Precision Livestock

Journal Homepage: sciencesforce.com/pl

Precision Liv. Vol. 1 (2024) 1–13

Paper Type: Original Article

Hybrid Deep Learning Approach for Milk Quality Prediction

Ahmed Tolba [1], Nihal N. Mostafa [2], Ali Wagdy Mohamed [3], and Karam Sallam [4,*]

[1] Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt; a.tolba24@fci.zu.edu.eg.

[2] Department of Computer Science, Misr higher institute for computer and commerce, Egypt; nihal.nabil@fci.zu.edu.eg.

[3] Operations Research Department, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza 12613, Egypt; aliwagdy@staff.cu.edu.eg.

[4] School of IT and Systems, University of Canberra, ACT 2601, Australia; karam.sallam@canberra.edu.au.

Received: 23 Sep 2023 **Revised:** 16 Dec 2023 **Accepted:** 04 Jan 2024 **Published:** 09 Jan 2024

Abstract

Milk quality prediction is considered a vital research area due to increase the need for obtain sustainable development goals. This study aims to predict milk quality by integrate gated recurrent units (GRUs) and residual network (ResNet). Our model was evaluated on milk quality prediction dataset with seven unique feature such as pH, temperature, taste, odor, fat, turbidity, and color. The prediction output is classified with high (Goog), Low (Bad), and Medium (Moderate) classes. Our model shows superior results with comparison with multi-layer perceptron (MLP), random forest (RF) and support vector machine (SVM). In terms of accuracy, precision, recall, and F1-score 0.996, 0.992, 0.992, 0.992.

Keywords: Deep Learning; Milk Quality; Gated Recurrent Units, Residual Network; ResNet.

1 | Introduction

Milk production is a significant contributor to household livelihoods, food security, and nutrition in developing countries, with around 150 million households engaged in this sector. However, recent decades have seen a rise in dairy production in developing countries due to factors such as poor-quality feed resources, diseases, limited access to markets, and low genetic potential of dairy animals. Many developing countries also have unfavorable climates for dairying. Some countries have a long history of milk production, while others have recently established significant dairy production. These countries are mostly located in the Mediterranean and Near East, Indian subcontinent, West African savannah regions, East African highlands, and parts of South and Central America [1].

However, because dairy products and milk are produced in enormous amounts, manually testing their quality is a challenging operation. pH, storage circumstances, and weather are also numerous factors, including the health of the cows, might have an indirect or direct impact on the quality of milk. Consequently, machine learning (ML) is utilized to learn detailed patterns related to the obtained data and replace manual testing in the process of forecasting milk quality.

Corresponding Author: karam.sallam@canberra.edu.au

https://doi.org/10.61356/j.pl.2024.1199

Existing and newly created ML models, which have been used for many years, are effective measures of milk quality. Frizzarin et al. [2] investigated the efficiency of several ML technique on milk samples from 622 cows. The study explore that statistical ML could improve prediction performance over the traditional partial least squares approach. Support vector machine (SVM) had the highest accuracy for most traits. The use of modern statistical machine learning methods for trait prediction from mid-infrared spectroscopy may improve prediction accuracy for some traits. Another study presented principal component analysis (PCA) and wavelet transform, two compressed learning approaches, to turn Mid-infrared spectroscopy (MIRS) data into a compressed domain. The study obtains near-lossless compression values for both strategies for transforming MIRS data while maintaining prediction accuracy for a set of milk quality traits [3].

Unfortunately, ML techniques have some limitation in milk quality prediction. With the use of a deep learning (DL) method, all these restrictions can be removed. ML limitations are concluded as follows [4].

- Additional dimensionality reduction techniques like PCA are necessary for other non-linear approaches like SVM, Artificial Neural Networks (ANN), and Random Forest (RF) that can be used to extract information from non-linear spectral data.
- To establish a connection between spectral data and chemical concentrations, these techniques necessitate prior knowledge of the data pertaining to neighboring wavelengths, absorption peaks, and noise since they reduce high-dimensional space to low-dimensional space.

DL is a subfield of ML that may be used in ML frameworks and allows for the learning of compositions at different levels. It does feature extraction and processing, learns autonomously from the data, and requires very little prior knowledge of the dataset [5]. One of the most popular DL algorithms is the convolutional neural network (CNN), which is utilized in computer vision, remote sensing, biomedical, and biometric applications [6]. Other widely used DL algorithms are Recurrent Neural Networks (RNN), Feedforward Neural Networks (FNN), Deep Belief Networks (DBN), and Stacked Autoencoders (SAE) [7].

RNNs shows high performance in time-series, voice recognition, and signal-processing applications. Utilizing RNN models in can be difficult task due to their insufficient ability to handle sequential data. The expanding or decreasing nature of gradients during model training causes an issue with the backpropagated gradients vanishing or bursting, even though they are crucial for data processing. Some variants of RNN model can eliminate these problems. Long Short-Term Memory (LSTM) is a variant of RNN that can store prior knowledge within nodes for longer periods. Gated Recurrent Units (GRUs) are another form of advanced RNN that reduce redundancies within data and use reset and update gates to save prior information. Recent studies have suggested that combining RNNs with CNN can help improve the overall performance of the DL models for various applications [8].

Thus, in this paper we aim to maintain a hybrid GRU and Residual Networks (ResNet) which is variant of CNN milk quality dataset to predict milk quality in terms of namely High (Goog), Low (Bad), and Medium (Moderate). The prediction is depending on seven features such as pH, Temperature, Taste, Odor, Fat, Turbidity, and Color. A comparison between multi-layer perceptron (MLP), RF, and SVM has been made to evaluate our proposed model.

Thus, the main contributions of this study are as follows:

- An integration between GRU and ResNet is proposed.
- The integration of a residual block into the architecture based on gated GRU enables the effective capture of both long-term and short-term dependencies in sequential data, while simultaneously resolving the problem of the vanishing gradient.
- The proposed model results show high performance compared to MLP, RF, and SVM.
- Our model can preserve weights from over-stack within early layers.

The remainder of the paper is divided as follows. Section 2 provides the background needed for this study. Section 3 presents the methodology of this study. Section 4 presents proposed model. Section 5 presents experimental results. Section 6 illustrates the conclusion and future directions of this proposal.

2 |Literature Review

Many studies use DL models to enhance the performance of milk quality prediction. Said et al. [9] introduced a semi-supervised DL algorithm for predicting milk fat content and water adulteration ratio in cow milk using 14 Neospectra spectrometers. The algorithm outperforms a supervised model with an R2 = 0.95 and RMSE = 0.22, despite a reduced referenced dataset.

Another study utilized CNNs to assign a quality rating to milk based on different sensor readings. When compared to conventional techniques, their DL model produced excellent accuracy and lower maintenance costs. But they also emphasized issues with vanishing gradients, overfitting, and processing demands when employing CNNs [10]. Bhavsar et al. [11] proposed ML technique to predict milk quality. The classification results 99.99% according to accuracy.

Mu et al. [12] developed an electronic nose (E-nose) using seven metal oxide semiconductor sensors to identify milk sources and estimate milk fat and protein content. The device is low-cost and non-destructive. The E-nose uses principal component analysis, linear discriminant analysis, and ML algorithms to construct a classification model. The SVM model with fusion features has the best accuracy of 95%. The RF models provide the best performance in estimating milk fat and protein content, providing a technical basis for predicting milk quality.

Vidhya et al. [13] introduced a functional prototype of an AC cross-conductance sensor for detecting milk adulteration. The sensor, consisting of four conductive brass electrodes, was designed and tested in a lab environment to detect drinking water and milk with varying impurities. The sensor demonstrated significant resolution and precision in distinguishing pure and tainted milk.

Another contribution integrated AfiLab real-time milk analyzer measures with the stacking ensemble learning technique for in-line daily monitoring of cheese-making traits in Holstein cattle. Data and samples were collected from 499 Holstein cows from two farms. The study assessed the predictive ability of different statistical methods, including elastic net, gradient boosting machine, extreme gradient boosting, and ANN, across different cross-validation scenarios. The results showed that including cow information in the AfiLab infrared prediction increased accuracy by 10.3% for traditional MCP, 13.8% for curd firming, 9.8% for CY, and 11.2% for REC traits. Combining in-line on-farm information with stacking ensemble ML is an effective alternative for obtaining robust daily predictions of milk cheese-making traits.

Mota et al. [14] explores the predictive ability of random forest (RF) and gradient boosting machine (GBM) and penalized regression against PLS regression for predicting three phenotypes in Holstein-Friesian cattle. The data set included 471 Holstein-Friesian cows and three target phenotypes: body condition score (BCS), blood β-hydroxybutyrate (BHB), and ϰ-casein expressed as a percentage of nitrogen (ϰ-CN). The elastic net (EN) was found to be the best regularization, increasing predictive ability by 5%. The GBM model outperformed other methods in predictive ability by around 4%, 1%, and 7% for EN, RF, and PLS, respectively. The study concluded that GBM is a promising method for obtaining more accurate FTIR-based predictions for different phenotypes in dairy cattle.

Slob et al. [15] analyzed the application of DL in dairy farm management. It identifies 427 papers, with 38 primary studies. Over half (55%) focus on disease detection, while milk production and quality are the main problems. Seventy-one independent variables are identified, with milking parameters and milk properties being the most common. Twenty-three algorithms have been identified, with decision tree-based and artificial neural network-based being the most used. Challenges encountered include feature selection, unbalanced data, problem size, overfitting/estimating, and parameter tuning. This is the first systematic literature review on ML in dairy farm management.

3 | Methodology

In this section, we will introduce the method proposed in this paper in detail, first an overview of GRU, and then ResNet.

3.1 | Integrate Gated Recurrent Units (GRUs)

Similar to LSTM, a GRU [16] is a type of recurrent neural network (RNN) structure. With fewer parameters and a simpler structure than LSTM, GRU is an advanced form of the LSTM model that runs faster on specific tasks due to its improved training efficiency. When training long-term sequences, GRUs are meant to help with the vanishing gradient issue. It has been extensively utilized in natural language processing (NLP) applications, including text synthesis, machine translation, and language modeling, among other sequence data modeling tasks. Figure 1 displays the GRU structural diagram.

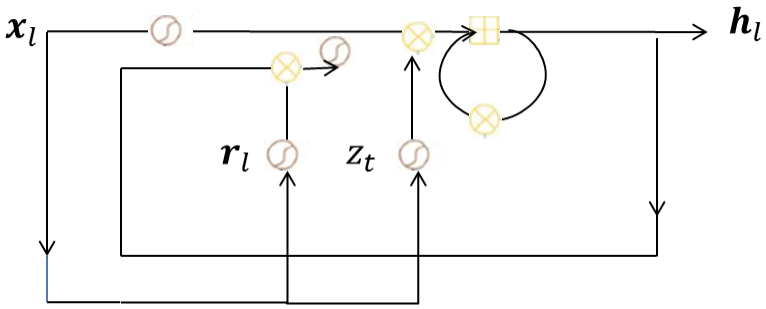


Figure 1. GRU architecture.

The GRU structure only includes two gate structures—a reset gate and an update gate—as opposed to the three gate structures of the input gate, output gate, and forget the entry of the LSTM structure. The update gate regulates the power of the last moment's hidden state on the current moment's output, while the reset gate controls the effect of the last moment's hidden state on the input of the current moment. In addition, GRU computes the current hidden state using a candidate hidden form.

The GRU model's formula is as follows:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (1)$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \quad (2)$$

$$\widetilde{h_t} = f(W_h x_t + U_h(r_t \odot h_{t-1})) \quad (3)$$

$$h_t = z_t \odot \widetilde{h_t} + (1 - z_t) \odot h_{t-1} \quad (4)$$

The formula is as follows: x_t is the input information; σ is the sigmoid activation function; r_t is the reset gate; z_t is the update gate; h_{t-1} is the hidden layer's state at the prior instant; $\widetilde{h_t}$ is the input x_t and the previous hidden layer; f is the tanh activation function; σ is the sigmoid activation function.

3.2 | Residual Network

ResNet [17] is a DL model which is designed to support hundreds or thousands of convolutional layers. It addresses the "vanishing gradient" problem by using "skip connections" to skip multiple layers and reuse activations from previous layers. This speeds up initial training and allows residual parts to explore more of the input image's feature space. Most ResNet models skip two or three layers at a time, with advanced models like HighwayNets learning "skip weights" dynamically.

ResNet architecture incorporates residual blocks, a key component. Older architectures like VGG16 used convolutional layers with batch normalization and nonlinear activation layers. However, increasing the number of layers can significantly improve CNN performance. ResNet architecture introduces an intermediate input to the output of convolution blocks. Residual block is shown in Figure 2.

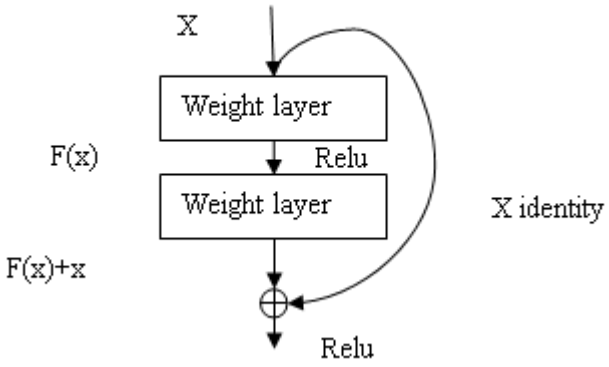


Figure 2. Residual network architecture.

4 | Proposed Method

This section will provide an overview of our proposed method, starting with the datasets that were selected to evaluate our proposed, then moving on to our proposed method.

4.1 | Dataset

The Kaggle Milk Quality dataset [18] was utilized to evaluate our proposed model. There are seven unique features present in the dataset, namely pH, Temperature, Taste, Odor, Fat, Turbidity, and Color. These criteria are employed to forecast the quality of milk. The category data denotes the specific grade or goal grade of the milk. There exist three unique categories, namely High (Goog), Low (Bad), and Medium (Moderate), as indicated by their respective numbers of 256, 429, and 374. The dataset comprises 1059 rows and 8 columns in total. There are seven classified qualities and one quantitative trait among all the traits. Table 1 shows the distribution of milk quality dataset.

Table 1. Milk quality prediction dataset distribution.

	Count	Mean	std	min	25%	50%	75%	max
pH	1059	6.63	1.399	3	6.5	6.7	6.8	9.5
Temperature	1059	44.23	10.1	34	38	41	45	90
Taste	1059	0.546	0.498	0	0	1	1	1
Odor	1059	0.432	0.495	0	0	0	1	1
Fat	1059	0.671	0.469	0	0	1	1	1
Turbidity	1059	0.491	0.5	0	0	0	1	1
Color	1059	251.8	4.307	240	250	255	255	255

4.2 | Experimental Details

In this study, an integration between GRU and Resnet has been introduced.

4.2.1 | Pre-processing Stage

Initially, verify the presence of null values. It has been determined that there are no missing values in any of the features.

In the Second stage, an examination of the data distribution reveals an imbalance in the dataset. Specifically, the labels (high, low, medium) include values of 256, 429, and 374, respectively. This issue hinders the model's ability to learn effectively. An effective approach to address this issue is to employ oversampling of the instances within the minority class. The SMOTE [5] technique is employed to choose examples that are in close proximity inside the feature space. This is achieved by drawing a line connecting the examples in the feature space and subsequently selecting a new sample at a certain location along this line. SMOTE initially chooses a random instance from the minority class and identifies its k closest neighbors from the minority class. The synthetic instance is generated by randomly selecting one of the k nearest neighbors b and building a line segment in the feature space by connecting a and b. The generation of synthetic instances involves the creation of a convex combination of two selected examples, denoted as a and b. Upon implementing this technique, the data achieves equilibrium among the three classes, resulting in a total of 429 data points for each class.

In the third stage, the object values in the 'Grade' feature should be replaced by category integer values ['low': 0 , 'medium': 1 , 'high': 2] as they are not comprehensible to a computer.

In the Fourth stage, feature values are scaled using the StandardScaler. In the field of machine learning, the StandardScaler is a preprocessing technique employed to normalize features. This is achieved by eliminating the mean and scaling the features to unit variance. The data is transformed in a manner that assigns a mean of zero and a standard deviation of one to each feature. The transformation performed expressed mathematically as:

$$Z = \frac{X - \mu}{S} \tag{5}$$

Where Z represent new samples, μ Mean and S Standard deviation.

Normalization facilitates faster convergence of algorithms and mitigates the dominance of features with larger scales over those with lower scales. The StandardScaler function is utilized to approximate the data to a standard normal distribution.

Finally Splitting the dataset, for DL experiment the dataset is divided into multiple subsets to facilitate training, testing and validation by 80%, 10 %, 10%. While in machine learning experiment, the dataset is divided into two parts training and testing by 80% and 20 % respectively.

4.2.2 | Model Architecture

The proposed model is based on the concept of the residual connection of GRU layers. A residual connection, often referred to as a skip connection, is a method employed in neural network architectures to add the initial input of a layer into its output. This approach aids in addressing the issue of the vanishing gradient problem and enhances the training process for deeper neural networks. GRU is a specific architecture from RNN. It is based on the gates mechanism where the system incorporates various gates to regulate the transmission of data inside the network, such as an update gate and a reset gate. The GRU model integrates the benefits of conventional RNNs with a more streamlined training procedure through the selective updating of its memory state.

The integration of a residual block into the architecture based on GRU enables the effective capture of both long-term and short-term dependencies in sequential data, while simultaneously resolving the problem of the vanishing gradient. This approach yields more precise outcomes and enhanced efficiency. The proposed architecture was demonstrated in Figure 3.

4.2.3 |Training Model

Our proposed model compiled to determine the loss function, Adam optimizer with learning rate 0.001, and metrics for evaluating performance. The Categorical cross entropy loss function is used to optimize the initial weights of certain DL models in order to increase classification accuracy. The loss function is mathematically defined as follows:

$$\text{Minimize: } \text{loss} = -\sum_{i=1}^{M} y_i \cdot \log \breve{y}_i \tag{6}$$

Where y_i represent real values and $\breve{y}_i$ represent predicted values.

Most DL models were trained with 70 epochs. In our model a mini-batch gradient descent technique was utilized to decrease the error calculated from the loss function (Categorical Cross Entropy), Where the batch size is 16.

The dataset is divided into multiple subsets to facilitate training, testing and validation by 80%, 10 %, 10% respectively. The training dataset is utilized to train DL models and subsequently assess their performance. In order to assess the model's performance and make necessary adjustments to its parameters, the validation dataset is employed. Finally, the use of the testing dataset is employed to evaluate the ultimate performance and transferability of the trained model on unseen data. Our proposed was compared with ML models such as RF Classifier and SVM are built with parameters (criterion='gini',max_depth=3) and kernel='linear' respectively . Figure 4 shows the general steps of our proposed model.

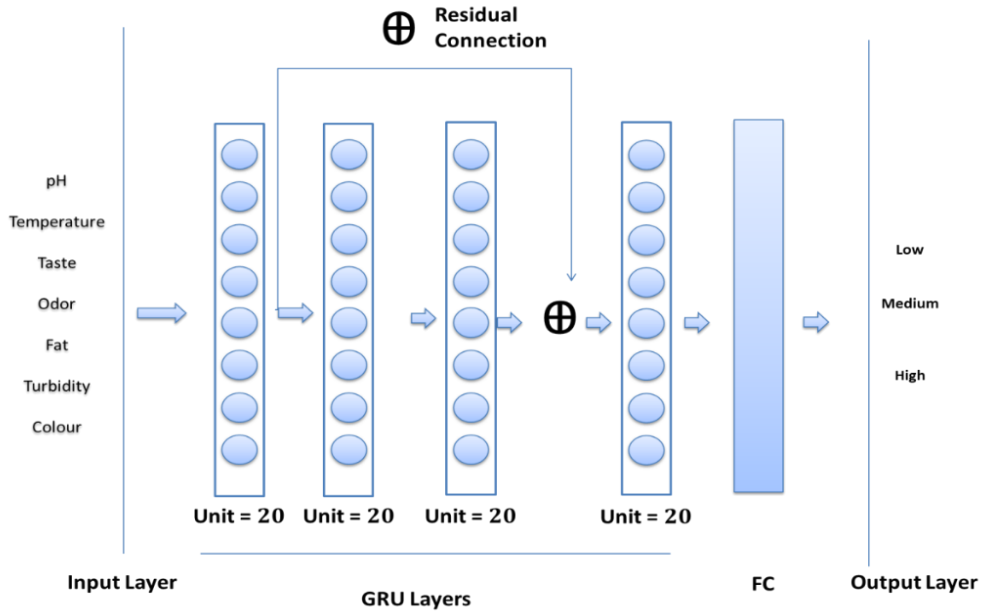


Figure 3. Our proposed model.

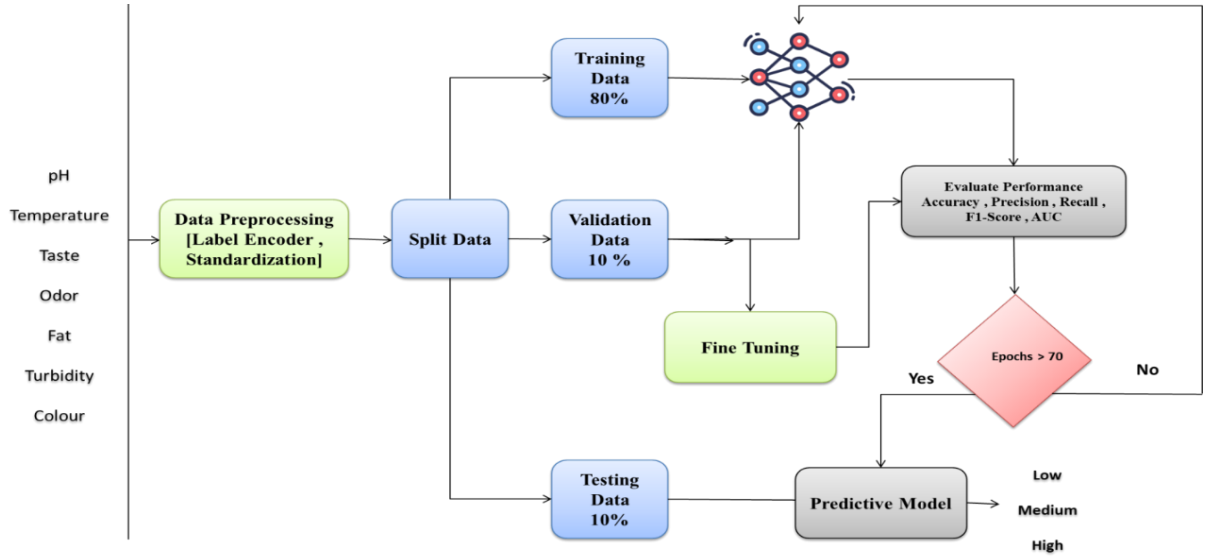


Figure 4. Deep learning pipeline for milk quaility prediction.

5 | Results and Discussion

This section investigates the performance of the proposed model using a widely used dataset, Milk Quality. In addition, it has been compared to several Machine learning and Deep learning models, such as RF Classifier, SVM, MLP and The Proposed Model. Those DL models are implemented in Python using the Kaggle platform and Keras API. The Adam optimizer was used to train the weights of those models for 70 epochs. And ML Models are implemented in Python using the Kaggle platform and scikit learn API. The performance indicators used to evaluate the performance of those models is Accuracy of a classifier is determined by the ratio of correct predictions to the total number of predictions. Precision measures the accuracy of a Classifier in predicting True Positives out of the total number of Positive Predicted instances. Recall quantifies the accuracy of a classifier in correctly identifying true positives based on the actual predicted values. The efficiency of the model is assessed by the F1 score, which is calculated based on the Precision and Recall metrics. AUC, It shows how well the model can discriminate between positive and negative examples. The equations 7, 8, 9, 10 and 11 represent the formulas for the performance parameters.

Numerical representation:

$$\textbf{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}. \tag{7}$$

$$\textbf{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{8}$$

$$\textbf{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{9}$$

$$\textbf{F1} - \textbf{score} = 2 * \frac{Precision \cdot Recall}{Precision + Recal}. \tag{10}$$

$$\textbf{AUC} = \frac{1 + \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{TN}} - \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}}{2} \tag{11}$$

Where TP, FN, TN, and FP represent true positive, false negative, true negative, and false positive, respectively.

Graphical Representation:

- **Confusion Matrix:** A confusion matrix is an illustration that provides a concise summary of the effectiveness of a ML or DL model when applied to a given dataset. The data presented illustrates the frequencies of accurate and inaccurate predictions generated by the model as shown in Figure 5. An ideal model achieves optimal performance by accurately predicting all occurrences of the true values in the dataset.

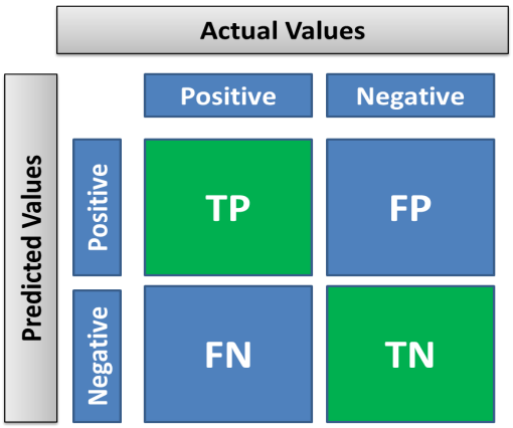


Figure 5. Confusion matrix.

- **ROC Curve:** The receiver operating characteristic (ROC) curve quantifies the performance of a model in classification tasks. It graphs the specificity (1 - false positive rate) against the sensitivity (true positive rate) at different classification thresholds. A higher receiver operating characteristic (ROC) curve indicates superior performance and is employed to evaluate the model's capacity to differentiate between positive and negative instances.
- **TSNE:** The t-SNE (t-Distributed Stochastic Neighbor Embedding) technique is commonly employed for the purpose of representing high-dimensional data in a lower-dimensional environment. It is beneficial for conducting exploratory data analysis. The objective is to create a map of data points in order to put related points in close proximity to each other. By categorizing similar classes following the prediction process, we may assess the model's efficacy by examining the heightened dissimilarity between related classes.

As shown by the data shown in Table 2, The "Proposed Model" demonstrates exceptional performance Milk Quality Prediction, as seen by its highest accuracy, precision, recall, F1-Score, and AUC accuracy score of 0.996, 0.992, 0.992, 0.992 and 0.994 respectively, surpassing all other models. Although the "Proposed Model" possesses the greatest number of parameters with 9,363, it also demonstrates a commendable level of time consumption 24.00 seconds for both training and inference in comparison to alternative models, indicating effective utilization of resources. The SVM and RF model have the lowest levels of accuracy, precision, recall, F1-Score, and AUC when compared to other models.

In Figures 6 and 7 Illustrate the performance evaluation of the proposed model is examined under the accuracy curve, loss curve, ROC curve, Confusion Matrix representation and TSNE histograms.

According to the accuracy curve (a) and loss curve (b), The training process commenced with an initial accuracy rate of 38.37% and a matching loss rate of 1.0966. Subsequently, the validation process yielded an accuracy rate of 55.81% and a loss rate of 1.0461. During the training process, there was a substantial improvement in both the accuracy of the training and validation data. By the end of the training, the accuracy reached its highest point at 100%. The loss exhibited a consistent decline during the training process, suggesting that the model's effectiveness in reducing prediction mistakes has significantly improved. The validation accuracy constantly maintained a high level, suggesting that the model effectively applied its knowledge to new and unseen data.

According to the receiver operating characteristic (ROC) curve (c) for a 3-class classification model It is seen that the macro average (AUC) attains a value of 99%, with all classes.

According to the Confusion Matrix Representation (d) and TSNE plot (e) : proved that the model achieved high accuracy, especially for classes Low and Medium, but had a minor confusion between classes Medium and high.

Table 2. Comparison between the proposed model and others in terms of various performance indicators.

Model Name	Consuming time	# Parameters	Accuracy	Precision	Recall	F1-Score	AUC
Proposed Model	24.00	9,363	0.996	0.992	0.992	0.992	0.994
MLP	14.38	50	0.933	0.922	0.922	0.921	0.941
RF	0.19	1062	0.914	0.920	0.920	0.910	0.996
SVM	0.02	2450	0.868	0.870	0.870	0.870	0.933

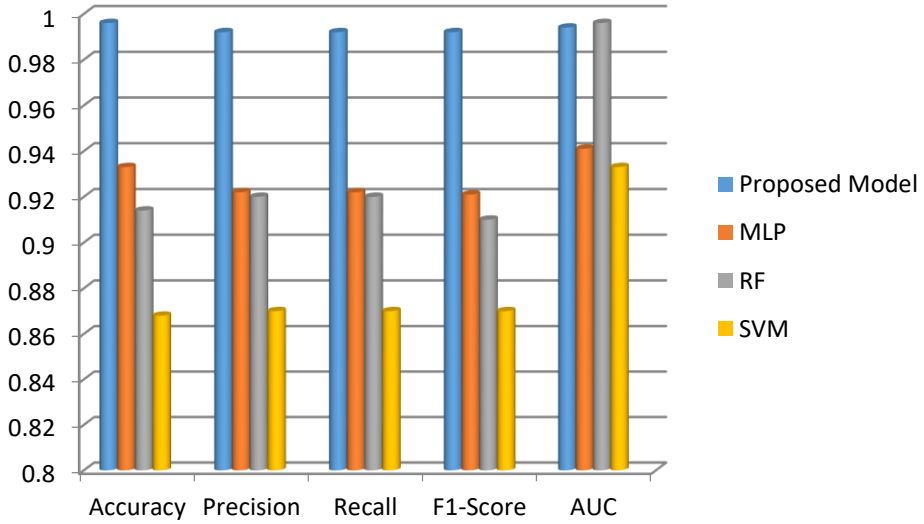
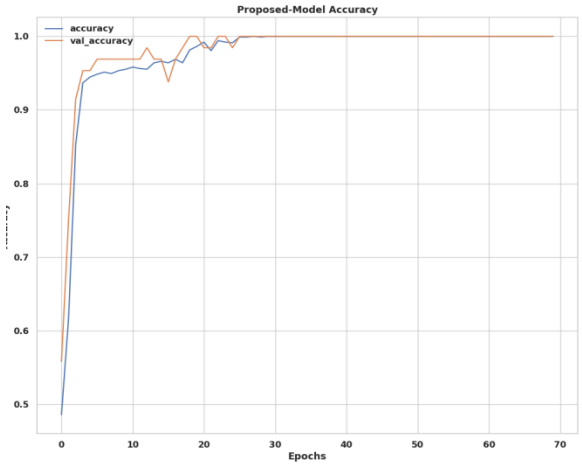
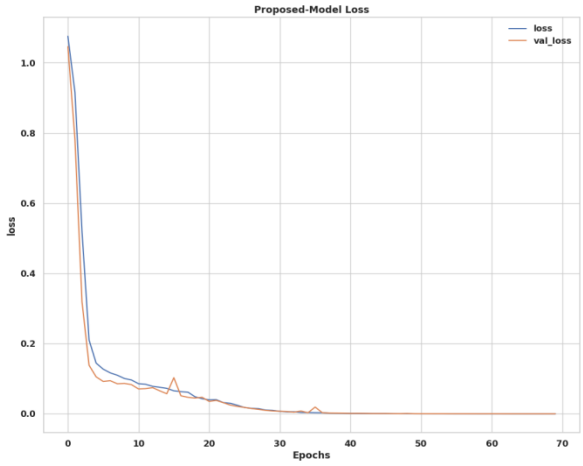


Figure 6. Proposed model results.

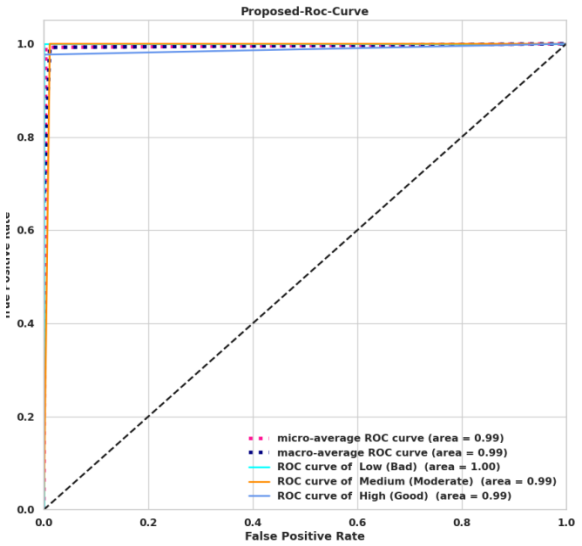
.



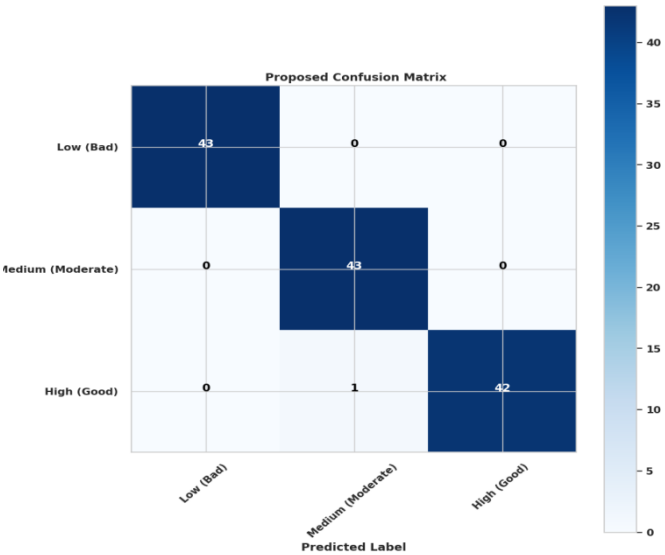
a) Accuracy curve



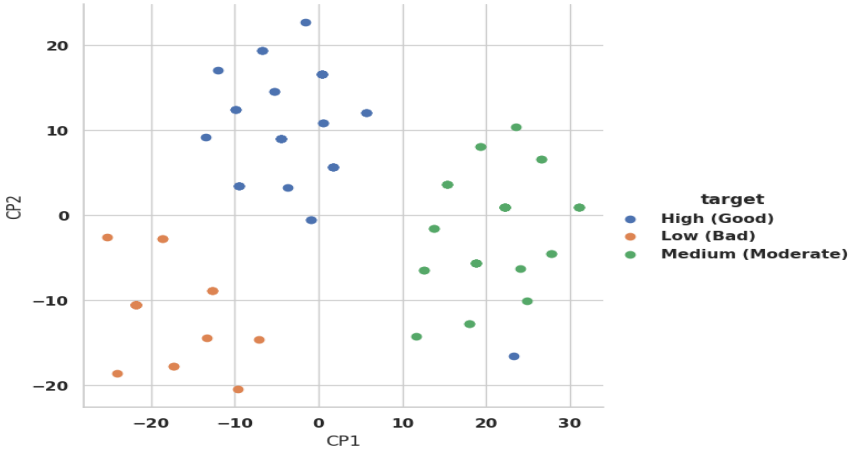
b) Loss curve



c) ROC curve



d) Confusion Matrix



e) TSNE

Figure 7. Performance evaluation of the proposed model under accuracy curve, loss curve, ROC curve, CM and TSNE events.

6 | Conclusion and Future Work

In this study, we proposed a novel DL technique by integrating gated recurrent unit (GRU) with residual network (ResNet). GRU enables the effective capture of both long-term and short-term dependencies in sequential data. Also, residual block eliminates the effect of the vanishing gradient problem. Our proposed model was compared with Muli-layer perceptron, random forest, and support vector machine. The proposed model shows superior results in terms of Accuracy, Precision, Recall, and F1-Score by 0.996, 0.992, 0.992, 0.992, respectively.

Further studies are needed to show performance with multiclass classification during class expansion. Also, more comparisons are needed with different DL architectures. More DL models are required to deal with real data that contains uncertainty. To reduce the memory footprint, model compression techniques will be used in future work to lower model complexity and potentially reduce energy use.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Author Contributaion

All authors contributed equally to this work.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Faye, B. and G. Konuspayeva, The sustainability challenge to the dairy sector–The growing importance of non-cattle milk production worldwide. International dairy journal, 2012. 24(2): p. 50-56.

[2] Frizzarin, M., et al., Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. Journal of Dairy Science, 2021. 104(7): p. 7438-7447.

[3] Vimalajeewa, D., C. Kulatunga, and D.P. Berry, Learning in the compressed data domain: Application to milk quality prediction. Information Sciences, 2018. 459: p. 149-167.

[4] Mhapsekar, R.U., et al., Hybrid Blended Deep Learning Approach for Milk Quality Analysis. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024.

[5] Goodfellow, I., Y. Bengio, and A. Courville, Deep learning. 2016: MIT press.

[6] Malek, S., F. Melgani, and Y. Bazi, One-dimensional convolutional neural networks for spectroscopic signal regression. Journal of Chemometrics, 2018. 32(5): p. e2977.

[7] Xu, L., et al., Using deep learning algorithms to perform accurate spectral classification. Optik, 2021. 231: p. 166423.

[8] Pouyanfar, S., et al., A survey on deep learning: Algorithms, techniques, and applications. ACM Computing Surveys (CSUR), 2018. 51(5): p. 1-36.
[9] Said, M., A. Wahba, and D. Khalil, Semi-supervised deep learning framework for milk analysis using NIR spectrometers. Chemometrics and Intelligent Laboratory Systems, 2022. 228: p. 104619.
[10] Kumar, J.S., et al. Quality assessment and grading of milk using sensors and neural networks. in 2022 International Conference on Electronics and Renewable Systems (ICEARS). 2022. IEEE.
[11] Bhavsar, D., et al., Milk Quality Prediction Using Machine Learning. EAI Endorsed Transactions on Internet of Things, 2024. 10.
[12] Mu, F., et al., Milk source identification and milk quality estimation using an electronic nose and machine learning techniques. Sensors, 2020. 20(15): p. 4238.
[13] Vidhya, S., et al. Milk quality prediction using supervised machine learning technique. in International Conference on Emerging Trends in Expert Applications & Security. 2023. Springer.
[14] Mota, L.F., et al., Evaluating the performance of machine learning methods and variable selection methods for predicting difficult-to-measure traits in Holstein dairy cattle using milk infrared spectral data. Journal of Dairy Science, 2021. 104(7): p. 8107-8121.
[15] Slob, N., C. Catal, and A. Kassahun, Application of machine learning to improve dairy farm management: A systematic literature review. Preventive Veterinary Medicine, 2021. 187: p. 105237.
[16] Li, X., CNN-GRU model based on attention mechanism for large-scale energy storage optimization in smart grid. Front. Energy Res, 2023. 11: p. 1228256.
[17] Targ, S., D. Almeida, and K. Lyman, Resnet in resnet: Generalizing residual architectures. arXiv preprint arXiv:1603.08029, 2016.
[18] Milk quality prediction dataset.