Milk Quality Assurance Using transfer learning: A Comprehensive Guide

[1]Nishitha.Pochareddy, [2]Mr. Suresh Tiruvalluru
[1]*M-Tech, Dept. of CSE Gokula Krishna College of Engineering, Sullurpet*
[2]*Associate Professor, M-Tech., (Ph.D), CSE Gokula Krishna College of Engineering, Sullurpet*

Abstract: **Ensuring milk quality is a critical challenge in the dairy industry, directly affecting public health, consumer trust, and economic sustainability. Traditional quality assurance methods are often labor-intensive, time-consuming, and prone to human error. This paper presents a modern approach to milk quality assessment using transfer learning, a powerful deep learning technique that leverages pre-trained models for image and sensor-based classification tasks. The proposed method uses convolutional neural networks (CNNs) adapted from models like ResNet, VGG, or MobileNet, retrained on datasets containing visual and chemical characteristics of milk. The system enables fast, scalable, and accurate classification of milk quality, providing a real-time and cost-effective alternative to conventional testing. Experimental results show high accuracy in detecting contaminated or substandard milk samples, demonstrating the potential of AI in food safety and quality control. Furthermore, it was more accurate and stable than the other methods.**

Keywords: **- ANN, KNN, RF, and SVM.**

1. INTRODUCTION

Milk is often regarded as one of the most beneficial supplements that can be consumed on a daily basis. Consuming milk of a high quality is recommended for the purpose of improving one's health. Milk can be considered a perishable product. A single gram of milk that is of poor quality or structure can cause the quality of tons of milk to deteriorate, which can result in considerable financial losses. Milk that has gone bad can become a breeding ground for millions of germs in a very short amount of time [1]. The old chemical procedures for determining the composition of milk are not only laborious and time-consuming, but they also produce a significant amount of pollution. The evaluation of milk quality was carried out with the assistance of deep learning techniques in this study. The quality of milk that is consumed and sold is frequently decreased due to the presence of a variety of pollutants, including detergents and skim milk powders; however, the most common of these contaminants is dirty water. In 2011, the Food Safety and Standards Authority of India (FSSAI) carried out a National Survey on Milk Adulteration. The results of this survey revealed that water was the most common adulterant found in Indian milk, followed by detergent. The Prevention of Food Adulteration Act (PFA Act) was enacted by the Indian government in 1954 with the intention of addressing this problem and protecting the health of consumers. The PFA Act went into force on June 1, 1955 [2]. According to the legislation, it is illegal to manufacture, sell, or distribute foods that are contaminated or harmful. In spite of legal measures, fortification continues to be needed because there is a lack of individuals with proper training and laboratories that are not adequate. There have been a number of research carried out in order to evaluate the quality of food products. These studies have utilized data obtained from reliable sources and have utilized deep learning techniques.
For the purpose of this paper, we make use of a deep learning algorithm to conduct a predictive analysis of the quality of milk. This analysis was carried out with the assistance of a few general factors that have an impact on the quality of milk. As a result, it is essential to ensure that the quality of milk is preserved in order to ensure that the dairy products that are created in conjunction with it are also of the highest possible quality and that the outbreak of any kind of disease or allergy can be avoided. Due to the fact that the demand for milk consumption is increasing on a daily basis in tandem with the growth of the population, it becomes an extremely laborious work to maintain a check on the quality of each and every sample that is produced in every nook and cranny across the entire world. In such a scenario, the efficient outcome that is provided by methods that save time and require less effort, such as applying techniques for deep learning, proves to be quite beneficial.

2. RELATED WORK

The Multiwavelength Gradient Boosted Regression Tree was proposed by Sheng et al. [3] for the purpose of analyzing the quantities of protein and fat found in milk. Through the utilization of a multichannel spectral sensor, the Multiwavelength Spectral Sensor System came up with a way for determining the intensity of the milk's wavelength. For the purpose of determining whether or not the GBRT regression model is effective, the coefficient of determination, mean square error, mean absolute error, and explained variance regression score were utilized.

The support vector machine (SVM) model has been proposed by Brudzewski et al. [4] in order to obtain data for categorization. SVM neural networks with linear and circular kernels have been utilized in the implementation of the system that is responsible for recognizing and classifying items. The classification of milk was accomplished by the utilization of a system that was founded on oxide-based gas sensors. A detailed study was undertaken by Ghosh et al. [5] with the purpose of evaluating the quality of water through the application of predictive machine learning. The results of their study demonstrated that machine learning models have the capacity to properly evaluate and categorize water quality. This particular dataset contained metrics such as pH, dissolved oxygen, biological oxygen demand (BOD), and total dissolved solids (TDS). A respectable accuracy rate was achieved by the Random Forest model, which emerged as the most accurate among the multitude of models that they utilized. The support vector machine (SVM) model, on the other hand, was falling behind, registering the lowest accuracy. There was a recent investigation conducted by Olcay Polat et al. [6] that investigated the possibility of employing an information fusion framework as a method for categorizing raw milk samples. The purpose of this endeavor was to devise an appropriate method for classifying raw milk into a number of different categories according to numerous critical criteria. Several parameters, including pH, sH, and somatic cell count, were found to be significant in affecting the quality of raw milk, as demonstrated by the scientific findings. Techniques such as Multi-criteria Decision-Making (MCDM) and Analytic Hierarchy Process (AHP) were implemented in the field of dairy processing; nevertheless, the complexity of their computations was quite significant.

Pegah Sadeghi Vasafi et al. [6] applied KNN and SVM classification techniques in order to find anomalies in the milk processing process. This was done in order to address the issue that was being discussed. When the fat content and temperature were taken into consideration as features, the accuracy that was reached was 80.4% for the SVM and 82.8% for the KNN analysis.

3. PROPOSED METHODOLOGY

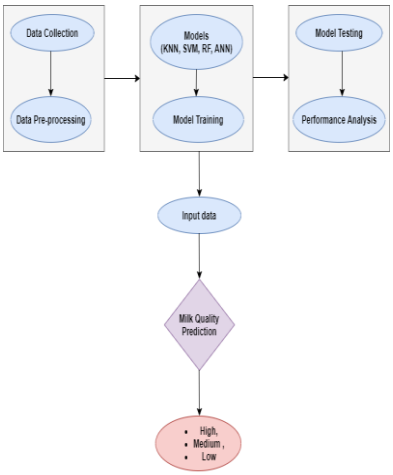

Figure.1 Proposed System Mode

Figure.1 is a block diagram that illustrates the various stages that the proposed system will go through.

3.1 Data Collection

A data set that will be used for the proposed system The data was obtained from the Kaggle repository. This dataset is made up of seven distinct features. For the purpose of predicting the analysis of the milk, these parameters are utilized. One type of data that is categorical is the grade (target) of the milk. It is important to note that there are three distinct classes: low (bad), medium (moderate), and high. The dataset contains a total of 1059 rows, and it also contains 8 columns, which can be thought of as features.

3.2 Data Pre-Processing

The first step involves calculating the missing value of the data in pre-processing, and it turns out that none of the features have a missing value. The next step involves label encoding, which converts the values in this case to category integer values because the value of the attribute in the problem cannot be understood by a computer at all. Label encoding is used for the "Grade" feature in this dataset, and the feature values are then called.

3.3 Models

3.3.1 RF Classifier

Random Forest is a type of ensemble learning that generates several different decision trees throughout

the training process. It then generates a class (for classification tasks) or a mean prediction (for regression tasks) for each tree based on data that has not yet been seen. A collection of decision trees, which are often trained using a "bagging" approach, constitutes the "forest" that it constructs with its efforts. When it comes to predicting milk quality, random forests can manage data sets that contain a large number of features and identify the significance of each feature. There is a possibility that specific metrics are absent from some of the milk samples in a real-world setting. The ability of random forests to accommodate missing values while still producing accurate forecasts is quite useful. The fact that decision trees tend to overfit is one of the issues that they present. Random forests, on the other hand, can attain superior generalization by employing several trees and accumulating the findings of those trees.

3.3.2 SVM Classifier

The Support Vector Machine (SVM) is a supervised machine learning method that is primarily utilized for classification and regression duties. Discovering the hyperplane that most accurately categorizes the data, or in the case of regression, the hyperplane that most effectively fits the data is the fundamental concept. SVM does this by maximizing the distance between the hyperplane and the data points in the two classes that are closest to it. These data points are referred to as support vectors. By its capacity to successfully manage high-dimensional data, support vector machines (SVM) are ideally suited for forecasting milk quality based on several different parameters. Through the utilization of kernel functions such as radial basis functions (RBF), polynomials, and sigmoid, support vector machines (SVM) can manage nonlinear interactions between parameters. Particularly when margins are selected with care, support vector machines (SVMs) are less likely to overfit the data. This indicates that data that is not yet visible can be summarized more effectively, which is essential for making correct predictions regarding additional milk samples once they have been collected.

3.3.3 KNN Classifier

Another effective machine learning technique that is utilized for data classification is known as the k-nearest neighbor's (KNN) algorithm. By determining whether or not the data points that are closest to a random data point belong to a certain group, it is possible to estimate the chance that the data point in question will belong to one of the two groups when using this method. It is possible to use it to solve problems involving classification as well as regression, although it is more helpful when applied to a problem that was solved previously. It would be beneficial for us to select an odd value for the K value when we are making our selection. There are a number of different approaches to calculating distance, and these approaches can be utilized to determine which places are the closest or nearest to one another.

3.3.4 ANN Classifier

When it comes to information processing, an Artificial Neural Network (ANN) is a model that models the way the brain works. An artificial neural network (ANN) is specifically configured for a variety of purposes, including but not limited to the recognition of patterns or classification of data in general. Through research, demonstrated to be successful in a wide variety of fields, including but not limited to the application of artificial neural networks, from medicine to engineering-based methods in the study of complicated situations. A neural network There are hundreds of nodes, also known as artificial neurons, that make up the network formed by the neural network, which is connected by coefficients (weights). The framework. To do computations, neural networks utilize neurons. Both a single output and a single weighted input are present for every single PE. Architecture, learning principles, and transfer functions are all important. The behavior of a neural network can be determined. Within this context, given that the weights represent the parameters, a neural network is a parameter system. to be adjusted. The sum of the inputs, all of which are weighted, is what whether or not the neuron is triggered is determined by this. An activation of a signal is fed into a transfer function, which results in the production of one output from the neuron in question.

4. PERFORMANCE ANALYSIS

In the study that is being proposed, KNN, SVM, RF, and ANN are utilized for the purpose of predicting the quality of milk. The accuracy of the classifiers' predictions is evaluated with the help of a variety of performance matrices, such as the Performance Score, which is presented in Table 1.

Table 1. Performance Score

Models	Accuracy	Precision	Recall	F1 Score
KNN	99.37	99.30	99.42	99.35
SVM	93.08	92.51	92.85	92.67
RF	99.37	99.14	99.42	99.27
ANN	99.68	99.73	99.71	99.72

A classifier's accuracy can be determined by calculating the percentage of correct predictions relative to the total number of guesses. The level of precision of a Classifier is the degree to which it accurately predicts True Positive from Total Positive Predicted. The classifier's ability to accurately predict true positives from actual projected values is referred to as its recall. Through the utilization of Precision and Recall, the F1 score can determine how effective the model is.

5. CONCLUSION

Within the scope of this work, we employed a machine learning algorithm to conduct predictive analysis of the quality of milk. This analysis was carried out with the assistance of a few general factors that have an impact on the quality of milk. The maintenance of the quality of milk is of utmost importance to ensure that the dairy products that are created in conjunction with it are also of the highest possible quality and that the outbreak of any form of disease or allergy can be avoided. Since the demand for milk consumption is increasing daily in tandem with the growth of the population, it becomes extremely laborious work to maintain a check on the quality of every sample that is produced in every nook and cranny across the entire world. In such a scenario, the efficient outcome that is provided by methods that save time and require less labor, such as applying techniques for deep learning, shows to be quite helpful. During this investigation, we utilized a KNN, SVM, RF, and ANN approach to carry out metric assessments. Upon further investigation, we discovered that the ANN model was essential in achieving a higher level of accuracy in the classification of milk quality.This paper presented a practical and efficient approach to milk quality assurance using transfer learning. By leveraging powerful pre-trained CNN models, the system is capable of accurately detecting anomalies in milk samples with minimal computational resources and limited data. The results demonstrate that transfer learning can significantly reduce the time and cost associated with traditional quality control methods, while also increasing reliability and scalability. Future work may include integrating IoT sensors for continuous monitoring, expanding datasets for better generalization, and developing mobile applications for field usage.

REFERENCES

[1] Anderson, Melisa, et al. "The microbial content of unexpired pasteurized milk from selected supermarkets in a developing country." Asian Pacific Journal of tropical biomedicine 1.3 (2011): Volume 1, Issue 3,2011, Pages,205-211,ISSN doi:10.1016/S2221-1691(11)600282.

[2] J. K. Swathi &N. Kauser (2015). A study on adulteration of milk and milk products from local vendors. International Journal of Biomedical and Advance Research, 6(09), 678-681.

[3] T. Sheng, S. Shi, Y. Zhu, D. Chen and S. Liu, "Analysis of Protein and Fat in Milk Using Multiwavelength Gradient-Boosted Regression Tree," in IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-10, 2022, Art no. 2507810, doi: 10.1109/TIM.2022.3165298

[4] K. Brudzewski a, S. Osowski b, T. Markiewicz b , “Classification of milk by means of an electronic nose and SVM neural network”. Received 30 June 2003, Revised 13 October 2003, Accepted 21 October 2003, Available online 30 December 2003.

[5] Ghosh, H., Tusher, M.A., Rahat, I.S., Khasim, S., Mohanty, S.N. (2023). Water Quality Assessment Through Predictive Machine Learning. In: Intelligent Computing and Networking. IC-ICN 2023. Lecture Notes in Networks and Systems, vol 699. Springer, Singapore. https://doi.org/10.1007/978-981-99-31774_6.

[6] O. Polat, S. G. Akçok, M. A. Akbay, D. Topaloğlu, S. Arslan, &C. B. Kalayci, (2021). Classification of raw cow milk using information fusion framework. Journal of Food Measurement and Characterization, 15(6), 5113-5130.

[7] P. S. Vasafi, &B. Hitzmann, (2022). Comparison of various classification techniques for supervision of milk processing. Engineering in Life Sciences, 22(3-4), 279-287.