UNIVERSITI TEKNIKAL MALAYSIA MELAKA

MILK QUALITY ANALYSIS WITH MULTIMODAL DATA FUSION: COMBINING IMAGE AND NUMERICAL FEATURES

PUTERI NUR FARZANAH FAGHIRA BINTI KAMARUDIN

MASTER OF SCIENCE IN ELECTRONIC ENGINEERING

2025

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

Faculty of Electronics and Computer Technology and Engineering

MILK QUALITY ANALYSIS WITH MULTIMODAL DATA FUSION: COMBINING IMAGE AND NUMERICAL FEATURES

Puteri Nur Farzanah Faghira Binti Kamarudin

Master of Science in Electronic Engineering

2025

MILK QUALITY ANALYSIS WITH MULTIMODAL DATA FUSION: COMBINING IMAGE AND NUMERICAL FEATURES

PUTERI NUR FARZANAH FAGHIRA BINTI KAMARUDIN

A thesis submitted
in fulfillment of the requirements for the degree of
Master of Science in Electronic Engineering

Faculty of Electronics and Computer Technology and Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2025

DECLARATION

I declare that this thesis entitled “Milk Quality Analysis with Multimodal Data Fusion: Combining Image and Numerical Features“ is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :.....

Name : Puteri Nur Farzanah Faghira Binti Kamarudin

Date : 12 September 2025

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Master of Science in Electronic Engineering

Signature :.....

Supervisor Name :Ir. Gs. Ts. Dr. Nik Mohd Zarifie bin Hashim

Date :13 September 2025

DEDICATION

To my beloved mother, Raja Nor Asikin Binti Raja Kamil and father, Kamarudin Bin Bahari.

ABSTRACT

Ensuring the quality of milk is a critical challenge in the food industry, with significant implications for consumer safety, economic efficiency, and supply chain reliability. Traditional methods of milk quality assessment, such as chemical analysis and sensory evaluations, while accurate, are labor-intensive, costly, and unsuitable for real-time or large scale applications. The growing adoption of artificial intelligence (AI) has introduced advanced methods for automating these processes, yet most existing AI-based approaches rely on single-modality data, such as either visual features or numerical measurements. These methods are often limited in their ability to capture the multidimensional nature of milk quality indicators, resulting in reduced prediction accuracy and robustness. This study seeks to address these limitations by developing a multimodal deep learning model that combines image and numerical data for classifying milk quality into three categories which are good, spoiling, and spoiled. This study employs intermediate fusion and late fusion techniques to combine the outputs of pre-trained models for each modality. This study also highlights the potential of multimodal deep learning in addressing the complex interplay of physical and chemical factors influencing milk quality. Results show that the intermediate fusion technique achieved an accuracy of 98.87% while late fusion technique using concatenation with proposed layers achieved an accuracy of 99.77% This proves that the multimodal framework outperforms single-modality approaches in terms of accuracy, scalability, and generalizability across diverse milk storage and spoilage conditions. By incorporating complementary data sources, the proposed framework achieves a holistic and automated approach to milk quality assessment, suitable for industrial-scale applications. Additionally, the study contributes to the advancement of fusion strategies in multimodal AI, demonstrating the efficacy of combining heterogeneous data for improved decision making. While the findings are promising, the research also identifies several areas for further investigation. Future work could explore the integration of additional modalities, such as odor sensors or spectral data, to further enhance classification performance. Extending the framework to other perishable goods could also validate its applicability in broader food quality assessment contexts. This study not only bridges an important gap in the literature but also sets a foundation for scalable, efficient, and robust AI-driven solutions in food safety and quality control.

ANALISIS KUALITI SUSU DENGAN PENCANTUMAN DATA MULTIMODAL: MENGGABUNGKAN CIRI IMEJ DAN BERANGKA

ABSTRAK

Menjamin kualiti susu merupakan satu cabaran kritikal dalam industri makanan, dengan implikasi besar terhadap keselamatan pengguna, kecekapan ekonomi, dan kebolehpercayaan rantaian bekalan. Kaedah tradisional untuk menilai kualiti susu seperti analisis kimia dan penilaian deria, walaupun tepat, memerlukan tenaga kerja yang tinggi, mahal, dan tidak sesuai untuk aplikasi masa nyata atau berskala besar. Penerapan kecerdasan buatan (AI) yang semakin berkembang telah memperkenalkan kaedah canggih untuk mengautomasikan proses ini. Namun begitu, kebanyakan pendekatan berasaskan AI yang sedia ada masih bergantung kepada data satu-modality, sama ada ciri visual atau pengukuran berangka. Pendekatan ini sering terhad dalam keupayaan untuk menangkap sifat multidimensi penunjuk kualiti susu, yang akhirnya mengurangkan ketepatan ramalan dan keteguhan model. Kajian ini bertujuan untuk menangani kekangan tersebut dengan membangunkan model pembelajaran mendalam multimodal yang menggabungkan data imej dan data berangka untuk mengklasifikasikan kualiti susu kepada tiga kategori iaitu baik, mula rosak, dan rosak. Kajian ini menggunakan teknik fusion pertengahan (intermediate fusion) dan fusion lewat (late fusion) untuk menggabungkan output daripada model pra-latih bagi setiap modality. Empat kaedah fusion diuji bagi fusion lewat iaitu penggabungan (concatenation), pengumpulan maksimum (max pooling), pengundian ensemble (voting ensemble), dan purata berwajaran (weighted averaging). Keputusan menunjukkan bahawa rangka kerja multimodal mengatasi pendekatan satu-modality dari segi ketepatan, kebolehsuaian skala, dan keumuman merentas pelbagai keadaan penyimpanan dan kerosakan susu. Penyelidikan ini menyerlahkan potensi pembelajaran mendalam multimodal dalam menangani interaksi kompleks antara faktor fizikal dan kimia yang mempengaruhi kualiti susu. Dengan menggabungkan sumber data yang saling melengkapi, rangka kerja yang dicadangkan ini menyediakan pendekatan yang menyeluruh dan automatik dalam penilaian kualiti susu yang sesuai untuk aplikasi skala industri. Selain itu, kajian ini menyumbang kepada kemajuan strategi fusion dalam AI multimodal dengan membuktikan keberkesanan gabungan data heterogen bagi meningkatkan proses membuat keputusan. Walaupun penemuan yang diperoleh adalah memberangsangkan, kajian ini juga mengenal pasti beberapa aspek untuk penyelidikan lanjut. Kajian masa hadapan boleh meneroka integrasi modality tambahan seperti sensor bau atau data spektrum bagi meningkatkan lagi prestasi klasifikasi. Memperluas rangka kerja ini kepada barangan mudah rosak yang lain juga dapat mengesahkan kebolehgunaan pendekatan ini dalam konteks penilaian kualiti makanan yang lebih meluas. Kajian ini bukan sahaja merapatkan jurang penting dalam literatur, tetapi juga meletakkan asas kepada penyelesaian AI yang boleh diskalakan, cekap dan teguh dalam keselamatan makanan dan kawalan kualiti.

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious, the Most Merciful. First and foremost, I would like to take this opportunity to express my sincere acknowledgment to my supervisor, Ir. Gs. Ts. Dr. Nik Mohd Zarifie bin Hashim, for their invaluable guidance, continuous support, and encouragement throughout my research journey. Their expertise, patience, and constructive feedback have been instrumental in shaping this work, and I am truly grateful for their unwavering dedication. I would also like to extend my heartfelt appreciation to my beloved parents, whose endless love, prayers, and sacrifices have been my greatest source of strength. Their unwavering support and belief in my abilities have been the driving force behind my academic pursuits, and I am forever indebted to them. Furthermore, I am deeply thankful to the Faculty of Electronics and Computer Technology and Engineering for providing the necessary resources and a conducive learning environment that has greatly contributed to the success of this research. My gratitude also goes to my lecturers, colleagues, and friends who have offered their support, insights, and motivation throughout this journey. Above all, I thank Allah (SWT) for granting me the strength, patience, and perseverance to complete this thesis. Without His guidance and blessings, none of this would have been possible.

TABLE OF CONTENTS

		PAGES
DECLARATION		
APPROVAL		
DEDICATION		
ABSTRACT		**i**
ABSTRAK		**ii**
ACKNOWLEDGEMENT		**iii**
TABLE OF CONTENTS		**iv**
LIST OF TABLES		**vi**
LIST OF FIGURES		**ix**
LIST OF ABBREVIATIONS		**xi**
LIST OF APPENDICES		**xiii**
LIST OF PUBLICATIONS		**xiv**
CHAPTER		
1.	**INTRODUCTION**	**1**
	1.1 Background	1
	1.2 Problem Statement	5
	1.3 Research Question	8
	1.4 Research Objective	10
	1.5 Scope of Research	10
	1.6 Thesis Outline	12
2.	**LITERATURE REVIEW**	**14**
	2.1 Introduction	14
	2.2 Traditional Methods for Milk Quality Assessment	15
	2.2.1 Traditional Methods	15
	2.2.2 Limitations	18
	2.3 Machine Learning and Deep Learning in Milk Quality Assessment	19
	2.3.1 Single-Modality Models	22
	2.3.2 Visual Analysis	25
	2.3.3 Numerical Analysis	25
	2.3.4 Limitations of Single-Modality Approaches	26
	2.4 Multimodal Deep Learning	29
	2.4.1 Early Fusion	31
	2.4.2 Intermediate Fusion	33
	2.4.3 Late Fusion	43
	2.4.4 Studies on All Fusion Techniques	50
	2.4.5 Challenges in Multimodal Deep Learning	54
	2.5 Related Works on Multimodal Deep Learning in Food Safety	55
	2.6 Classification Metrics	59
	2.7 Summary	60
3.	**METHODOLOGY**	**62**
	3.1 Introduction	62

3.2 Multimodal Dataset Construction and Preparation 64
3.3 Classification Metrics 76
3.3.1 Confusion Matrix 76
3.3.2 Accuracy 78
3.3.3 Precision 78
3.3.4 Recall 79
3.3.5 F1-Score 79
3.4 Qualitative Milk Quality Analysis 80
3.5 Quantitative Milk Quality Analysis 83
3.5.1 Intermediate Fusion Technique 84
3.5.2 Late Fusion Technique 87
3.6 Summary 95

4. RESULT AND DISCUSSION 97
4.1 Introduction 97
4.2 Milk Qualitative Analysis 97
4.3 Milk Quantitative Analysis via Intermediate Fusion 107
4.4 Milk Quantitative Analysis via Late Fusion 119
4.4.1 Visual Data Analysis 119
4.4.2 Numerical Data Analysis 121
4.4.3 Late Fusion Results 123
4.5 Intermediate Fusion versus Late Fusion 147
4.6 Imbalance dataset 150
4.7 Summary 151

5. CONCLUSION AND RECOMMENDATIONS FOR FUTURE RESEARCH 152
5.1 Introduction 152
5.2 Summary of the Research Objectives 152
5.3 Research Contributions 155
5.4 Pratical Implications and Beneficiaries 157
5.5 Limitions of The Present Study 159
5.6 Future Works 160
5.7 Summary 161

6. REFERENCES 163
APPENDICES 184

LIST OF TABLES

TABLE	TITLE	PAGE
Table 2.1	Summary of traditional methods used in milk spoilage analysis	17
Table 2.2	Summary of MMDL that utilizes intermediate fusion	42
Table 2.3	Summary of MMDL that utilizes late fusion	49
Table 2.4	Summary of MMDL that utilizes multiple fusion	53
Table 2.5	Summary of studies that utilize MMDL in food safety	58
Table 3.1	Controlled samples preparation	65
Table 3.2	Sample of numerical data	69
Table 3.3	Numerical representation for storage condition	69
Table 3.4	Numerical representation for exposure status	70
Table 3.5	Numerical representation for odor	70
Table 3.6	Summary of visual dataset	72
Table 3.7	Visual data split into training, validation and testing	73
Table 3.8	Summary of numerical dataset	75
Table 3.9	Numerical data split into training, validation and testing	75
Table 3.10	The four components of confusion matrix	76
Table 3.11	Confusion matrix	77
Table 3.12	Numerical data split into training, validation and testing	87
Table 3.13	Hyperparameters for each visual model	89
Table 3.14	Epoch and batch size for first pair	92
Table 4.1	Overall result of ground truth versus human analysis (image only)	102
Table 4.2	Overall result of ground truth versus human analysis (image and numerical)	104
Table 4.3	Image only versus image and numerical	105

Table 4.4 Accuracy percentage and model training time for intermediate fusion 107

Table 4.5 Confusion matrices for intermediate fusion at 10 epochs 111

Table 4.6 Confusion matrices for intermediate fusion at 25 epochs 112

Table 4.7 Confusion matrices for intermediate fusion at 50 epochs 113

Table 4.8 Classification report for intermediate fusion model (10 epochs) 116

Table 4.9 Classification report for intermediate fusion model (25 epochs) 117

Table 4.10 Classification report for intermediate fusion model (50 epochs) 118

Table 4.11 Training and validation accuracy for Random Forest and XGBoost 123

Table 4.12 Accuracy percentage and model training time for late fusion (concatenation with proposed layers) 124

Table 4.13 Confusion matrices for late fusion (concatenation with proposed layers) 127

Table 4.14 Classification report for concatenation with proposed layers (10 epochs) 130

Table 4.15 Classification report for concatenation with proposed layers (25 and 50 epochs) 131

Table 4.16 Model combination. 132

Table 4.17 Accuracy percentage and model training time for concatenation with proposed layers using other pairs of model combination (Epoch=10, batch size=64) 133

Table 4.18 Accuracy percentage and model training time for late fusion (concatenation) 135

Table 4.19 Confusion matrix for late fusion (concatenation) 136

Table 4.20 Classification report for late fusion (concatenation) 137

Table 4.21 Accuracy percentage and model training time for late fusion (max pooling) 138

Table 4.22 Confusion matrix for late fusion (max pooling) 139

Table 4.23 Classification report for late fusion (max pooling) 140

Table 4.24 Accuracy percentage and model training time for late fusion (voting ensemble) 141

Table 4.25 Confusion matrix for late fusion (voting ensemble) 142

Table 4.26 Classification report for late fusion (voting ensemble) 143

Table 4.27 Accuracy percentage and model training time for late fusion (weighted average) 144

Table 4.28 Confusion matrix for late fusion (weighted average) 145

Table 4.29 Classification report for late fusion (weighted average) 146

Table 4.30 Optimum model training time for each technique 149

Table 4.31 Pros and cons of intermediate and late fusion 150

LIST OF FIGURES

FIGURE	TITLE	PAGE
Figure 3.1	Overall proposed method workflow	63
Figure 3.2	Milk samples in (a) room temperature, and (b) in refrigerator	67
Figure 3.3	Setup of milk sample collection	67
Figure 3.4	(a) Sample image from the top-view image of milk in carton, and (b) top-view image of milk in glass cup	68
Figure 3.5	(a) A chocolate drink, and (b) a scenery	71
Figure 3.6	Image augmentation techniques	72
Figure 3.7	Example: question 4 from Section 1, where * indicates the question is compulsory to be answered	81
Figure 3.8	Example: question 4 from Section 2, where * indicates the question is compulsory to be answered	82
Figure 3.9	Workflow for Intermediate Fusion	85
Figure 3.10	Intermediate Fusion	86
Figure 3.11	Workflow for Late Fusion	88
Figure 3.12	Concatenation with proposed layers	91
Figure 3.13	Concatenation	93
Figure 3.14	Max pooling with proposed layers	94
Figure 3.15	Majority voting	94
Figure 3.16	Weighted average	95
Figure 4.1	Gender distribution in the survey	98
Figure 4.2	Age distribution in the survey	99
Figure 4.3	Occupation among respondents	100
Figure 4.4	Milk drinking frequency among respondents	101

Figure 4.5 Effect of epoch and batch size on model training time (intermediate fusion) 110

Figure 4.6 Visual models' accuracy 120

Figure 4.7 Numerical models' accuracy 122

Figure 4.8 Effect of epoch and batch size on model training time for late fusion (concatenation with proposed layers) 126

Figure 4.9 Accuracy percentage, Intermediate Fusion versus Late Fusion 148

LIST OF ABBREVIATIONS

UTeM	-	Universiti Teknikal Malaysia Melaka
AI	-	Artificial Intelligence
MMDL	-	Multimodal Deep Learning
MBS	-	Micro Biological Survey
SCC	-	Somatic Cell Counts
VOC	-	Volatile Organic Compound
PCA	-	Principal Component Analysis
NFC	-	Near Field Communication
PDA	-	Polydiacetylene
ML	-	Machine Learning
DL	-	Deep Learning
SVM	-	Support Vector Machines
ResNet	-	Residual Networks
GRU	-	Gated Recurrent Units
NIR	-	Near-Infrared
VC-SVM	-	Variable Cluster–Support Vector Machine
CNN	-	Convolutional Neural Network
ASD	-	Autism Spectrum Disorder
STFT	-	Short-Time Fourier Transform
PU	-	Paderborn University
AD	-	Alzheimer's Disease
KCCA	-	Kernel Canonical Correlation Analysis

ADNI	-	Alzheimer’s Disease Neuroimaging Initiative
NSCLC	-	Non-Small Cell Lung Cancer
ADC	-	Adenocarcinoma
SQC	-	cell carcinoma
MINT	-	Multi-stage INTermediate fusion
EEG	-	Electroencephalogram
ET	-	Eye Tracking
MTNet	-	Multimodal Transformer Network
AFF	-	Atypical Femur Fractures
NFF	-	Normal Femur Fractures
EHR	-	Electronic Health Record
AUC	-	Area Under the Curve
SNP	-	Single Nucleotide Polymorphism
MCI	-	Mild Cognitive Impairment
CN	-	cognitively normal
TCN	-	Temporal Convolutional Networks
DNN	-	Deep Neural Network

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
Appendix A	Milk Analysis via Qualitative Assessment: Section 1	184
Appendix B	Milk Analysis via Qualitative Assessment: Section 2 (sample Question 1)	185
Appendix C	Milk Analysis via Qualitative Assessment: Section 3 (sample Question 1)	186

LIST OF PUBLICATIONS

The followings are the list of publications related to the work on this thesis:

Journal Publication:

i. **P. N. F. F. Kamarudin**, N. M. Z. Hashim, 2025. Design and Application of a Custom Late Fusion Layer for Image-Numerical Milk Quality Analysis. *International Journal of Research and Innovation in Social Science (IJRISS)*, Volume 9 Issue 8 (August 2025). (Publication Process)

ii. **P. N. F. F. Kamarudin**, N. M. Z. Hashim, M. M. Ibrahim, M. D. Sulistiyo, 2024. Milk Spoilage Classification through Integration of RGB and Thermal Data Analysis. *International Journal of Computing and Digital Systems*, 15(1), pp. 1839-1851.

Proceeding Publication:

i. **P. N. F. F. Kamarudin**, N. M. Z. Hashim, N. Z. Abd Rahman, 2025. Qualitative Validation for Multimodal Deep Learning in Milk Quality Classification, *International Conference on Computer, Information Technology and Intelligent Computing 2025 (CITIC)*, 2025. (Publication Process)

ii. **P. N. F. F. Kamarudin**, N. Z. Abd Rahman, N. M. Z. Hashim, 2025. CNN-Powered Real-Time Classification of Milk Spoilage via RGB Images, *International Conference on Computer, Multimedia University Engineering Conference 2025 (MECON)*, 2025. (Publication Process)

iii. **P. N. F. F. Kamarudin**, N. M. Z. Hashim, 2023. Milk Qualitative Analysis for School Milk Program, *APS Proceedings Volume 8*

Innovation and Invention Competition Award:

i. Gold Medal Award in Malaysian Grand Invention Expo 2023 for the innovation project of Milk Qualitative Analysis for School Milk Program (M.Q.A-4-S.M.P)

ii. Gold Award in Jejak Inovasi UTeM 2023 for the innovation project of Q.I.S.M.A

CHAPTER 1

INTRODUCTION

1.1 Background

Artificial intelligence (AI) has become a transformative force across numerous industries, revolutionizing fields ranging from healthcare (Dipietro et al., 2024; Murugan et al., 2024) and autonomous systems (Aruna et al., 2024; Dakic et al., 2024) to agriculture (Ballester et al., 2024; Chen et al., 2024 b; Logeshwaran et al., 2024) and manufacturing (Orabi et al., 2024). Central to this revolution is deep learning, a subset of AI that excels in extracting meaningful patterns from large and complex datasets. Deep learning's ability to model intricate relationships in data has made it the foundation for numerous innovative solutions. Among its recent breakthroughs, multimodal deep learning has emerged as a particularly powerful approach for addressing complex classification and decision-making problems (Mathematics et al., 2024).

Multimodal deep learning is characterized by its ability to process and integrate information from multiple data sources, or modalities, which may include visual, numerical, textual, or audio data. By combining these diverse forms of data, the technique leverages the complementary strengths of each modality, capturing a richer and more holistic representation of the underlying phenomenon. This integration enables multimodal systems to outperform single-modality approaches, which often suffer from incomplete or biased representations (Singh et al., 2023). Moreover, multimodal deep learning enhances

robustness by compensating for missing or noisy data in one modality with information from others, making it particularly suited to real-world applications (Wei et al., 2024).

Multimodal deep learning is particularly valuable in applications where diverse types of data contribute complementary insights, enabling a deeper understanding of complex phenomena. By combining distinct data sources, this approach allows systems to exploit the strengths of each modality, creating a more comprehensive and accurate representation than single-modality methods can achieve (Singh et al., 2024). In the context of food quality assessment, multimodal techniques are especially impactful, as they can integrate multiple forms of information, such as visual data capturing physical characteristics like color, texture, or structural changes, and numerical data reflecting environmental or chemical properties such as pH levels, temperature, or storage time.

Milk, being a highly perishable and widely consumed food product (Boudahri et al., 2022; De Klerk et al., 2022), exemplifies an application area where such integration is critical. As milk quality is influenced by numerous factors, such as microbial activity, storage conditions, and environmental exposure, relying on a single modality can result in incomplete or unreliable assessments. Visual indicators like discoloration or changes in texture may suggest spoilage, but these alone are insufficient for precise classification. Similarly, numerical measurements such as pH or temperature provide valuable quantitative insights but may not capture the full picture. Combining these modalities through multimodal deep learning can address these limitations, offering a more robust and holistic assessment.

Accurate and efficient milk quality evaluation is essential not only for ensuring consumer safety but also for reducing economic losses and maintaining public trust in food

supply chains (Kumar et al., 2024; Nukapeyi et al., 2024). By leveraging multimodal data, advanced AI systems can contribute to the development of scalable and automated quality control solutions, which are vital for meeting the demands of modern food production and distribution systems. Traditional methods of milk quality assessment, such as laboratory analyses (Ibrahim et al., 2023; Mahrous et al., 2023) and sensory evaluations (Kim et al., 2022), have long been the standard for ensuring product safety and freshness. These techniques typically involve precise chemical tests, microbial analysis, or expert sensory judgment to determine spoilage or contamination levels. While effective, these approaches are inherently resource-intensive, requiring specialized equipment, trained personnel, and considerable time to produce results. In large-scale production and supply chain contexts, these limitations can lead to inefficiencies, delays, and increased costs, underscoring the need for faster, more scalable solutions.

AI-driven approaches, particularly those leveraging machine learning and deep learning, have emerged as transformative alternatives to traditional methods (Temilade Abass et al., 2024). By automating quality assessment processes, these technologies can significantly reduce the time and resources required, enabling real-time monitoring and rapid decision-making. However, many current AI implementations focus on single-modality data, such as analyzing visual features like color changes or sediment formation, or processing numerical attributes such as pH levels, temperature, and storage time. While these single-modality approaches have shown promise, they often fail to capture the intricate relationships between diverse data types, limiting their overall accuracy and reliability (Dubey, 2023).

Multimodal techniques offer a compelling solution by combining image data with numerical information, leveraging the strengths of both modalities to deliver more robust predictions (Dao, 2022). For instance, visual data can detect subtle changes in milk's physical appearance, such as discoloration or texture changes, which might indicate spoilage. At the same time, numerical data provides critical quantitative insights, such as deviations in pH levels or temperature thresholds, which are not observable through images alone. By integrating these data sources through advanced deep learning architectures, multimodal methods can achieve a more holistic understanding of milk quality, reducing false positives and negatives and increasing the reliability of predictions.

Despite its potential, the application of multimodal deep learning in milk quality classification remains underexplored, with existing research often constrained by a reliance on single-modality approaches. These methods typically analyze either visual data, such as images capturing physical changes in the milk, or numerical data, such as measurements of pH, temperature, or storage conditions. While these single-modality approaches can provide valuable insights, their isolated nature often results in suboptimal outcomes, as they fail to capture the full spectrum of information that multimodal data offers. This limitation highlights the need for more sophisticated solutions that integrate multiple data types for a more comprehensive analysis.

Multimodal deep learning has the potential to address these shortcomings by combining visual and numerical modalities, creating a synergistic framework that enhances classification accuracy and robustness. For example, visual data might identify surface discoloration or textural changes, while numerical data provides quantifiable metrics like temperature fluctuations or pH levels, which are critical indicators of spoilage. Together,