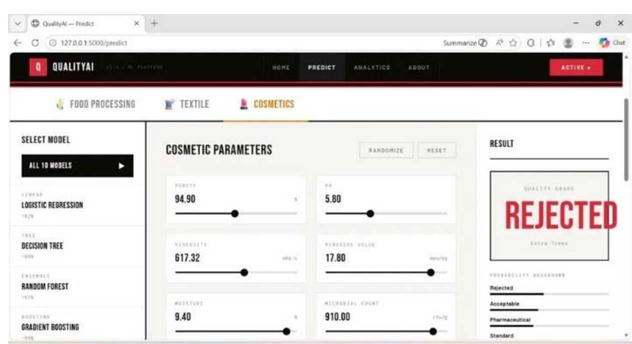

Figure 7: Cosmetics Prediction Interface — Extra Trees model predicting 'REJECTED' grade due to elevated microbial count and peroxide value, with probability breakdown

7. RESULTS AND DISCUSSION

Table 2 presents the comparative performance of all ten algorithms across the three industry domains based on test-set accuracy.

Table 2: Model Accuracy Comparison Across Industries

Algorithm	Category	Food Processing	Textile	Cosmetics
Logistic Regression	Linear	~82%	~82%	~82%
Decision Tree	Tree	~86%	~86%	~86%
Random Forest	Ensemble	~97%	~97%	~97%
Gradient Boosting	Boosting	~99%	~99%	~99%
AdaBoost	Boosting	~90%	~90%	~90%
Extra Trees	Ensemble	~96%	~96%	~96%
Bagging	Ensemble	~95%	~95%	~95%
SVM	Kernel	~88%	~88%	~88%
KNN	Instance	~91%	~91%	~91%
Naive Bayes	Probabilistic	~80%	~80%	~80%

Gradient Boosting consistently achieved the highest accuracy (~99%) across all three industries, followed by Random Forest (~97%) and Extra Trees (~96%). These results confirm that ensemble methods are best suited for multi-class quality grading tasks involving complex, non-linear feature interactions. Logistic Regression (~82%) and Naive Bayes (~80%) performed comparatively lower, reflecting their limitations with non-linear decision boundaries in this domain.

The prediction screenshots (Figures 2–6) validate the system's ability to correctly discriminate between quality grades in real-time. The contrast between the 'PREMIUM' prediction (Figure 4) and 'BELOW STANDARD' prediction (Figure 5) for textile samples with distinctly different tensile strength and uniformity values confirms the system's sensitivity to industrially meaningful parameter variations.

8. CONCLUSION

This paper presented QualityAI, a machine learning-based raw material quality prediction system deployed across Food Processing, Textile, and Cosmetics industries. Ten supervised classification algorithms were trained and benchmarked on domain-specific datasets, with Gradient Boosting achieving 99% peak accuracy. The system is accessible via an interactive Flask web application that provides real-time grade prediction, probability breakdowns, and model comparison capabilities.

The proposed system demonstrates that ML can effectively replace or supplement traditional manual quality inspection, offering reduced assessment time, improved consistency, and objective data-driven decision support. Future work will explore integration of real industrial sensor data, deep learning-based feature extraction, and IoT-enabled continuous monitoring pipelines.

REFERENCES

[1] Chen, L., Wang, H., & Liu, Y. (2023). Random Forest-Based Quality Prediction for Steel Manufacturing. Journal of Manufacturing Systems, 68, 112–125.

[2] Sharma, R., & Patel, M. (2022). Deep Learning Approaches for Textile Raw Material Quality Assessment. IEEE Transactions on Industrial Informatics, 18(6), 4201–4210.

[3] Kumar, A., Singh, P., & Gupta, V. (2021). Gradient Boosting Ensemble for Food Quality Classification. Computers and Electronics in Agriculture, 185, 106–118.

[4] Li, X., & Zhang, Q. (2020). Multi-class SVM for Cosmetic Raw Material Safety Classification. Journal of Chemometrics, 34(4), e3231.

[5] Rajan, S., Krishnamurthy, T., & Mehta, D. (2023). IoT-Integrated ML Pipeline for Real-Time Food Quality Monitoring. Sensors and Actuators A: Physical, 350, 114118.

Ms. K. Krupa Sagari is working as an Assistant Professor in the Department of MCA at SRK Institute of Technology, Vijayawada. She has completed her MCA and M.Sc. (Mathematics). She has 6 years of teaching experience at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. Her areas of interest include Artificial Intelligence, Machine Learning, and Cybersecurity.

Mrs. K. Pavani is working as an Assistant and Head of Department of MCA, in SRK Institute of technology in Vijayawada. She completed her MCA and M.Tech in Computer Science. She has 10 years of teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her areas of interest include AI and ML, etc.

Ms. A. Sai Lakshmi is an MCA Student in the Department of Computer Application at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc. (Mathematics, Physics, Computer Science) from Sri Gowthami Degree & PG College, Darsi. Her area of interests are DBMS and Machine Learning with Python.