Applied Science and Engineering Progress, Vol. 20, No. 3, 2027, 8286

Research Article

A comparative evaluation of agentic AI-assisted model and visual inspection for a deep learning analysis model: An empirical study of lime quality classification in compliance with Codex Standard 217-1999

Alongkorn Klaiklueng
Advanced Industrial Engineering Management Systems Research Center, Department of Industrial Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

Chavittha Kengpol*
Nonthaburi Wittayalai School, Nonthaburi, Thailand

* Corresponding author. E-mail: chavittha.kengpol@gmail.com
DOI: 10.14416/j.asep.2026.08.014
Received: 30 April 2026; Revised: 11 June 2026; Accepted: 9 July 2026; Published online: 25 August 2026

Abstract
Accurate quality classification is essential in fruit trade, particularly for peel-based grading. However, conventional visual inspection is often subjective and inconsistent. The literature review identifies a research gap: no Codex-aligned lime grading model has applied an agentic AI-assisted model while systematically comparing regression models. This gap highlights the need to align data structure, model assumptions, and predictive performance in the development of standard-aligned classification models. The objective of this research is to design a deep learning (DL)-based image analysis model incorporating an agentic AI-assisted model for fruit quality classification and to evaluate its performance against conventional visual inspection. Experimental results show that the proposed model achieves over 90% classification accuracy, significantly outperforming visual inspection (55%). The model integrates Visual Geometry Group 19 (VGG19) for the image classification model with ordinal logistic regression (OLR) to classify lime quality attributes. VGG19 reached 94% accuracy, while OLR showed significant predictors and strong ordinal associations, confirming its suitability for Codex-aligned quality grading. The model also provides explainable outputs through measured lime attributes and class probabilities. The contribution of this research is the development of a Codex-aligned lime quality classification model that integrates a Convolutional Neural Network (CNN)-based deep learning model, OLR, and a Large Language Model (LLM)-driven agentic AI-assisted model for automated and interpretable quality assessment based upon Codex Standard 217-1999. Its advantage lies in improving inspection standardization, reducing subjectivity, and enhancing grading efficiency. The benefit of the proposed model is improving quality control reliability while reducing missed produce sales opportunities for farmers.

Keywords: Agentic AI-assisted model, Classification, Codex standard, Deep learning, Regression, Visual inspection

1 Introduction

An agentic AI-assisted model is an artificial intelligence approach that enables systems to autonomously perceive, reason, and make decisions in complex environments [1]. Unlike conventional AI, which primarily focuses on prediction or classification, the agentic AI-assisted model integrates perception with reasoning to support more structured and adaptive quality decisions. The advantage of an agentic AI-assisted model lies in its ability to continuously learn, adapt, and improve decision-making while reducing human intervention and minimizing errors caused by bias and fatigue. As a result, it enhances the reliability, objectivity, and standardization of quality assessment [1].

This advantage is particularly important in fruit grading, as many fruits share similar external characteristics, such as color, size, and surface defects, making differentiation through visual inspection difficult. The limitation of visual inspection lies in its dependence on human perception, which can vary among inspectors and lead to subjective judgment, inconsistency, and errors caused by fatigue. Previous studies reported that human accuracy ranges from 70% to 80%, with missed defect rates ranging from

20% to 30% [2]. These findings highlight the need for a more reliable, accurate, and standardized grading approach. Using an agentic AI-assisted model can improve fruit grading by reducing human involvement and delivering more consistent results, while enhancing accuracy, quality control, and standardization compared to traditional visual inspection. Among agricultural products, limes are an appropriate case study for evaluating the effectiveness of an agentic AI-assisted model due to their high economic value and strong reliance on external visual attributes. Limes (*Citrus aurantifolia* Swingle) are globally important crops, widely used across culinary cultures and valued for their high nutritional content, particularly vitamin C. As illustrated in Figure 1, lime cultivation is distributed worldwide, with a global market value estimated at USD 20.2 billion in 2024 [3]. Mexico is the leading producer and exporter, followed by Spain and Turkey [4].

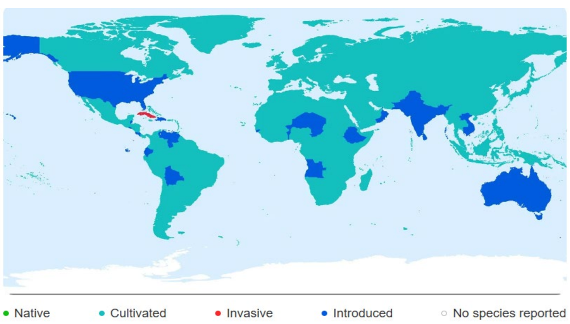

Figure 1: Global distribution and occurrence status of lime [3].

The market value of limes is primarily determined by external attributes, such as uniform green color, standardized size, and the absence of defects, as defined in Codex Standard 217-1999. The standard classifies limes into Extra class, Class I, and Class II, as illustrated in Figure 2. While the quality levels are ordinal, the underlying feature data, such as color, brightness, and surface defects, are continuous variables that exhibit uneven distributions and overlap, particularly among adjacent classes. This condition contributes to ambiguity and inconsistency in visual inspection. These characteristics indicate that the data contain both ordinal structure and continuous properties, making it necessary to compare mathematical models such as multinomial logistic regression (MLR) and ordinal logistic regression (OLR) to determine the most appropriate model for lime quality classification under Codex Standard 217-1999.

The literature review identifies a research gap, as no evidence has been found of a Codex-aligned lime quality classification model that applies an agentic AI-assisted model, integrates CNN-based deep learning, and compares regression models. This gap highlights the need to align data structure, model assumptions, and predictive performance in the development of standard-aligned classification models. In this research, a CNN extracts features such as peel color, texture, and defects, while an OLR model classifies lime quality data. In addition, comparative evaluations between visual inspection and agentic AI-assisted models under standardized conditions remain limited, highlighting the significance of this study.

The objective of this research is to design a deep learning-based image analysis model incorporating an agentic AI-assisted model for fruit quality classification and to evaluate its performance against conventional visual inspection. The contribution of this research is the development of a Codex-aligned and interpretable lime quality classification model that integrates a CNN-DL model, OLR, and an LLM-driven agentic AI-assisted model for practical and automated quality assessment. The benefit is improving the accuracy, reliability, and explainability of quality control beyond visual inspection while maintaining compliance with Codex Standard 217-1999 and reducing missed sales opportunities for farmers.

Figure 2: Codex based upon lime quality classes.

2 Literature Review

2.1 *Artificial intelligence, deep learning*

Artificial intelligence (AI) enables machines to perform perception and reasoning tasks in a human-like manner. Deep learning (DL), a subfield of AI, uses multi-layer neural networks to automatically learn complex patterns from data [1], [5], as illustrated in Figure 3. Due to its ability to extract discriminative features directly from

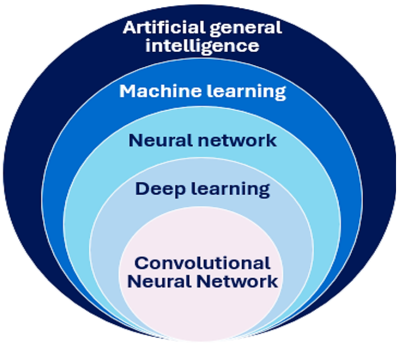

Figure 3: Hierarchical relationship of artificial intelligence technologies [1].

images, DL has become an effective approach for fruit quality assessment and image classification tasks [6]. Recent studies have also demonstrated the effectiveness of IoT-assisted and AI-based intelligent decision-making in smart agriculture applications, including context-aware fertilizer recommendation, evapotranspiration estimation, and precision irrigation systems [7], [8]. In addition, recent research has highlighted the growing role of DL in IoT-enabled precision agriculture and crop monitoring systems for sustainable agricultural management [9], [10].

Convolutional neural networks (CNNs), a widely used DL architecture, perform hierarchical feature extraction for image classification, as shown in Figure 4. CNN-based models have demonstrated strong performance in fruit grading and agricultural image analysis [6], with architectures such as Visual Geometry Group 19 (VGG19), AlexNet, and Residual Network 50 (ResNet50) widely adopted for image-based classification tasks. Despite their high predictive performance, CNN models are often criticized for their limited interpretability and inability to perform reasoning beyond pattern recognition. Recent advances in Large Language Models (LLMs) have enabled the development of an agentic AI-assisted model capable of explainable decision-making through multi-step workflows. These capabilities enable intelligent decision-making by integrating prediction results, mathematical inference, and domain-specific standards before reaching a final decision [10]-[12].

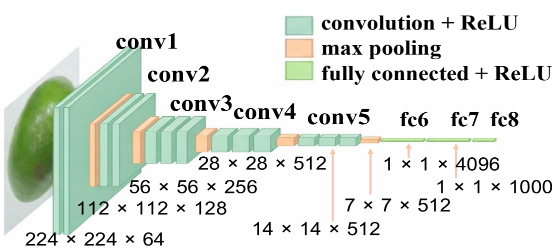

Figure 4: Convolutional neural network (CNN).

2.1.1 Agentic AI-assisted model

An agentic artificial intelligence (AI)-assisted model refers to AI systems that can autonomously perceive, reason, plan, and execute actions toward achieving specific goals through multi-step interactions with external tools and environments [1]. This characteristic distinguishes it from traditional AI, which typically produces outputs directly from inputs in a single step [1]. In contrast, an agentic AI-assisted model operates through iterative processes involving reasoning, planning, execution, memory, and validation [13]. The differences between traditional AI and agentic AI-assisted models are summarized in Table 1 [14].

This concept improves decision-making processes in complex tasks by integrating reasoning with structured actions, allowing AI systems to process information through multiple steps before making decisions. Recent studies have highlighted the effectiveness of an agentic AI-assisted model in coordinating multiple analytical tools, enabling dynamic decision-making through multi-step analytical and validation processes [15]. Unlike traditional deep learning systems, an agentic AI-assisted model can evaluate intermediate outputs and adapt its decision process according to task requirements. However, the application of agentic AI-assisted models in fruit quality classification remains limited, especially in tasks requiring standard-based quality assessment such as Codex-based lime classification.

Table 1: Comparison of AI and agentic AI-assisted models [1].

Aspect	Traditional AI	Agentic AI -assisted model
Operation	Single-step	Multi-step
Decision process	Based upon predefined models	Involves reasoning and planning
Flexibility	Low	High
Task complexity	Suitable for simple, well-defined tasks	Suitable for complex tasks
Interaction	Limited	Can interact with tools and environments
Limitations	Lacks multi-step reasoning	More complex and resource-intensive

2.2 *Lime and Codex Standard 217-1999*

Lime (*Citrus aurantifolia* Swingle) is a valuable citrus crop and an important export commodity, particularly in Mexico [16]. It is a medium-sized shrub that produces round green fruits, as shown in Figure 5. Postharvest quality, especially peel color, strongly influences market acceptance, with deep green fruit being preferred over yellow fruit. Improper temperature control can accelerate deterioration or induce chilling injury. The Codex Standard 217-1999 defines minimum quality requirements and grade categories, including Extra class, Class I, and Class II, primarily based upon peel appearance [17]. However, lime quality classification still relies on manual visual inspection, which is often subjective and inconsistent. Therefore, there is a need to apply artificial intelligence techniques to improve the accuracy and consistency of classification according to Codex criteria.

Figure 5: Lime fruit and lime plants.

2.3 *Models for mathematical quality evaluation*

Multiple linear regression (MLR) is widely applied in agricultural quality evaluation to predict continuous fruit quality attributes by combining multiple measurable variables into a single quantitative output, as expressed in Equation (1). MLR has been used in fruit quality studies to estimate internal fruit quality attributes such as firmness and other measurable properties relevant to postharvest grading [18].

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \qquad (1)$$

where Y is the predicted quality indicator, β_0 is the intercept, β_i are regression coefficients, x_i are measurable features, and ε is the error term. This model enables the estimation of fruit quality attributes such as firmness that are difficult to measure directly [18] and can integrate imaging techniques to convert visual data into quantitative indicators. Model performance is evaluated using R^2, Root Mean Square Error (RMSE) and correlation coefficients.

For ordered quality classification, ordinal logistic regression (OLR) is used to transform feature inputs into cumulative probabilities of belonging to each quality level, as expressed in Equation (2). OLR has been successfully applied in fruit grading tasks involving ordered quality categories, demonstrating its suitability for agricultural classification problems in which grade labels follow a ranked structure [19].

$$Logit[P(Y \leq j|x)] = \theta_j - \beta^T x,\ j = 1, \ldots, k \qquad (2)$$

where Y denotes the ordinal quality grade, j is the grade level, x is the predictor vector, θ_j defines the decision threshold between adjacent grades, and β represents the regression coefficients. The model converts features into ordered probabilities and assigns the most likely class based upon threshold comparison, making it suitable for grading tasks [20]. Model performance is evaluated using accuracy and ordinal association measures, including Somers' D, Goodman and Kruskal's gamma, and Kendall's Tau-a, to assess agreement with standard grade rankings [21].

Nevertheless, a research gap remains, as few studies have compared MLR and OLR for fruit quality classification. Most studies apply these models separately, so it remains unclear which is more appropriate for ordered quality classification. This limitation highlights the need to align data structure, model assumptions, and predictive performance in the development of standard-aligned classification models. It also provides an important theoretical perspective by explaining how outcome scale and predictor overlap influence the suitability of MLR and OLR for agricultural quality assessment.

2.4 *Cohen's kappa*

Cohen's Kappa is a statistical measure of inter-rater agreement for categorical data that adjusts for chance agreement [22]. It is widely used to evaluate consistency between visual inspection and automated classification systems in image analysis and quality assessment applications. Previous studies have applied Cohen's Kappa to assess the reliability of classification results beyond simple accuracy, particularly when comparing visual inspection outcomes with reference standard classifications [23]. Kappa values range from −1 to 1, where values of 0.61–0.80 indicate substantial agreement and values of 0.81–1.00 indicate almost perfect agreement. Therefore, Cohen's Kappa is appropriate for evaluating the level of agreement between visual inspection results and standard classifications.

2.5 *Visual inspection*

Visual inspection is a widely used approach in quality control and maintenance across various industries, involving the detection of defects and assessment of product quality based on observable surface characteristics. It plays a critical role in high-risk sectors such as aerospace and manufacturing, where even minor defects can affect safety and performance. However, traditional visual inspection is typically performed manually, making it prone to subjectivity, fatigue, and inconsistency, which can reduce reliability and repeatability of quality assessment [24].

Recent advancements in computer vision and artificial intelligence have enabled automated visual inspection systems that provide higher accuracy, consistency, and efficiency than manual visual inspection [25]. These advantages make such systems particularly suitable for agricultural applications, where product quality is often determined by visual attributes and large volumes of produce require rapid and standardized assessment. In tasks such as Codex-based lime quality classification, inspection requires not only visual feature recognition but also structured quality assessment based upon standardized criteria. Nevertheless, limited studies have compared AI-based systems and visual inspection under standardized conditions, highlighting a significant research gap [26]. Therefore, an agentic AI-assisted model, which integrates deep learning with mathematical models for quality assessment, is considered a promising approach for agricultural quality assessment.

2.6 *Research gaps and proposed direction*

Overall, the literature indicates that the integration of agentic AI-assisted models, CNNs, and regression models for Codex-aligned quality assessment remains limited. The research gaps are presented below.

- Although Codex standards are clearly defined, quality grading still relies on manual visual inspection and lacks AI-based classification models aligned with Codex standards, resulting in inconsistency and subjectivity.
- Only a limited number of studies have examined linear and ordinal regression models, highlighting the urgent need to align data structure, model assumptions, and predictive performance for Codex-aligned lime grading models.
- Although mathematical models enhance interpretability, there remains a lack of systematic integration between CNN-extracted features and agentic AI-assisted model quality assessment approaches aligned with Codex quality criteria.
- Studies involving the comparative evaluation of AI-based systems and visual inspection in agricultural classification remain limited, particularly for Codex-based lime quality assessment.

3 Materials and methods

This research develops a lime quality classification model consisting of 6 steps, as shown in Figure 6.

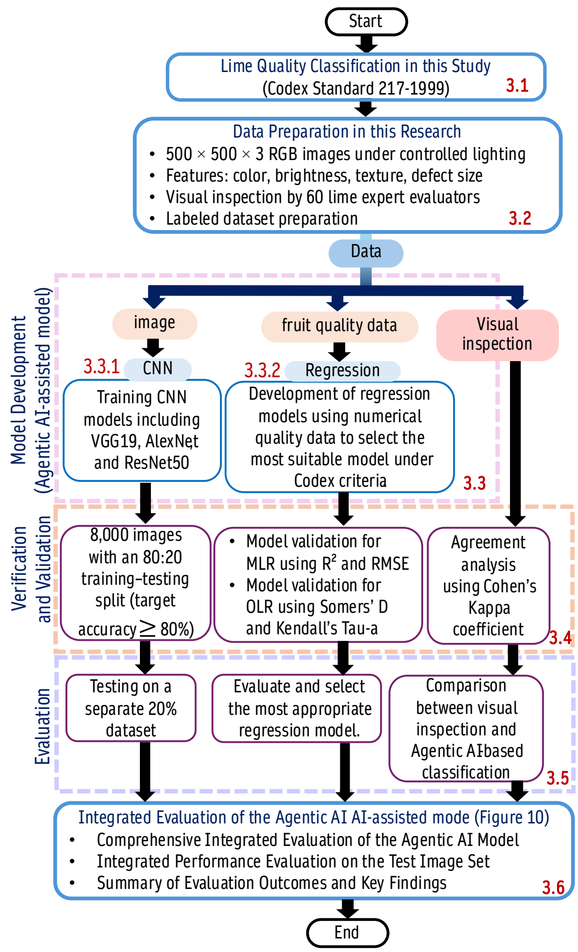

Figure 6: Research methodology.

3.1 *Lime quality classification in this study*

This study classifies lime peel into five quality groups: Extra class, Class I, and Class II based upon Codex Standard quality grades, together with Overripe and Decayed categories representing limes below the minimum quality requirements. This extended classification improves coverage for practical quality assessment, as illustrated in Figure 7.

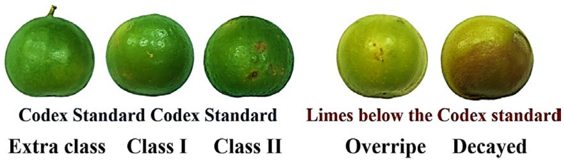

Figure 7: Examples of limes used as reference criteria.

3.2 *Data preparation in this research*

Lime images (500 × 500 × 3 RGB) are captured in a controlled studio environment to ensure consistent lighting and background conditions, with four images captured per lime sample. Visual inspection is conducted manually by three evaluators with expertise in lime quality standards to establish consensus-based ground-truth labels. Important lime quality attributes, including peel color, brightness, surface texture, and defect size, are measured using a standard colorimeter instrument and recorded to support feature extraction and regression model development. The dataset collection and validation procedures are conducted in accordance with Codex Alimentarius guidelines to ensure the accuracy and reliability of the quality labels.

Additionally, visual inspection data are collected from 60 expert lime evaluators using the same lime image assessment test based on Codex standards for comparison with the proposed AI model. The data are used to evaluate inter-rater consistency using Cohen's Kappa coefficient [22], and the results are further discussed in terms of visual inspection and compared with the proposed agentic AI-assisted model in Sections 4.1 and 4.5.

3.3 *Model development (Agentic AI-assisted model)*

This study develops an agentic AI-assisted model that integrates image perception, CNN-based feature extraction, quality assessment, and LLM-assisted explanation for automated lime classification.

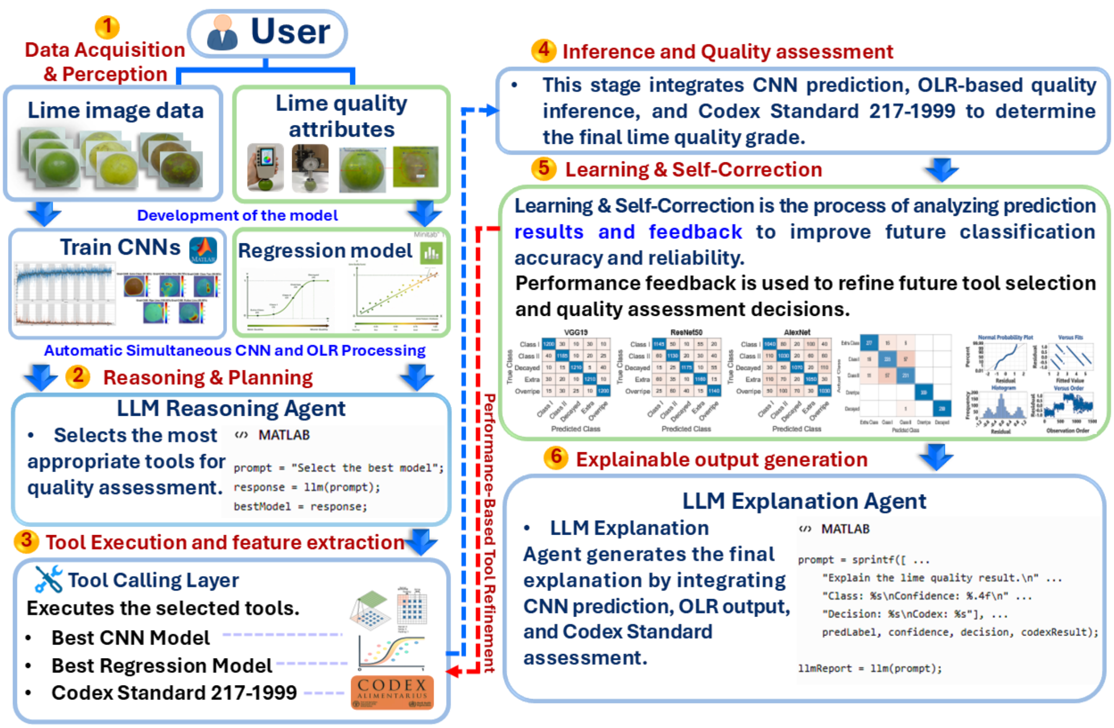

Figure 8: Agentic AI-assisted model for lime quality classification.

As illustrated in Figure 8, the model consists of 6 stages: (1) Data Acquisition and perception, (2) Reasoning and planning, (3) Tool execution and feature extraction, (4) Inference and quality assessment, (5) Learning and Self-Correction, and (6) Explainable output generation.

3.3.1 Development and training of CNN models

In this stage, corresponding to Section 3.3.1 in Figure 6 and Stage 1 (Data Acquisition and Perception) in Figure 8, the agentic AI-assisted model utilizes CNNs to extract features from lime images. This study trains and compares multiple CNN architectures, including VGG19, AlexNet, and ResNet50, to identify the most appropriate model for lime quality classification. These models are selected due to their different capabilities in extracting visual features related to color, surface texture, and defects for quality classification.

The CNN models are trained using transfer learning with data augmentation techniques, including image flipping, rotation, translation, and scaling, to improve model generalization [28]. Training is performed using the Adam optimizer for 100 epochs with a constant learning rate of 1×10^{-7} on an NVIDIA RTX 3060 Ti GPU. The experiments are implemented in MATLAB R2022a.

The selected CNN model analyzes raw image data to classify lime images into preliminary quality classes, including Extra Class, Class I, Class II, Overripe, and Decayed, based upon the Codex Standard for limes. The dataset includes diverse lime samples with variations in color, texture, defect severity, and ripeness to improve classification robustness before being used in the subsequent quality assessment process. In addition, Grad-CAM visualizations are employed to support the interpretability of classification results and improve the transparency of the model [29].

3.3.2 Development of the regression model

In this stage, corresponding to Section 3.3.2 in Figure 6, the agentic AI-assisted model operates through Stage 2 (Reasoning and Planning), Stage 3 (Tool Execution and Feature Extraction), and Stage 4 (Inference and Quality Assessment) in Figure 10. During Stage 2 (Reasoning and Planning), the LLM Reasoning Agent analyzes the measured lime quality attributes described in Section 3.2 together with CNN classification outputs to identify variables associated with each quality class, formulate an appropriate quality assessment strategy, and determine the analytical tools required for subsequent quality evaluation.

Although the quality levels are ordinal, the lime quality data, such as color, brightness, and defects, are continuous variables that can overlap across adjacent quality classes.

In Stage 3 (Tool Execution and feature extraction), the variables selected by the LLM Reasoning Agent (Llama 3.2 via Ollama) are analyzed using two regression models, namely MLR and OLR, to identify the most appropriate model for quality assessment. This comparison is necessary because the predictor variables are continuous, whereas the quality grades are ordinal.

Next, in Stage 4 (Inference and Quality Assessment), the selected regression model and analyzed lime data from Minitab 20.2 are used to develop predictive equations for explaining quality outcomes and identifying the characteristics of each quality class. These equations also support Codex-based lime quality classification and assessment.

3.4 *Verification and validation*

The verification and validation process evaluates the agentic AI-assisted model across perception, reasoning, quality assessment, and explainable output generation.

For the perception component, CNN models including VGG19, AlexNet, and ResNet50 are evaluated using a dataset of 8,000 lime images collected under controlled lighting conditions, consisting of five quality classes with 1,600 images per class. The dataset is divided into training and testing sets using an 80:20 split [27], with images from the same lime assigned to only one dataset to prevent data leakage and ensure reliable evaluation of classification accuracy and feature extraction performance.

MLR and OLR models are validated for inference and quality assessment. MLR is evaluated using R^2 and RMSE, while OLR is evaluated using Somers' D and Kendall's Tau-a. The results from both models are then used to determine the more appropriate model for lime quality classification.

The verification and validation process also supports Stage 5 (Learning and Self-Correction) by analyzing classification results, confidence scores, model agreement, and user feedback to identify potential classification errors and improve the quality assessment process.

3.5 *Evaluation*

The evaluation is conducted using a test dataset comprising 20% of the total data and separated from the training set [21] to assess the performance of the agentic AI-assisted model in classifying lime images into five quality classes.

In this stage, the performance of CNN models, including VGG19, AlexNet, and ResNet50, is evaluated and compared to determine the model with the highest classification accuracy. MLR and OLR models are also evaluated and compared to identify the most appropriate approach for quality assessment and Codex-based lime quality classification.

3.6 *Integrated model evaluation*

In the agentic AI-assisted model, Stage 6 (Explainable Output Generation) employs the LLM Explanation Agent to generate explanations of the classification results. CNN-based perception is used to extract lime features, while CNN and regression models are integrated to estimate quality outcomes, probabilities, and final quality grades according to Codex criteria. The outputs present the predicted quality class together with images and numerical indicators.

Model performance is evaluated using a test dataset to assess accuracy in lime quality grading. Visual inspection results are compared with CNN predictions to evaluate the perception capability of the agentic AI-assisted model. Regression models are validated separately to assess their performance in explaining lime quality attributes within the proposed model.

4 Results and Discussion

4.1 *Visual inspection scores and kappa analysis*

Visual inspection is conducted in two stages: three evaluators with expertise in lime quality standards establish consensus-based ground-truth labels for dataset preparation, while visual inspection scores from 60 expert lime evaluators using the same image dataset and Codex-based grading criteria are collected for comparison with the proposed agentic AI-assisted model. The visual inspection scores show moderate variability with slight positive skewness, as illustrated in Figure 9, with a mean score of 55.17 and a standard deviation of 14.19, reflecting subjective differences among evaluators. The average visual inspection accuracy of 55% is used as a baseline for comparison with the proposed model in Section 4.5.

Cohen's Kappa is used to assess inter-rater reliability beyond chance agreement. Moderate to substantial agreement is found for Extra class ($\kappa \approx 0.595$) and Decayed ($\kappa \approx 0.615$), indicating consistent identification of premium and defective categories. In contrast, Class I and Class II show only fair agreement ($\kappa \approx 0.31$–0.34), likely due to subtle visual differences and subjective judgment in intermediate grades. This inconsistency in borderline classes reduces classification reliability and underscores the need for an agentic AI-assisted model to improve objectivity, consistency, and explainability in lime quality assessment.

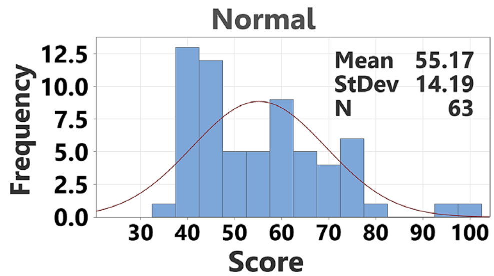

Figure 9: Distribution of visual inspection scores.

4.2 *Model development (Agentic AI-assisted model)*

4.2.1 Development and training of CNN models

This stage corresponds to Stage 1 (Perception) of the agentic AI-assisted model, in which CNNs analyze lime images and classify them into quality classes before subsequent reasoning, planning and quality assessment processes.

Multiple CNN architectures, including VGG19, ResNet50, and AlexNet, are trained and evaluated using accuracy, precision, recall, F1-score, and confusion matrices to identify the most appropriate model for lime quality classification. As shown in Table 2 and Figure 10. VGG19 can achieve the highest performance with approximately 94% accuracy, followed by ResNet50 (90%) and AlexNet (82%). The confusion matrices further indicate that VGG19 provides the most accurate and consistent class-wise predictions with fewer misclassifications, particularly between visually similar categories such as Class I and Class II, as illustrated in Figure 11.

Table 2: Performance evaluation results of CNN models.

Model	Accuracy	Precision	Recall	F1-score
VGG19	94%	0.938	0.938	0.938
ResNet50	90%	0.899	0.898	0.899
AlexNet	82%	0.817	0.816	0.816

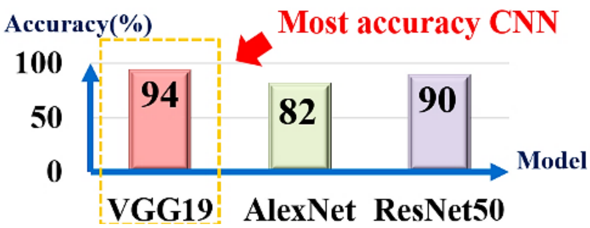

Figure 10: Accuracy comparison of CNN models.

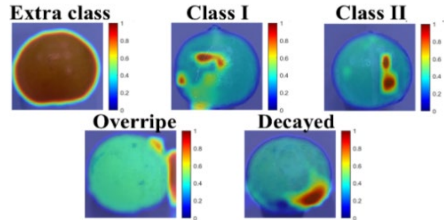

Figure 12: Grad-CAM results of VGG19 for lime quality classification.

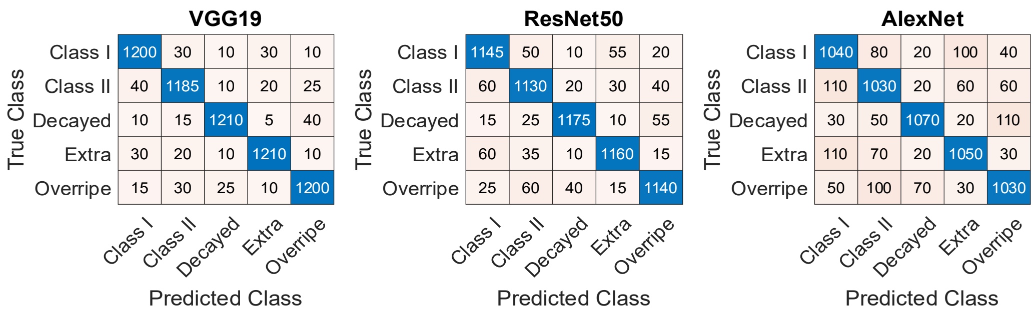

Figure 11: Confusion matrices of VGG19, ResNet50, and AlexNet for lime quality classification.

Therefore, VGG19 is selected as the primary CNN model due to its superior classification performance and effective feature extraction. Similarly, Kumar and Chinnasamy [12] demonstrated that CNN-based models are effective for tomato leaf disease classification, supporting the use of deep learning for agricultural image analysis. ResNet50 also demonstrates strong performance through residual learning, although with higher computational complexity, while AlexNet offers lower complexity but reduced classification accuracy.

Grad-CAM visualization is applied to the VGG19 model, as shown in Figure 12, to improve model transparency and interpretability by identifying image regions that contribute to classification decisions [29]. The results indicate that the model focuses on peel color, ripeness characteristics, and defect regions relevant to Codex-based lime quality grading, consistent with recent explainable CNN studies in agriculture [6].

Within the proposed agentic AI-assisted model, CNN-based image classification serves as Stage 1 (Perception) of the proposed model. The results indicate that more accurate CNN models, particularly VGG19, provide reliable classification outputs that support the subsequent reasoning and planning processes (Stage 2) and quality assessment procedures in later stages of the workflow.

4.2.2 Development of the regression model

The developed regression models show that both MLR and OLR effectively explain the relationship between lime attributes and quality levels, with all variables being statistically significant (p-value < 0.001). Consistent with Section 3.3.2, the agentic AI-assisted model utilizes regression analysis throughout Stages 2 to 4 to support Codex-based lime quality classification and assessment. For the MLR model obtained from Stage 4, the regression equation is expressed as follows.

$$Y = 6.986 - 0.02417(x_1) + 0.01659(x_2) + 0.02079(x_3) - 0.03486(x_4) - 0.005933(x_5)$$

where x_1, x_2, x_3, x_4, and x_5 represent brightness, green value, yellow value, firmness, and peel defect, respectively. The model demonstrates high predictive performance, with an R^2 value of 92.90%, indicating strong explanatory power for continuous quality variation.

The coefficients indicate that brightness and firmness negatively influence the predicted quality score, whereas green and yellow values contribute positively. Substituting representative sample values (x_1= 45, x_2 = -25, x_3 = 28, x_4 = 130, x_5 = 0) yields Y = 1.53, corresponding to the

Extra class. However, because MLR produces continuous outputs, additional thresholding is required for discrete classification, which can introduce ambiguity in borderline cases.

The residual analysis in Figure 13 suggests that the MLR model reasonably satisfies the normality assumption, as indicated by the Normal Probability Plot (a) and Histogram (c). However, the Versus Fits plot (b) and Versus Order plot (d) suggest possible non-linearity and potential violations of the independence assumption, indicating that the model can not fully capture the underlying relationships in the data.

The OLR model obtained from Stage 4 showed a statistically significant fit (p-value < 0.05) with strong ordered associations between the response variable and predicted probabilities (Somers' D = 0.91, Gamma = 0.91, and Tau-a = 0.73). These results confirm a strong relationship between lime attributes and quality grades. In the agentic AI-assisted model, the OLR model supports quality assessment in Stage 4, while its outputs are interpreted by the LLM Explanation Agent during output generation in Stage 6, enabling structured classification aligned with Codex standards, particularly in borderline cases where visual inspection produces inconsistent judgments. The OLR model is expressed through the following cumulative logit equations.

$$\text{Logit}[P(Y \le 1)] = -43.48 + 0.31(x_1) - 0.76(x_2) - 0.16(x_3) + 0.26(x_4) + 0.22(x_5)$$

$$\text{Logit}[P(Y \le 2)] = -67.76 + 0.31(x_1) - 0.76(x_2) - 0.16(x_3) + 0.26(x_4) + 0.22(x_5)$$

$$\text{Logit}[P(Y \le 3)] = -73.66 + 0.31(x_1) - 0.76(x_2) - 0.16(x_3) + 0.26(x_4) + 0.22(x_5) \quad (4)$$

$$\text{Logit}[P(Y \le 4)] = -76.42 + 0.31(x_1) - 0.76(x_2) - 0.16(x_3) + 0.26(x_4) + 0.22(x_5)$$

where x_1–x_5 represent peel brightness, green value, yellow value, firmness, and peel defect size. For example, substituting the values for an Extra class lime ($x_1 = 45$, $x_2 = -25$, $x_3 = 28$, $x_4 = 130$, $x_5 = 0$) into the OLR cumulative logit equations yields a very large positive logit for the $Y \le 1$ equation, as shown in the following.

$$\eta = +0.31(45) - 0.76(-25) - 0.16(28) + 0.26(130) + 0.22(0) = 62.27$$

Then, $\text{Logit}[P(Y \le 1)] = -43.48 + 62.27 = 18.79$

Therefore, $P(Y \le 1) = \frac{e^{18.79}}{1+e^{18.79}} \approx 1.00$.

Table 3: Class-wise performance evaluation of the OLR model for lime quality classification.

Class	Quality Grade	Precision	Recall	F1-score
1	Extra Class	0.911	0.926	0.919
2	Class I	0.757	0.757	0.757
3	Class II	0.783	0.773	0.778
4	Overripe	1.000	1.000	1.000
5	Decayed	1.000	0.997	0.998

This result indicates that the sample is classified into the Extra class with near certainty. Such high-confidence predictions reflect the strong separability of quality classes in the feature space.

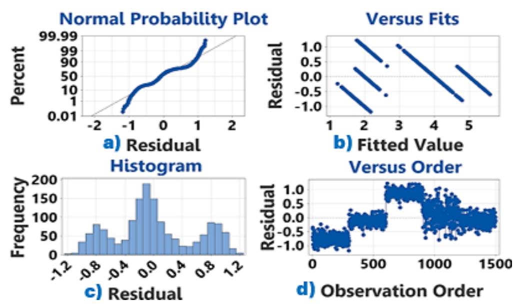

Figure 13: Residual diagnostic plots for the MLR model.

The distribution analysis in Figure 14 shows separation between higher-quality classes (Extra Class, Class I, and Class II) and lower-quality classes (Overripe and Decayed), particularly in firmness and brightness, reflecting a clear ordinal structure with limited overlap among classes. Consistently, Table 3 demonstrates that the OLR model also achieves high overall classification accuracy with strong class-wise performance based upon confusion matrices, precision, recall, and F1-score, confirming its suitability for ordered quality classification. Overripe and Decayed achieves near-perfect performance values close to 1.00, while Extra Class also shows high performance with an F1-score of 0.919. Only slight

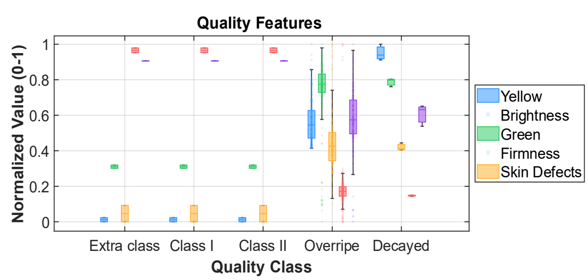

Figure 14: Distribution of lime quality features across quality classes.

misclassification is observed between neighboring grades such as Class I and Class II, reflecting their similar visual characteristics. Overall, both MLR and OLR demonstrate strong predictive performance; however, OLR is more appropriate for ordered and overlapping grade structures. Similar findings have been reported in previous apple grading studies using ordinal regression [19]. Within the agentic AI-assisted model, OLR improves decision consistency, reduces ambiguity in borderline cases, and supports structured Codex-compliant lime quality classification.

4.3 *Verification and validation*

For Stage 1 (Perception), CNN models including VGG19, AlexNet, and ResNet50 achieve accuracy above 80%, confirming reliable perception. For Stage 5 (Learning and Self-Correction), evaluation results from the MLR and OLR models are used to validate performance and support quality assessment refinement. MLR demonstrates strong predictive accuracy with high R^2 and low RMSE, while OLR shows statistically significant predictors and strong ordinal association measures, confirming reliable classification of ordered quality grades.

The validation results improve classification accuracy, reliability, and model generalization, supporting the Learning and Self-Correction process [27].

4.4 *Evaluation*

The evaluation is conducted using a test dataset comprising 20% of the total data to determine the most appropriate models for lime quality classification within the proposed agentic AI-assisted model.

For the perception component, VGG19, AlexNet, and ResNet50 are evaluated based on classification accuracy. VGG19 achieves the highest accuracy (94%), outperforming the other architectures; therefore, it is selected as the perception model, consistent with the findings of Kumar and Chinnasamy [12]. In Stage 5 (Learning and Self-Correction), MLR and OLR are evaluated to determine the most appropriate classification model. Although MLR shows strong predictive performance (R^2 = 92.90%), it requires thresholding for discrete classification. In contrast, OLR demonstrates strong ordinal consistency (Somers' D = 0.91; Kendall's Tau-a = 0.70) while directly producing probability-based outputs for ordered quality levels. Therefore, OLR is selected as the regression model, consistent with previous fruit grading studies using ordinal regression [19].

4.5 *Integrated model evaluation*

The developed classification model integrates the selected VGG19 model with the OLR model. The VGG19 model autonomously extracts image-based features for preliminary lime classification, while measured lime attributes, including peel color, brightness and defect size, are used as inputs for the OLR model to generate probability values for each quality level. This architecture supports Stage 6: Explainable Output Generation of the agentic AI-assisted model process. The model provides explainable outputs through the predicted quality class, measured lime attributes, and OLR probability values interpreted by the LLM Explanation Agent. These outputs enhance transparency and reliability, enabling systematic verification and interpretation of classification results based upon image-derived and numerical information.

The developed conceptual model is evaluated using the test dataset for lime quality classification. Compared with conventional visual inspection, the proposed agentic AI-assisted model provides more consistent and objective results, achieving 90% classification accuracy compared with 55% for visual inspection by 60 expert lime evaluators. While Kumar and Chinnasamy [12] reported 99% accuracy for tomato leaf disease classification using CNN models, the presented study demonstrates that AI can also achieve reliable Codex-based lime quality grading.

5. Conclusions

The objective of this research is to develop an agentic AI-assisted model for Codex-based lime quality classification by integrating CNN-based perception, OLR-based quality assessment, and LLM-driven reasoning and explanation mechanisms. The literature review identifies a research gap, as there is limited evidence of applying agentic AI-assisted model concepts to support quality assessment in Codex-based fruit classification systems. The results demonstrate that the proposed model achieves high classification performance, with VGG19 attaining 94% accuracy and significantly outperforming visual inspection, which achieves 55% accuracy. The contribution of this research is the development of a Codex-aligned lime grading model integrating CNN, OLR, and LLM-driven agentic AI-assisted model for automated and interpretable quality assessment. Theoretically, this research contributes to the development of an agentic AI-assisted model by demonstrating the importance of aligning data structure, model selection, and grading standards, while showing that OLR is more appropriate than MLR for ordered and overlapping Codex quality grades. Practically, this research demonstrates the potential of an agentic AI-assisted model to support objective and reliable lime quality grading in real-world agricultural practice. The advantage lies in delivering explainable quality assessments through an agentic AI-assisted model. The benefit of the proposed model is improving accuracy, reliability, and explainability in quality grading while reducing missed sales opportunities for farmers. A limitation of this research is that the model is developed using lime images under controlled conditions, limiting its applicability to other fruits and environments. Future research can be extended through external validation of the proposed model, together with the integration of aroma, taste, and juice content using multimodal sensing and 3D fruit analysis to enhance quality assessment and support LLM-driven reasoning.

Acknowledgments

The authors would like to express their sincere gratitude to the manufacturers and all expert evaluators mentioned in this research for their valuable information, advice, and comments. Additionally, they extend their heartfelt thanks to the Faculty of Engineering at King Mongkut's University of Technology North Bangkok for their unwavering support in the completion of this research.

Author Contributions

A.K.: provided the project’s conceptualization, methodology, and research design; conducted data curation and data analysis.; C.K.: contributed to the research design; conducted the investigation and data curation. Both authors were also responsible for writing the original draft and performing all subsequent reviews and editing. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors utilized the ChatGPT tool to enhance the language and readability of the manuscript.

References

[1] M. Abou Ali, F. Dornaika, and J. Charafeddine, “Agentic AI: A comprehensive survey of architectures, applications, and future directions,” *Artificial Intelligence Review*, vol. 59, Art. no. 11, 2026, doi: 10.1007/s10462-025-11422-4.

[2] J. E. See, “Visual inspection reliability for precision manufactured parts,” *Human Factors*, vol. 57, no. 8, pp. 1427–1442, 2015, doi: 10.1177/0018720815602389.

[3] PictureThis, “Key lime citrus,” 2026. [Online]. Available: https://www.picturethisai.com/wiki/Citrus_aurantiifolia.html. Accessed: Jan. 26, 2026.

[4] World Integrated Trade Solution (WITS), “Lemons and limes, fresh or dried exports by country 2024,” 2024. [Online]. Available: https://wits.worldbank.org. Accessed: Jan. 14, 2026.

[5] N. Parashar, P. Johri, A. A. Khan, N. Gaur, and S. Kadry, “An integrated analysis of yield prediction models: A comprehensive review of advancements and challenges,” *Computers, Materials & Continua*, vol. 80, no. 1, pp. 389–425, 2024, doi: 10.32604/cmc.2024.050240.

[6] V. K. Vishnoi, K. Kumar, B. Kumar, S. Mohan, and A. A. Khan, “Detection of apple plant diseases using leaf images through convolutional neural network,” *IEEE Access*, vol. 11, pp. 6594-6609, 2023, doi: 10.1109/ACCESS.2022.3232917.

[7] A. A. Khan, M. Faheem, R. N. Bashir, C. Wechtaisong, and M. Z. Abbas, “Internet of Things (IoT) assisted context aware fertilizer recommendation,” *IEEE Access*, vol. 10, pp. 129505–129520, 2022, doi: 10.1109/ACCESS.2022.3228160.

[8] A. A. Khan, M. A. Nauman, R. N. Bashir, R. Jahangir, R. Alroobaea, A. Binmahfoudh, M. Alsafyani, and C. Wechtaisong, “Context aware evapotranspiration (ETs) for saline soils reclamation,” *IEEE Access*, vol. 10, pp. 110050–110064, 2022, doi: 10.1109/A.CCESS.2022.3206009.

[9] R. N. Bashir, F. A. Khan, A. A. Khan, M. Tausif, M. Z. Abbas, M. M. A. Shahid, and N. Khan, “Intelligent optimization of reference evapotranspiration (ETo) for precision irrigation,” *Journal of Computational Science*, vol. 69, Art. no. 102025, 2023, doi: 10.1016/j.jocs.2023.102025.

[10] S. C. Ramu, D. Raman, K. Ramana, A. V. Krishna, A. A. Khan, S. Khan, and S. M. Abdu, “A lightweight deep learning and whale

optimization framework for sustainable precision agriculture," *Discover Computing*, vol. 29, Art. no. 73, 2026, doi: 10.1007/s10791-026-09952-8.
[11] K. Vijaipriya, N. Moses, and P. A. Michael, "Efficient multi-task learning in multi-user multiple input multiple output systems integrated orthogonal frequency division multiplexing systems: A hybrid amalgamated convolutional neural network–bidirectional long short-term memory approach," *Applied Science and Engineering Progress*, vol. 19, no. 2, Art. no. 7942, 2026, doi: 10.14416/j.asep.2025.11.005.
[12] S. R. Kumar and M. P. Chinnasamy, "AI-Driven Detection of Tomato Leaf Diseases for Sustainable Agriculture," *Applied Science and Engineering Progress*, vol. 18, no. 4, Art. no. 7825, 2025, doi: 10.14416/j.asep.2025.06.007.
[13] D. B. Acharya, K. Kuppan, and B. Divya, "Agentic AI: Autonomous intelligence for complex goals—A comprehensive survey," *IEEE Access*, vol. 13, pp. 18912–18936, 2025, doi: 10.1109/ACCESS.2025.3532853.
[14] J. Lee and H. Su, "Agentic AI for smart manufacturing," *Manufacturing Letters*, vol. 46, pp. 92–96, 2025, doi: 10.1016/j.mfglet.2025.10.013.
[15] A. Loaiza-Bonilla, C. Yost, S. Kurnaz, E. Tuysuz, N. G. Thaker, D. Giritlioglu, and J. P. Noel Meza, "Transforming oncology clinical trial matching through neuro-symbolic, multi-agent AI and an oncology-specific knowledge graph: A prospective evaluation in 3804 patients," *ESMO Real World Data and Digital Oncology*, vol. 12, Art. no. 100706, 2026, doi: 10.1016/j.esmorw.2026.100706.
[16] FAS-USDA, "Production: Lemons & Limes." [Online]. Available: https://www.fas.usda.gov/data/production/commodity/0572120. Accessed: Feb. 18, 2026.
[17] Codex Alimentarius Commission, *Standard for Mexican Limes (CXS 217-1999)*, Rome, Italy, 1999.
[18] A. M. Torkashvand, A. Ahmadi, and N. L. Nikravesh, "Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network (ANN) and multiple linear regressions (MLR)," *Journal of Integrative Agriculture*, vol. 16, no. 7, pp. 1634–1644, 2017, doi: 10.1016/S2095-3119(16)61546-0.
[19] X. Qu, L. Ma, T. Shen, X. Han, X. Yan, and S. Bi, "Ordinal regression-extreme learning machine based apple grading method," in *Proceedings of the 2022 China Automation Congress (CAC)*, 2022, pp. 3507–3511, doi: 10.1109/CAC57257.2022.10054740.
[20] Q. Dong, X. Chen, and B. Huang, *Data Analysis in Pavement Engineering: Methodologies and Applications*. Cambridge, U.K.: Woodhead Publishing, 2023.
[21] P. A. Gutiérrez, M. Pérez-Ortiz, J. Sánchez-Monedero, F. Fernández-Navarro, and C. Hervás-Martínez, "Ordinal regression methods: Survey and experimental study," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 1, pp. 127–146, 2016, doi: 10.1109/TKDE.2015.2457911.
[22] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960, doi: 10.1177/001316446002000104.
[23] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012, doi: 10.11613/BM.2012.031.
[24] O. Tasneem and R. Pieters, "Human–robot collaborative visual inspection with Large Language Models," *Robotics and Computer-Integrated Manufacturing*, vol. 98, Art. no. 103154, 2026, doi: 10.1016/j.rcim.2025.103154.
[25] A. Corallo, V. Del Vecchio, A. Di Prizio, and M. Buscicchio, "Benchmarking AI-based visual inspection systems for aerospace quality control: A multi-vendor comparative evaluation," *Procedia Computer Science*, vol. 277, pp. 1621–1630, 2026, doi: 10.1016/j.procs.2026.02.200.
[26] L. Ierimonti, F. Mariani, F. Ubertini, and I. Venanzi, "A dynamic Bayesian networks-based approach for data fusion of model-based SHM and visual inspections in post-tensioned bridges," *Structures*, vol. 86, Art. no. 111389, 2026, doi: 10.1016/j.istruc.2026.111389.
[27] M. A. Saputra and I. Nurhaida, "Signature originality verification using a deep learning approach," *Electronic Journal of Education, Social Economics and Technology*, vol. 5, no. 1, pp. 19–29, 2024, doi: 10.33122/ejeset.v5i1.310.
[28] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, Art. no. 60, 2019, doi: 10.1186/s40537-019-0197-0.
[29] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.