International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

Check for updates

Predicting food adulterants in milk using Support Vector Machine (SVM)

Prakash R *, Shantha kumar S kapse, Umar Ameen and Chandan Hegde

Department of MCA Surana College (Autonomous) Bangalore, Karnataka, India.

International Journal of Science and Research Archive, 2024, 13(01), 2257–2268

Publication history: Received on 19 August 2024; revised on 02 October 2024; accepted on 05 October 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.1.1848

Abstract

Milk adulteration is a significant global issue, particularly in emerging nations, where inadequate monitoring and unhygienic conditions prevail. The adulteration of milk with various chemicals, such as urea, water, skimmed milk powder, sugar, and detergent, poses serious health risks, including heart problems, diarrhea, CNS disorders, irritation, and gastrointestinal disorders. Traditional detection methods are labor-intensive and require sophisticated equipment, which limits their practical application. This study aims to develop a machine learning-based approach to detect milk adulteration using attributes like Solids- Not-Fat (SNF), fat, Corrected Lactometer Reading (CLR), Total Solids (TS), temperature, and protein content. Various machine learning models were employed and evaluated for their performance, including Logistic Regression, Decision Trees, SVM, and Random Forests. The findings demonstrate that machine learning can effectively identify adulteration types, providing a foundation for the dairy industry's practical and automated detection systems. This research comprehensively reviews common milk adulterants and highlights advanced detection methods to ensure milk quality and safety.

Keywords: SVM; Solids-Not-Fat (SNF); Corrected Lactometer Reading (CLR); Total Solids (TS); Temperature; Protein content

1. Introduction

India holds a significant position in global milk production, contributing 17% of the world's supply. It accounts for 90% of the global buffalo milk production and ranks second in cow's milk production, with an output of 54 million metric tons, just behind the USA. Despite being the largest exporter of milk and skimmed milk powder, India still imports certain milk varieties [1].

The production and export of milk in India are managed through an efficient supply chain that connects various companies, facilitating distribution. Over six billion people worldwide consume milk and its products. Milk is composed of various components including lactose, water, fat, protein, milk sugar, and minerals, with trace amounts of phospholipids, vitamins, and enzymes [2].

Rajasthan, Maharashtra, Uttar Pradesh, Madhya Pradesh, Punjab, Haryana, Karnataka, and Andhra Pradesh are among the states with the highest milk production rates [1]. However, the quality of milk has been compromised over time due to adulteration, which involves adding cheaper substances to increase quantity and profit. Water is a common adulterant, reducing the milk's nutritional value and potentially introducing water-borne diseases if the added water is contaminated.

To better understand milk adulterants and their impact on human health, research has been conducted to assess the quality of raw milk.

* Corresponding author: Prakash R

Copyright © 2024 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

Milk is a vital source of essential nutrients such as calcium, vitamins, and proteins, making it a staple in diets worldwide. However, the integrity of milk has been compromised by adulteration, where substances are added to increase quantity and extend shelf life. Common adulterants include water, sugar, detergent, and urea, posing significant health risks, particularly to children and the elderly. Traditional methods for detecting these adulterants involve chemical analysis and physical testing, which are not only time-consuming but also require specialized laboratory equipment and trained person- nel.

Figure 1 State-wise milk production contribution

The advent of machine learning offers a promising alternative, leveraging computational power to analyze complex datasets and identify patterns indicative of adulteration. This research aims to explore the application of machine learning techniques in detecting milk adulteration based on key milk attributes: Solids-Not-Fat (SNF), fat, Corrected Lactometer Reading (CLR), Total Solids (TS), temperature, Ph value, and protein content. By developing and evaluating various machine learning models, this study seeks to provide a reliable and efficient method for ensuring milk quality and safety.

1.1. Existing quality assurance techniques

Traditional methods for detecting milk adulteration have primarily relied on chemical and physical tests [3]. These methods, while accurate, are often limited by their need for sophisticated laboratory equipment and the expertise required to conduct and interpret results. For instance, the lactometer test measures the density of milk to detect water adulteration, while chemical reagents can be used to identify the presence of starch, detergent, or urea. However, these methods are not only time-consuming but also impractical for widespread, routine testing in the dairy industry [4].

Machine learning, on the other hand, has shown significant promise in various fields of food quality control, including the detection of contaminants and the classification of food types [5]. Studies have demonstrated the effectiveness of machine learning algorithms in identifying adulterants in other food products such as honey, olive oil [6], and spices.

Despite these advancements, there is a notable gap in the literature specifically addressing the use of machine learning for milk adulteration detection. This study aims to fill this gap by systematically evaluating the performance of different machine learning models in detecting various types of milk adulteration using individual milk attributes.

Aim and objectives

The primary objective of this research paper is to develop and validate machine learning models for the efficient and accurate detection of milk adulteration based on key milk attributes, to provide a reliable and automated system for ensuring milk quality and safety in the dairy industry.

2. Methodology

2.1. Dataset collection

The dataset analyzed in this study includes 1000 milk samples gathered over five consecutive days, with collections made both in the morning and evening from distinct dairy societies. Each sample has been classified as either ”Good” or ”Bad” based on a thorough quality assessment. The primary attributes measured are Solids-Not-Fat (SNF), Fat, Corrected Lactometer Reading (CLR), total solids (TS), temperature, protein content, pH, and the quantity in liters. These attributes were specifically selected due to their significance in milk composition and their potential to reveal signs of adulteration.

Below is the table detailing the locations of the dairy societies and the number of milk samples collected each day:

Table 1 Data collection details

SLNO	Dairy name	No of Sample
1	Channenahalli [599] Milk Producers Cooperative Society Ltd (Under BMUL)	188
2	Channenahalli [599] Milk Producers Cooperative Society Ltd (Under BMUL)	192
3	Channenahalli [599] Milk Producers Cooperative Society Ltd (Under BMUL)	185
4	Channenahalli [599] Milk Producers Cooperative Society Ltd (Under BMUL)	190
5	Channenahalli [599] Milk Producers Cooperative Society Ltd (Under BMUL)	188

The table below outlines the key attributes measured in each milk sample:

Table 2 Attribute details

Attribute	Description
Sample ID	Unique identifier for each sample
Source	The origin of the milk sample (e.g., Source A, Source B)
SNF (%)	Solids-Not-Fat content
Fat (%)	Fat content percentage
CLR	Corrected Lactometer Reading
TS (%)	Total solids percentage
Temperature (°C)	Temperature at the time of collection
Protein (%)	Protein content percentage
Quantity (L)	Volume in liters
Ph	Ph value at the time of collection
Result	Quality assessment (Good/Bad)

2.2. Machine Learning Models

- SVM
- Decision Trees
- Logistic Regression
- Convolutional Neural Networks (CNNs)

Given the dataset containing features like Quantity, Fat, CLR, SNF, Protein, ST, pH, Temperature, Rate, and Amount, SVM is the best choice compared to other models. SVMs are adept at handling complex, non-linear data and creat- ing effective decision boundaries, making them superior for this dataset. Unlike CNNs, which require extensive data and computational power, and traditional methods like Linear Regression and Logistic Regression, which have limitations with linear relationships and multicollinearity, SVMs provide robust performance without overfitting, unlike Decision Trees. Thus, SVMs offer the most reliable and efficient solution for this dataset [6] [7].

2.3. SVM model

Support Vector Machines (SVMs) are versatile machine- learning algorithms designed for both classification and regression tasks. Their unique execution approach sets them apart from other algorithms, making them popular for handling both continuous and categorical variables.

An SVM model functions by representing different classes in a hyperplane within a multidimensional space. This hyperplane is iteratively adjusted by the SVM to minimize errors. The primary aim of SVM is to partition the dataset into distinct classes while finding the optimal separating hyperplane [8].

- Support Vectors are the data points closest to the hy- perplane, and they are crucial in defining the separating boundary.
- Hyperplane:This is a decision plane that separates dif- ferent classes within the dataset. It acts as the dividing line or space between various class groups.
- Margin: This term refers to the distance between the two closest lines or planes that separate the classes. It can be measured as the perpendicular distance from the line to the support vectors. A larger margin indicates a better separation between classes, while a smaller margin suggests less effective separation

The goal of an SVM is to find the maximum-margin hyperplane (MMH) that divides the classes most effectively. This is achieved through two main steps:

- Iterative Hyperplane Generation: SVM generates mul- tiple hyperplanes to find the best separation between classes.
- Optimal Hyperplane Selection: The hyperplane that provides the most effective separation of the classes is then selected.

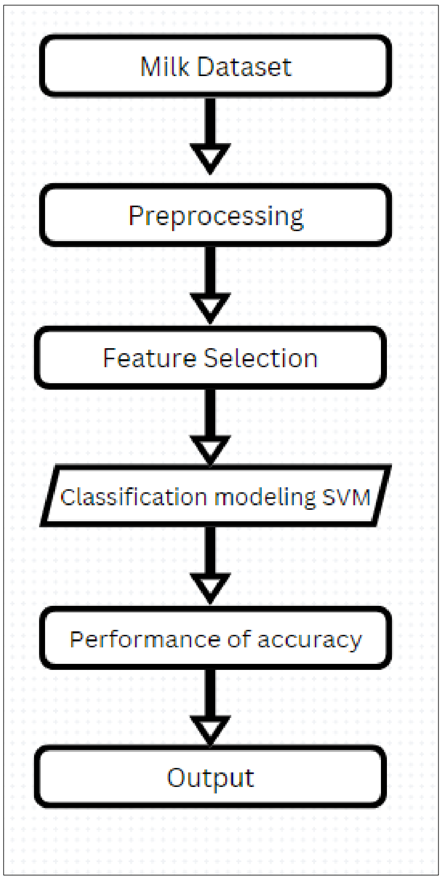

Figure 2 Flowchart of model working

2.4. Feature Selections

Feature selection is a critical step in machine learning, particularly for developing accurate models for detecting milk adulteration. The choice of features directly impacts the model's performance by influencing its ability to distinguish between adulterated and pure milk. Below, we outline the features selected for various types of milk adulteration and the rationale behind their inclusion.

In high-quality milk, specific average values are key indi- cators of its purity and nutritional worth. Generally, Solids-Not-Fat (SNF) content averages around 8.5%, Fat content is about 3.5%, and the Corrected Lactometer Reading (CLR) is approximately 26. Moreover, the average protein content is around 3.2%, the pH level typically ranges from 6.6 to 6.8, and the Total Solids (TS) content is about 12.5%. These averages ensure that the milk meets the essential standards for safety and quality [9] [10].

- Feature Collection for Water Adulteration: SNF, CLR, Fat, pH [11] [12] [13]
  - SNF (Solid-Not-Fat): This parameter is crucial for de- tecting water adulteration as the presence of excess water dilutes the solid content in milk. A decrease in SNF typically indicates the addition of water [1].
  - CLR (Corruption Level Ratio): CLR measures the ratio of various milk components. A lower CLR may indicate dilution with water, affecting the milk's consistency and composition.
  - Fat: Fat content is a key indicator of milk quality. Water adulteration usually results in a reduced fat percentage, making this feature essential for detection [14]
  - pH: The pH level can change with water addition, as water can alter the acidity of milk. Monitoring pH helps in identifying potential anomalies in milk composition.
- Feature Collection for Urea Adulteration [2] [15]: SNF, ST (Specific Temperature), pH, Protein [11] [12] [13]
  - SNF (Solid-Not-Fat): Urea adulteration can lead to changes in the SNF content. Urea increases the SNF levels artificially, which can be detected by analyzing deviations in this feature.
  - ST (Specific Temperature): Urea affects the thermal properties of milk. Specific temperature readings can help in identifying inconsistencies caused by urea.

- pH: Urea addition can alter the pH of milk. Monitor- ing pH levels helps in detecting any unusual deviations indicative of urea presence.
- Protein: Urea affects the protein content in milk. By analyzing protein levels, one can detect abnormalities caused by urea adulteration.

- Feature Collection for Sugar Adulteration [2]: SNF, CLR [11] [12] [13]
  - SNF (Solid-Not-Fat): The addition of sugar can affect the SNF levels in milk. Increased sugar content often leads to changes in SNF, making it a useful feature for detection [16].
  - CLR (Corruption Level Ratio): Sugar adulteration can alter the ratio of milk components, which is reflected in CLR. Monitoring CLR helps in identifying deviations caused by sugar.
- Feature Collection for Detergent Adulteration [2]: SNF, CLR, pH [11] [12] [13]
  - SNF (Solid-Not-Fat): Detergent adulteration impacts the SNF levels by altering the milk's composition. Monitor- ing SNF helps in detecting such adulteration.
  - CLR (Corruption Level Ratio):Detergents can change the composition of milk, which affects CLR. Anomalies in CLR readings can indicate detergent presence.
  - pH: Detergents can cause changes in the pH level of milk. Tracking pH variations helps in identifying detergent contamination.

3. Results

3.1. Comprehensive Data Depiction

3.1.1. Water Dilution Deception

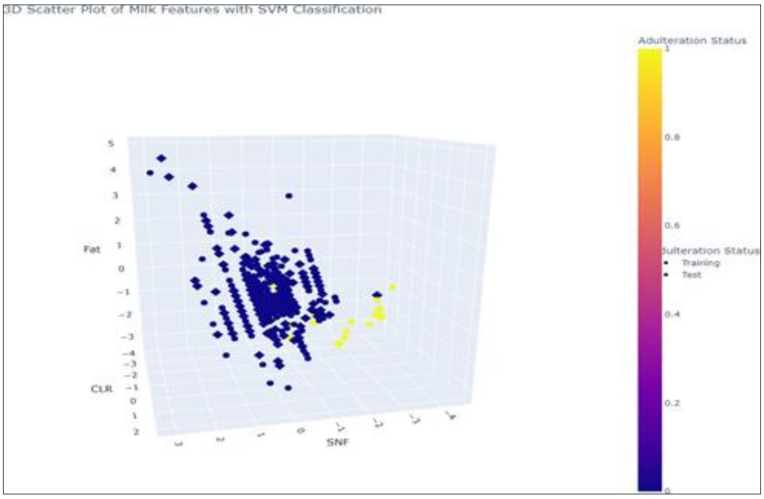

Figure 3 The 3D plot of milk Adulteration

The graphs illustrate the detection of water adulteration in milk using an SVM model, focusing on key parameters such as SNF, CLR, Fat, and pH. The first scatter plot shows the relationship between Fat and pH, where the adulteration status is color-coded. Here, most adulterated samples (status

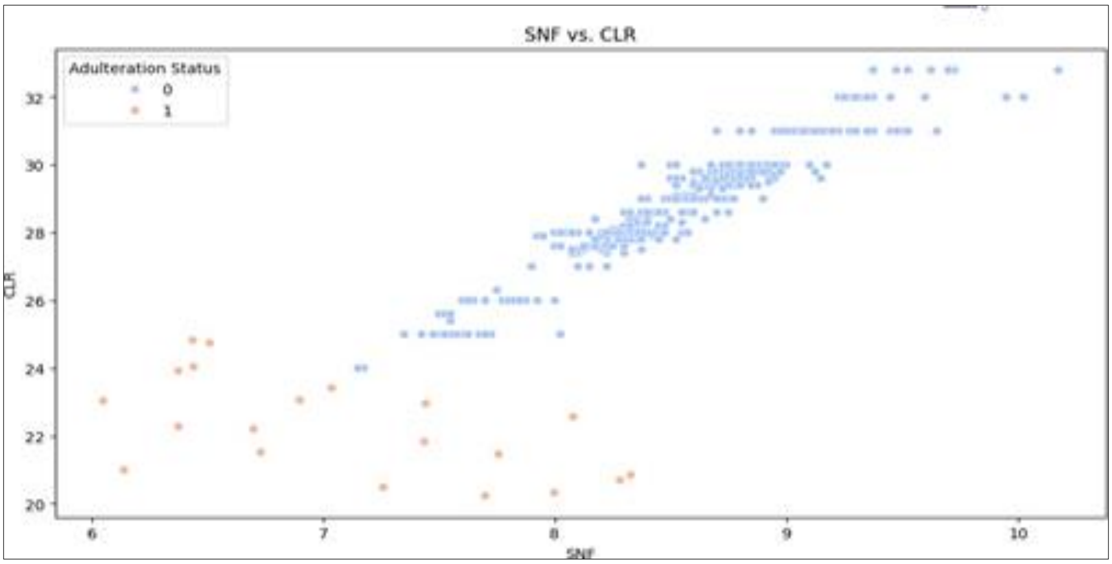

Figure 4 2D Scatter Plot for SNF vs. CLR

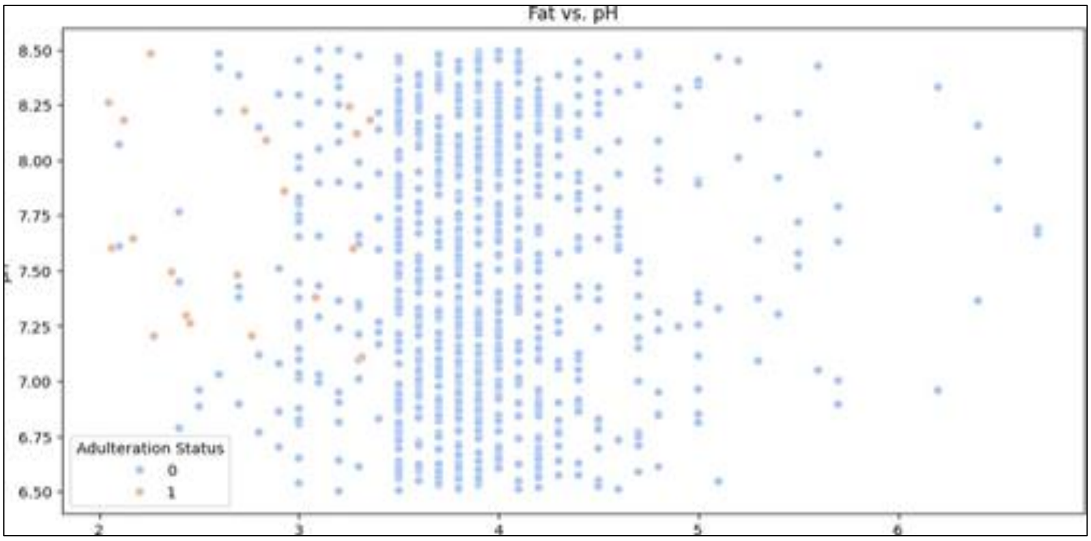

Figure 5 2D Scatter Plot for Fat vs. pH

- Cluster at lower pH and Fat values, indicating a significant deviation from the norm. The second plot, depicting SNF versus CLR, reveals a clear separation between adulterated and non-adulterated samples, with adulterated ones falling below the threshold values (SNF ¡ 8.5, CLR ¡ 26). The 3D scatter plot further confirms these findings, providing a spatial represen- tation where adulterated samples are distinctly separated from non-adulterated ones based on SNF, CLR, and Fat attributes. These visualizations effectively highlight the SVM model's ability to distinguish adulterated milk by identifying patterns and anomalies in key compositional parameters.

3.1.2. Urea Contamination Deception

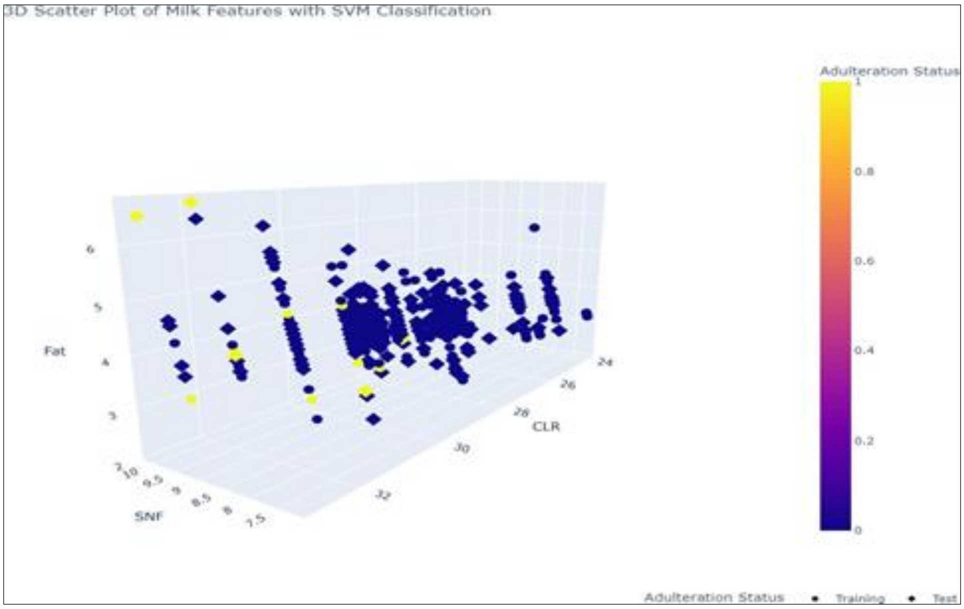

Figure 6 The 3D plot of Urea Adulteration

The pie chart demonstrates that 95.1% of the milk sam- ples are non-adulterated, while only 4.9% are adulterated, highlighting a significant class imbalance in the dataset. This imbalance is critical to consider when interpreting the model's performance. The 3D scatter plot focuses on the SVM model's classification based on SNF, CLR, and Fat attributes. The plot shows a clear distinction between adulterated and non- adulterated samples, with non-adulterated samples (dark blue) clustered separately from adulterated ones (yellow). Adulter- ated samples generally exhibit lower SNF and CLR values with varied Fat content. Despite the imbalance, the SVM model effectively identifies adulteration patterns, though some overlap suggests areas for potential misclassification. These visualizations collectively emphasize the model's strength in capturing key features for milk adulteration detection.

3.1.3. Detergent Contamination Deception

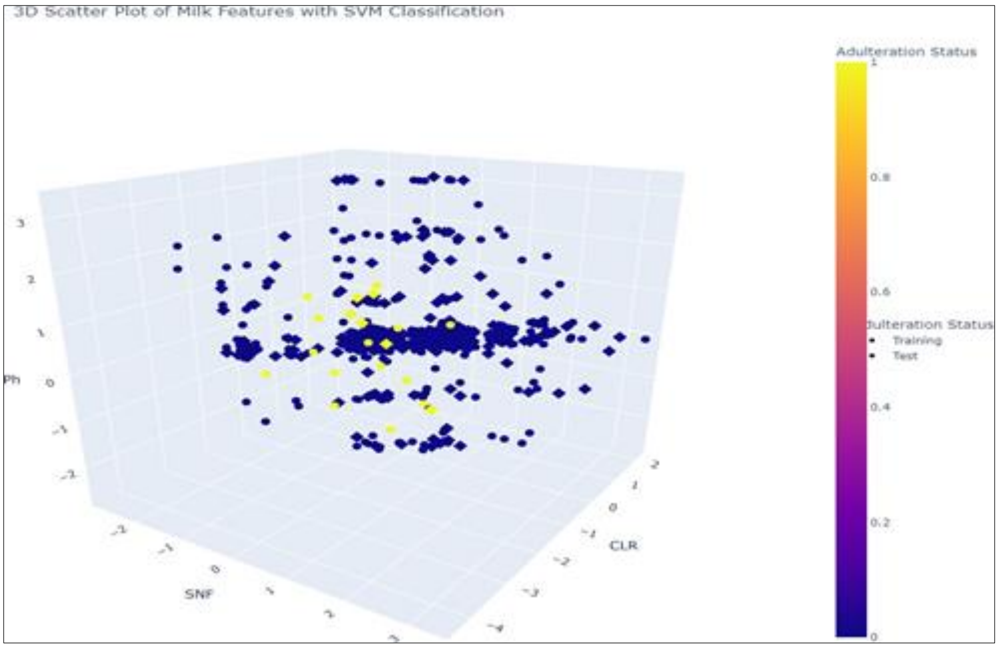

Figure 7 The 3D plot of Detergent Adulteration

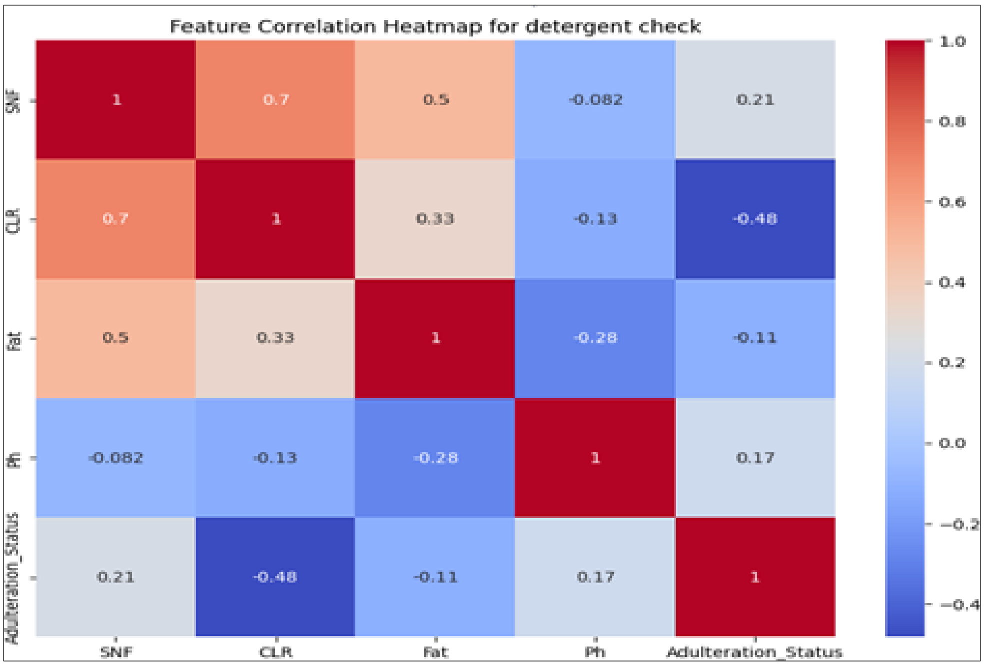

Figure 8 Feature correlation heatmap for detergent check

The feature correlation heatmap shows the relationships between various features such as SNF, CLR, Fat, and Ph in determining the Adulteration Status. Notably, there are positive correlations between SNF and CLR (0.7) and SNF and Fat (0.5), whereas CLR and Adulteration Status have a notable negative correlation (-0.48). The 3D scatter plot visualizes the milk features (SNF, CLR, Ph) with SVM classification, distinguishing between training and test data points. Adulteration status is indicated by color intensity, with training points in blue and test points in yellow. This plot highlights how the SVM model separates the data based on these features, allowing for a visual assessment of the classification boundaries and the distribution of adulteration status within the feature space.

3.1.4. Sugar Contamination Deception

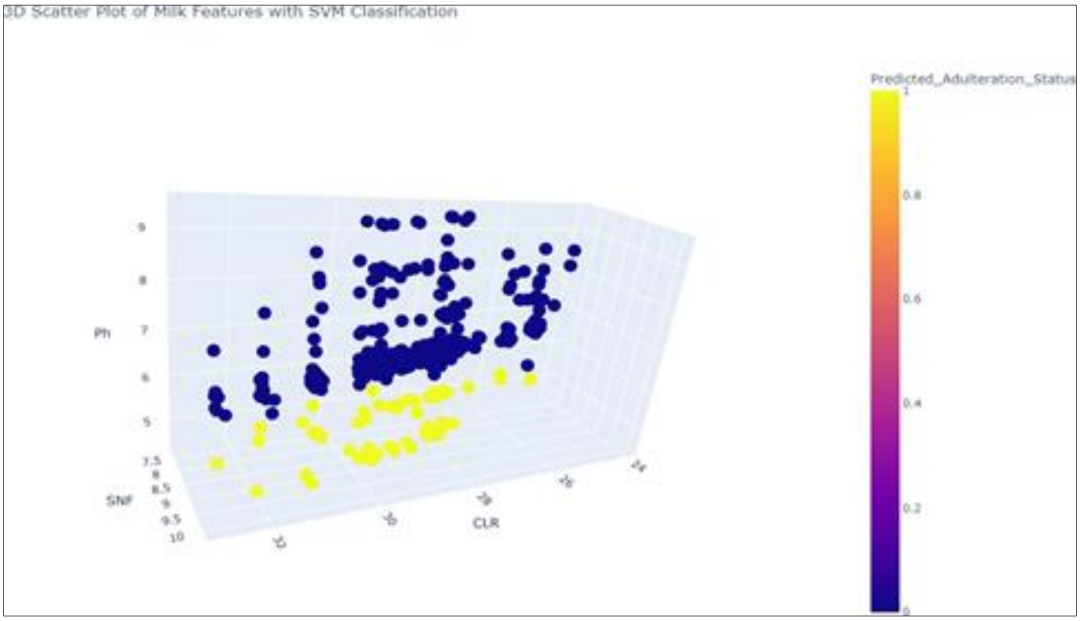

Figure 9 The 3D plot of Sugar Adulteration

This 3D scatter plot visualizes milk features classified by an SVM (Support Vector Machine) model. The axes represent three different milk features: CLR (a measure of milk density) on the x-axis, SNF (Solids-Not-Fat) on the y-axis, and pH level on the z-axis. The data points are coloured based on the predicted adulteration status, with yellow indicating a higher likelihood of adulteration and blue indicating a lower likelihood. The plot shows a clear separation between the two classes, suggesting that the SVM model has successfully identified distinct patterns in the milk features that correspond to adulterated and non-adulterated samples.

4. Discussion

Table 3 Performance Evaluation Results

SL No	Adulteration Type	Detection Status (0 as not adulterated 1 As adulterated	(%) Precision	Recall (%)	F1 score (%)	Accuracy
1	Water	0	0.99	1	1	1
2	Water	1	1	0.86	0.92	1
3	Sugar	0	0.94	1	0.97	0.95
4	Sugar	1	1	0.64	0.78	0.95
5	Urea	0	0.95	0.99	0.97	0.94
6	Urea	1	0.78	0.44	0.56	0.94
7	Detergent	0	0.96	0.99	0.98	0.96
8	Detergent	1	0.79	0.54	0.60	0.96

The model demonstrates exceptional performance in detect- ing water and sugar adulteration, achieving high precision, recall, and F1 scores for both. It accurately identifies non- adulterated samples of both types, with high overall accuracy. However, its effectiveness diminishes when detecting urea adulteration, where precision and recall are notably lower, indicating a higher likelihood of missing adulterated samples. For detergent adulteration, the model's performance is the weakest, with reduced precision and recall, suggesting that it struggles more with distinguishing adulterated from non- adulterated samples. Overall, while the model excels with water and sugar, it requires enhancement, particularly for detecting urea and detergent adulteration, to improve its re- liability and accuracy.

4.1. Water Adulteration

- Non-Adulterated (Score 0): The model shows excellent performance with a precision of 0.99, recall of 1.00, F1 score of 1.00, and accuracy of 1.00. This indicates that the model reliably identifies non-adulterated water samples with minimal false positives or negatives.
- Adulterated (Score 1): For adulterated samples, the model maintains high accuracy at 1.00 but has slightly lower recall (0.86) and F1 score (0.92), which suggests that while it is very accurate overall, it might miss some adulterated samples.

4.2. Sugar Adulteration

- Non-Adulterated (Score 0): The model performs well with a precision of 0.94, recall of 1.00, F1 score of 0.97, and accuracy of 0.95, indicating it effectively identifies non-adulterated sugar samples with high reliability.
- Adulterated (Score 1): The model achieves a precision of 1.00, recall of 0.64, F1 score of 0.78, and accuracy of 0.95. This shows it is effective at identifying adulterated sugar samples, although it might miss a few cases compared to its performance with non-adulterated samples.

4.3. Urea Adulteration

- Non-Adulterated (Score 0): Precision of 0.95, recall of 0.99, F1 score of 0.97, and accuracy of 0.94 show that the model is quite reliable in identifying non-adulterated urea samples, though with slightly lower accuracy compared to water and sugar.

- Adulterated (Score 1): The model's performance drops with precision at 0.78, recall at 0.44, F1 score at 0.56, and accuracy of 0.94. This indicates a significant challenge in detecting adulterated urea, with a higher likelihood of missing adulterated samples.

4.4. Detergent Adulteration

- Non-Adulterated (Score 0): The model has a precision of 0.96, recall of 0.99, F1 score of 0.98, and accuracy of 0.96, showing strong performance in detecting non- adulterated detergent samples.
- Adulterated (Score 1): For adulterated samples, precision is 0.79, recall is 0.54, F1 score is 0.60, and accuracy is 0.96. These lower scores indicate difficulties in iden- tifying adulterated detergent samples accurately, with a higher risk of false negatives.

Linearly separable data. The model maximizes the margin between the support vectors of the two classes, ensuring robust classification even with new data.

4.5. Contributions

The research contributes to the field of food quality control by introducing a novel approach to milk adulteration detection using machine learning. It provides a foundation for further studies and the development of practical applications in the dairy industry [?]. The findings underscore the importance of leveraging computational techniques to enhance food safety and quality assurance. The study also offers practical insights for dairy producers, policymakers, and regulatory bodies, highlighting the potential of machine learning to improve milk quality testing processes. By providing a reliable and efficient method for detecting adulteration, this research paves the way for safer and more transparent dairy products, ultimately benefiting consumers and the industry as a whole.

5. Conclusion

Machine learning, specifically using SVM models, provides a promising avenue for detecting milk adulteration with efficiency, accuracy, and practicality over traditional methods. The evaluated models show significant potential for real-world application, supporting enhanced quality control in the dairy industry. For water adulteration, the SVM model excels in identifying non-adulterated samples with near-perfect metrics, while slightly less effective in detecting adulterated cases. In sugar adulteration, the model effectively identifies both non-adulterated and adulterated samples, though with lower recall for adulterated samples. The model reliably identifies non-adulterated urea samples but struggles with adulterated ones, indicating a need for further refinement due to higher false negatives. For detergent adulteration, the model performs well with non-adulterated samples but shows lower precision, recall, and F1 scores for adulterated samples. Continuous improvement and validation of these SVM models are essential for robust detection across all adulteration types, enhancing the integrity and safety of milk products in the market.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] [Online]. Available: https://www.nddb.coop/information/stats/milkprodstate.

[2] S. Patari, P. Datta, and P. S. Mahapatra, "3d paper-based milk adul- teration detection device," Scientific Reports, vol. 12, no. 1, p. 13657, 2022.

[3] N. Nayak, "Detection of different adulterant in different milk samples,"

[4] Research and Reviews: J. Dairy Sci. Tech, vol. 6, no. 3, pp. 17–29, 2018.

[5] A. K. Yadav, M. Gattupalli, K. Dashora, and V. Kumar, "Key milk adulterants in india and their detection techniques: A review," Food Analytical Methods, vol. 16, no. 3, pp. 499–514, 2023.

[6] Y. Lin, J. Ma, Q. Wang, and D.-W. Sun, "Applications of machine learning techniques for enhancing nondestructive food quality and safety detection," Critical Reviews in Food Science and Nutrition, vol. 63, no. 12, pp. 1649–1669, 2023.

[7] [Online]. Available: https://link.springer.com/article/10.1186/s13040- 019-0200-5

[8] Y. Yakar and K. Karadagˇ, “Identifying olive oil fraud and adulteration using machine learning algorithms,” Qu´ımica Nova, vol. 45, no. 10, pp. 1245–1250, 2022.

[9] M. Frizzarin, I. C. Gormley, D. Berry, T. B. Murphy, A. Casa, A. Lynch, and S. McParland, “Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods,” Journal of Dairy Science, vol. 104, no. 7, pp. 7438–7447, 2021.

[10] J. G. LINN, “Factors affecting the composition of milk from dairy cows,” 1988. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK218193/

[11] [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK218193/

[12] M. Kamthania, J. Saxena, K. Saxena, and D. Sharma, “Milk adultration: Methods of detection &remedial measures,” International Journal of Engineering and Technical Research, vol. 1, pp. 15–20, 2014.

[13] T. Azad and S. Ahmed, “Common milk adulteration and their detection techniques,” International Journal of Food Contamination, vol. 3, pp. 1–9, 2016.

[14] F. Safety and S. A. of India (FSSAI), “Milk and milk products,” https://www.fssai.gov.in/upload/uploadfiles/ files/MILK AND MILK PRODUCTS.pdf, 2024, accessed: 2024-09-17.

[15] M. Kamthania, J. Saxena, K. Saxena, and D. Sharma, “Milk adultration: Methods of detection &remedial measures,” International Journal of Engineering and Technical Research, vol. 1, pp. 15–20, 2014.

[16] T. Azad and S. Ahmed, “Common milk adulteration and their detection techniques,” International Journal of Food Contamination, vol. 3, pp. 1–9, 2016.

[17] S. S. Bhat, S. K. Hegde, V. S. Bhat, K. Ramya, and P. S. Jodalli, “Acido- genic potential of plain milk, milk with sugar, milk with cornflakes and milk cornflakes with sugar: a comparative study,” International Journal of Clinical Pediatric Dentistry, vol. 9, no. 3, p. 218, 2016