

## Computer vision algorithms for characterizing and classifying agro-exportable grains through size analysis and clustering techniques in Peru's agricultural sector

Algoritmos de visão computacional para a caracterização e classificação de grãos agroexportáveis por meio de análise de tamanho e técnicas de agrupamento no setor produtivo do Peru

Algoritmos de visión computacional para la caracterización y clasificación de granos agroexportables mediante análisis de tamaño y técnicas de agrupamiento en el sector produtivo del Perú

**Betzabeth Abigail Arias Rivas**

<https://orcid.org/0009-0002-9434-4842>  
Universidad Nacional de Cañete UNDC, Peru

**Julio Cesar Machaca Mamani**

<https://orcid.org/0000-0003-4866-3885>  
Universidad Nacional de Cañete UNDC, Peru  
[jmachaca@undc.edu.pe](mailto:jmachaca@undc.edu.pe) (correspondence)

**Rocio Cahuana Lipa**

<https://orcid.org/0000-0002-7671-5585>  
Universidad Nacional de Cañete UNDC, Peru

**Manuel Alberto Luis Manrique Nugent**

<https://orcid.org/0000-0002-0816-2499>  
Universidad Nacional de Cañete UNDC, Peru

### ABSTRACT

**Introduction:** The implementation of machine learning and deep learning algorithms has revolutionized quality control processes in the agroindustry, providing precise and efficient solutions for grain classification and evaluation. The use of computer vision, combined with advanced algorithms, enables defect detection, measurement of physical characteristics, and optimization of batch standardization for both domestic and international markets. **Objective:** To review and analyze the application of ML algorithms in grain classification through digital image processing, comparing their accuracy, efficiency, and processing time to identify the most suitable strategy for productive environments. **Materials and Methods:** A systematic review was conducted in Scopus, IEEE Xplore, ScienceDirect, and MDPI, considering publications from 2017 to 2024. Inclusion criteria were: use of RGB or hyperspectral images, application of CNN, SVM, K-means, or Random Forest algorithms, and reporting of quantitative metrics. Review articles, studies not applied to grains, or those without experimental validation were excluded. **Results and Conclusions:** CNN achieved the highest accuracy ( $\approx 97\%$ ), making it ideal for detailed classification, although it requires high computational resources and longer training times. SVM and Random Forest demonstrated a balance between accuracy ( $\approx 91\text{--}92\%$ ) and efficiency, making them suitable for medium-sized datasets. K-means stood out for its speed and simplicity, although with lower accuracy ( $\approx 88\%$ ) as it is an unsupervised method. Emerging trends point to the integration of hyperspectral vision, transfer learning, and hybrid approaches to optimize the balance between accuracy, speed, and operational feasibility, thereby enhancing the competitiveness of the agroindustry in global markets.

**Keywords:** Algorithms, Agroindustry, Classification, Computer Vision, Hyperspectral Vision.

### ARTICLE HISTORY

**Received:** 11-04-2025

**Revised Version:** 19-09-2025

**Accepted:** 25-11-2025

**Published:** 25-12-2025

**Copyright:** © 2025 by the authors

**License:** CC BY-NC-ND 4.0

**Manuscript type:** Article

### ARTICLE INFORMATION

#### Science-Metrix Classification (Domain):

Applied Sciences

#### Main topic:

AI and agriculture

#### Main practical implications:

The findings guide agroindustry decision-making by matching algorithms to operational constraints, enabling faster, more accurate grain quality control, reduced inspection costs, improved batch standardization, and enhanced compliance with market requirements.

#### Originality/value:

This review synthesizes comparative evidence on multiple ML approaches for grain classification, contrasting accuracy, efficiency, and computational demands, while highlighting emerging trends such as hyperspectral imaging and hybrid models.

### RESUMO

**Introdução:** A implementação de algoritmos de aprendizado automático (machine learning e deep learning) revolucionou os processos de controle de qualidade na agroindústria, oferecendo soluções precisas e eficientes para a classificação e avaliação de grãos. O uso da visão computacional, combinado com algoritmos avançados, permite detectar defeitos, medir características físicas e otimizar a padronização de lotes destinados aos mercados nacionais e internacionais. **Objetivo:** Revisar e analisar a aplicação de algoritmos de ML na classificação de grãos por meio do processamento de imagens digitais, comparando sua precisão, eficiência e tempo de processamento para identificar a estratégia mais adequada em ambientes produtivos. **Métodos:** Foi realizada uma revisão sistemática no Scopus, IEEE Xplore, ScienceDirect e MDPI, considerando publicações entre 2017 e 2024. Os critérios de inclusão foram: uso de imagens RGB ou hiperespectrais, aplicação de algoritmos CNN, SVM, K-means ou Random Forest e relatório de métricas quantitativas. Foram excluídos artigos de revisão, trabalhos não aplicados a grãos ou sem validação experimental. **Resultados e conclusões:** A CNN apresentou a maior precisão ( $\approx 97\%$ ), sendo ideal para classificação detalhada, embora com altos requisitos computacionais e tempos prolongados de treinamento. SVM e Random Forest mostraram um equilíbrio entre precisão ( $\approx 91\text{--}92\%$ ) e eficiência, adequando-se a conjuntos de dados médios. K-means se destacou por sua rapidez e simplicidade, embora com menor precisão ( $\approx 88\%$ ) por não ser supervisionado. As tendências emergentes apontam para a integração de visão hiperespectral, transfer learning e abordagens híbridas, com o objetivo de otimizar o equilíbrio entre exatidão, velocidade e viabilidade operacional, fortalecendo a competitividade da agroindústria nos mercados globais.

**Palavras-chave:** Algoritmos, Agroindústria, Classificação, Visão Artificial, Visão Hiperespectral.

### RESUMEN

**Introducción:** La implementación de algoritmos de aprendizaje automático (machine learning y deep learning) ha revolucionado los procesos de control de calidad en la agroindustria, ofreciendo soluciones precisas y eficientes para la clasificación y evaluación de granos. El uso de visión por computador, combinado con algoritmos avanzados, permite detectar defectos, medir características físicas y optimizar la estandarización de lotes destinados a mercados nacionales e internacionales. **Objetivo:** Revisar y analizar la aplicación de algoritmos de ML en la clasificación de granos mediante procesamiento de imágenes digitales, comparando su precisión, eficiencia y tiempo de procesamiento para identificar la estrategia más adecuada en entornos productivos. **Material y métodos:** Se realizó una revisión sistemática en Scopus, IEEE Xplore, ScienceDirect y MDPI, considerando publicaciones entre 2017 y 2024. Los criterios de inclusión fueron: uso de imágenes RGB o hiperespectrales, aplicación de algoritmos CNN, SVM, K-means o Random Forest, y reporte de métricas cuantitativas. Se excluyeron artículos de revisión, trabajos no aplicados a granos o sin validación experimental. **Resultados y conclusiones:** CNN presentó la mayor precisión ( $\approx 97\%$ ), siendo ideal para clasificación detallada, aunque con altos requerimientos computacionales y tiempos prolongados de entrenamiento. SVM y Random Forest mostraron un balance entre precisión ( $\approx 91\text{--}92\%$ ) y eficiencia, adecuándose a datasets medianos. K-means destacó por su rapidez y simplicidad, aunque con menor precisión ( $\approx 88\%$ ) al no ser supervisado. Las tendencias emergentes apuntan a la integración de visión hiperespectral, transfer learning y enfoques híbridos, con el fin de optimizar el equilibrio entre exactitud, velocidad y viabilidad operativa, fortaleciendo la competitividad de la agroindustria en mercados globales.

**Palabras clave:** Algoritmos, Agroindustria, Clasificación, Visión Artificial, Visión Hiperespectral.

## INTRODUCTION

Peru has established itself as one of the world's leading producers and exporters of high-quality grains, with products such as *Coffea arabica* (coffee), *Theobroma cacao* (cocoa), *Chenopodium quinoa* (quinoa), *Amaranthus caudatus* (kiwicha), *Zea mays everta* (giant corn from Cusco), *Phaseolus vulgaris* (beans), *Glycine max* (soybeans), and other Andean grains with growing demand in European, Asian, and North American markets (García et al., 2021; López & Jimenez, n.d.; Shobana et al., 2022). These products not only represent a significant source of income for the agro-industrial sector, but also reinforce the country's agricultural identity, with exports exceeding hundreds of thousands of tons annually, with coffee, cocoa, and quinoa being the most internationally positioned (Capajaña & Capajaña, 2020; María et al., 2024). The visual quality of these beans, especially their size, uniformity, and absence of impurities and defects, largely determines their acceptance in premium markets and their classification within the phytosanitary and commercial regulations required by international buyers (Pastor et al., 2022; Paz et al., 2018; Shobana et al., 2022). However, the traditional grain classification process in Peru is largely based on manual inspection, a method that, although economical, has significant limitations: high subjectivity, variability between operators, low productivity, and difficulties in standardizing large-volume batches (Swathi et al., 2025; Tello et al., 2024). This problem is exacerbated by the rapid growth of agricultural exports, which requires more reliable, faster, and scalable quality control systems. Given this scenario, the use of computer vision algorithms is positioned as a key tool for modernizing the productive sector, allowing for automated and objective analysis of attributes such as size and shape, as well as the application of clustering techniques to segment grains into categories of quality, caliber, or export grade (Alshahrani et al., 2023; King et al., 2022).

Computer vision systems leverage digital image processing to identify morphometric characteristics (e.g., area, perimeter, aspect ratio, equivalent diameter) with high precision, using operations such as thresholding, edge detection, and morphological operations. The data is then processed using unsupervised clustering algorithms such as k-means, DBSCAN, or hierarchical clustering, which classify the grains into homogeneous groups without direct human intervention (Cinar & Koklu, 2022; Reza et al., 2019; Saxena et al., n.d.). This technological integration reduces selection errors, optimizes production flow, and ensures standardized traceability. In addition, the incorporation of machine learning and deep learning models, especially convolutional neural networks (CNNs), has improved accuracy in complex tasks such as the detection of surface defects or deformations, achieving greater reproducibility in the face of variations in color and texture typical of grains grown in heterogeneous conditions (Adrah et al., 2021; de C. Motta et al., 2024; Dhakal et al., 2023; Ibrahim et al., 2019; Naik et al., 2024). These technologies allow size analysis to be combined with advanced clustering metrics to classify, in real time, large volumes of coffee, cocoa, quinoa, or giant corn from Cusco, ensuring consistency in the post-harvest process (Chen et al., 2024; Verma et al., 2023).

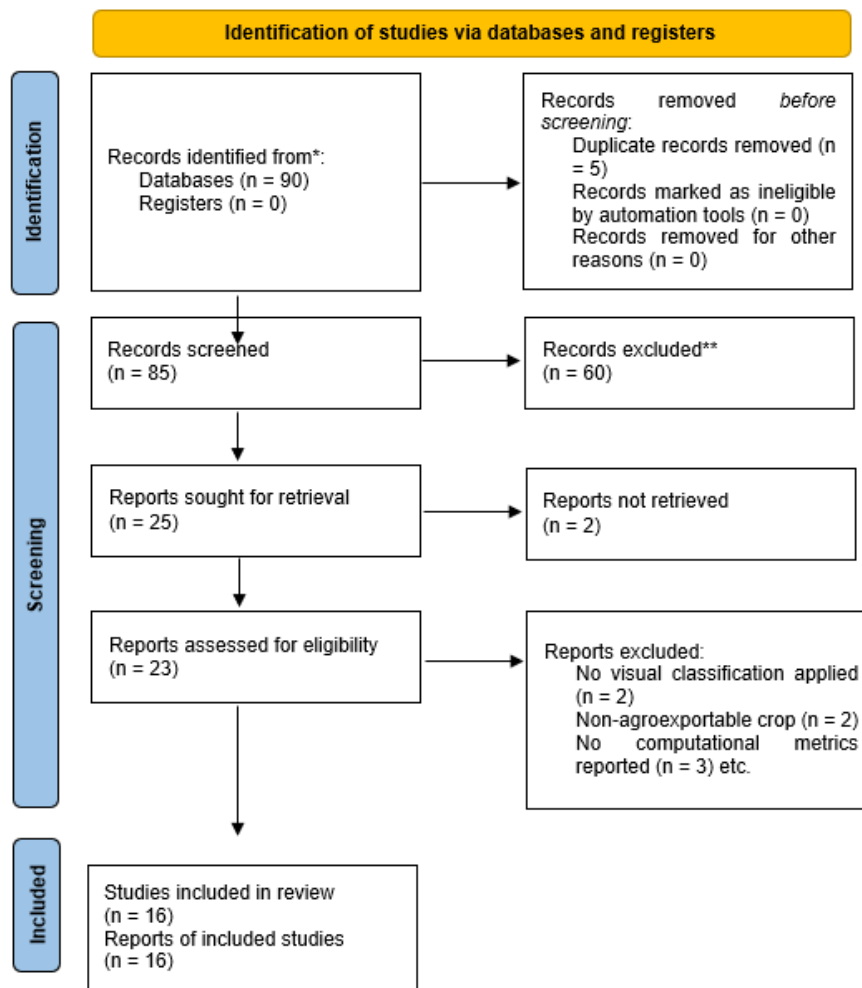
However, the current challenge in Peru lies in transferring these technologies to the real production environment, as many solutions reported in the scientific literature have been designed under laboratory conditions, with little adaptation to the variability of local crops, limited infrastructure, and the need for low operating costs (Pastor et al., 2022). This systematic review seeks to analyze, synthesize, and critically evaluate computational algorithms based on size and clustering techniques applied to agro-exportable grains, with the aim of identifying robust, scalable methodologies that are adaptable to the Peruvian agro-industrial context (Ramdhani & Alamsyah, 2023; Saxena et al., n.d.). It also addresses emerging trends, such as hyperspectral imaging, transfer learning, and integration with Industry 4.0 technologies, which are transforming the agro-export value chain in developing countries.

This review aims to critically analyze and compare the most widely used computer vision algorithms in the visual classification of agri-food grains for export, with an emphasis on their applicability, accuracy, computational efficiency, and feasibility of implementation in the Peruvian agro-industrial context. It seeks to answer the question: Which algorithms offer an optimal balance between accuracy and operating cost for implementation in automatic classification systems in post-harvest plants in Peru?

## METHODS

A systematic review of the literature was conducted between 2017 and 2024 in the IEEE Xplore, Scopus, ScienceDirect, and MDPI databases. The following keywords were used: 'grain classification', 'computer vision', 'deep learning', 'CNN', 'SVM', 'K-means', 'agroexport'. The inclusion criteria were: studies with reported accuracy metrics, application to agricultural products, and use of computational algorithms. Articles without full access, without experimental validation, or that focused on non-exportable crops were excluded. Fifteen articles were selected after applying the PRISMA search and selection flow.

Figure 1. PRISMA diagram



Source: authors' development

The studies selected for this review met clearly defined criteria to ensure the relevance and quality of the evidence analyzed. First, only articles published between 2017 and 2024 were considered, in order to ensure the technological relevance of the algorithms evaluated. Priority was given to research applied to agro-export crops of high relevance to the Peruvian context, such as coffee, cocoa, quinoa, rice, and corn. In terms of computational approaches, studies were selected that implemented representative and widely used algorithms in computer vision, including convolutional neural networks (CNN), support vector machines (SVM), unsupervised clustering techniques such as K-means, and ensemble methods such as Random Forest. Studies were also required to use RGB images as a data source, either from public databases or generated in real or laboratory environments. Finally, only those studies that reported key quantitative metrics, such as accuracy, F1-score, recall, and computational efficiency, were included, thus allowing for a rigorous comparison of the performance of each algorithm.

## RESULTS AND DISCUSSION

### Critical comparison between algorithms

Although CNNs achieve accuracies greater than 97%, they require high-performance equipment and lengthy training, which limits their adoption in rural areas without access to GPUs. On the other hand, algorithms such as K-means allow for efficient grouping by size ( $\approx 88\%$ ) with minimal computational load, although they are sensitive to batch variability. SVM represents an interesting balance, achieving up to 92% accuracy in contexts with medium-sized datasets. Random Forest offers robustness in the presence of noise and unbalanced data, ideal for variable environments such as those found in the highlands.

**Table 1.** Comparison of computer vision algorithms applied to the classification and quality control of Andean grains

Algorithm	Accuracy	Processing	Main application	Studies
<b>CNN (Deep Learning)</b>	90–99	Medium–High	Classification by size and defects; detection of foreign bodies in coffee, cocoa, and quinoa	<ul style="list-style-type: none"> <li>• (Yasar, 2024)</li> <li>• (El Akrouchi et al., 2025a)</li> <li>• (Avila-George et al., 2023)</li> </ul>
<b>SVM (Machine Learning)</b>	90–96	Low–Medium	Size separation, classification of rice, bean, and cereal varieties	<ul style="list-style-type: none"> <li>• (Naik et al., 2024)</li> <li>• (Chen et al., 2024)</li> <li>• (Ibrahim et al., 2019)</li> <li>• (Ardeshirifar, 2024)</li> </ul>
<b>K-means (Clustering)</b>	85–90	Low	Grouping by size, exploratory segmentation of batches	<ul style="list-style-type: none"> <li>• (Reza et al., 2019)</li> <li>• (Cinar &amp; Koklu, 2022; Zhang et al., 2023)</li> </ul>
<b>Random Forest (Ensemble)</b>	88 – 98	Medium	Overall classification, high robustness in the presence of noise and imbalances	<ul style="list-style-type: none"> <li>• Cinar &amp; Koklu (2022)</li> </ul>

**Note:** Data based on studies from 2019–2025; the choice of algorithm depends on the balance between accuracy, resources, and application

Comparison of algorithms shows variable performance in accuracy, processing complexity, and agro-industrial applications. Convolutional neural networks (CNN) stand out with 90% to 99% accuracy, making them ideal for complex tasks such as detecting defects and foreign bodies in coffee, cocoa, and quinoa, although they require medium to high computational resources (Yasar, 2024; El Akrouchi et al., 2025; Avila-George et al., 2023). The SVM algorithm, with an accuracy range between 90% and 96%, is efficient for classification by size and variety in rice, beans, and cereals, requiring less computing power (Naik et al., 2024; Chen et al., 2024; Ibrahim et al., 2019; Ardeshirifar, 2024). K-means, with an accuracy between 85% and 90%, is used in segmentation and exploratory clustering, although its unsupervised nature limits its applications in direct classification (Reza et al., 2019; Cinar & Koklu, 2022; Zhang et al., 2023).

Finally, Random Forest offers between 88% and 98% accuracy, being robust to noise and unbalanced data, with average processing (Cinar & Koklu, 2022). Overall, the data suggest that the choice of algorithm should balance accuracy, computational cost, and the nature of the task, prioritizing CNN or Random Forest for high-demand export systems and SVM or K-means for environments with lower technological capacity.

**Table 2.** Main mathematical formulas of algorithms used in computer vision

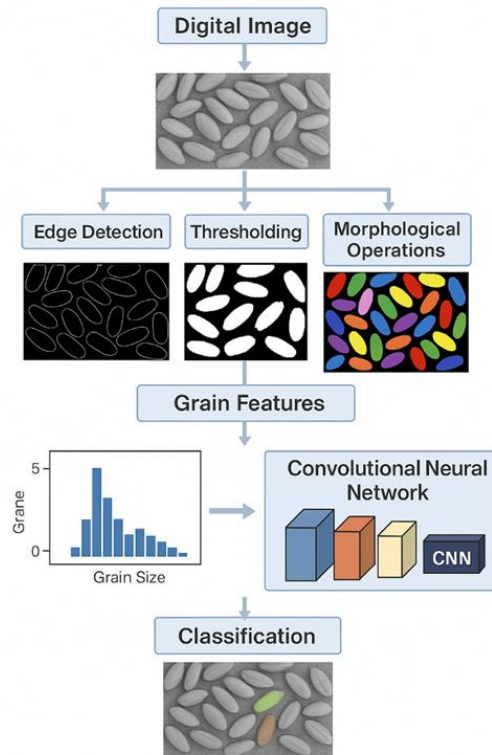
Algorithm	Mathematical Form	Reference
<b>CNN (Convolutional Neural Network)</b>	Operation of convulsion $S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i-m, j-n) K(m, n)$	LeCun et al. (1998)
<b>SVM (Support Vector Machine)</b>	Linear classifier with maximum margin $\min_{\mathbf{w}, b} \frac{1}{2} \ \mathbf{w}\ ^2 \quad \text{s.a.} \quad y_i (\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad \forall i$	Cortes & Vapnik (1995)
<b>K-means</b>	Intracluster inertia objective function $J = \sum_{k=1}^K \sum_{x_i \in C_k} \ x_i - \mu_k\ ^2$	MacQueen (1967)
<b>Random Forest</b>	Prediction by majority vote of T trees $\hat{y} = \text{mode} \{ h_t(x) \}_{t=1}^T$	Breiman (2001)

**Note:** Data based on Breiman, 2001; Cortes & Vapnik, 1995; LeCun et al., 1998; MacQueen, 1967.

A comparative analysis of the mathematical formulas associated with the main computer vision algorithms reveals substantial differences in their complexity and the type of problems they address. The convolution operation in CNNs, proposed by LeCun et al. (1998), forms the basis for the hierarchical extraction of spatial features, which explains its high

performance in object classification and detection tasks, albeit at the cost of high computational requirements. In contrast, the SVM formulation, introduced by Cortes and Vapnik (1995), optimizes a maximum margin hyperplane, offering robust solutions in high-dimensional spaces, especially in contexts with limited data sets. The K-means algorithm, described by MacQueen (1967), is based on minimizing intra-cluster distance, making it an efficient tool for unsupervised segmentation, although its performance depends on the correct choice of the number of clusters. Finally, the majority voting strategy in Random Forest, proposed by Breiman (2001), ensures high generalization capacity and resistance to overfitting by combining multiple weak classifiers. This diversity of mathematical foundations underscores the importance of selecting the most appropriate algorithm based on the type of data, available computational resources, and the specific objectives of the agro-industrial application.

**Figure 2.** Grain processing and classification flow using computer vision and convolutional neural networks (CNN).



**Source:** authors' development based on literature.

In grain classification using computer vision, the most used algorithms are convolutional neural networks (CNN) and their variants such as AlexNet, ResNet, and DenseNet, due to their high accuracy in detecting size, defects, and shape. Classic methods such as Support Vector Machine (SVM) and clustering techniques such as K-means continue to be widely used due to their low computational cost and good performance in environments with limited data (El Akrouchi et al., 2025b; Sharon Ayme Aguirre Matos, n.d.; Shobana et al., 2022). Random Forest and XGBoost are preferred in supervised classification stages when features have already been extracted using morphometric analysis or descriptors such as HOG and SIFT (Breiman, 2001; Busi et al., 2024; Laudari et al., 2022; Taspinar et al., 2022). These techniques have demonstrated efficiencies of over 90% in real applications on grains such as coffee, cocoa, quinoa, and rice, according to recent literature (Choi et al., 2025; Miranda et al., 2024; Salas & Salas, 2022).

### Comparison of efficiency and accuracy

The radar chart (Figure 2) illustrates the comparison between the CNN, SVM, K-means, and Random Forest algorithms, considering three key parameters: accuracy, efficiency, and processing speed. According to Yasar (2024), CNN has the highest accuracy ( $\approx 97\%$ ), making it the preferred option for complex grain classification tasks; however, its speed and efficiency are lower, as it requires greater computing power and longer training times. In contrast, K-means, as reported by (Reza et al., 2019), excels in its high efficiency and speed ( $\approx 95\%$ ), making it ideal for rapid segmentation, although its accuracy is moderate ( $\approx 88\%$ ) as it is not a supervised method. SVM, supported by studies such as (Ibrahim et al., 2019), is positioned as a balanced method, with an accuracy of  $\approx 92\%$  and good efficiency in contexts with medium-sized datasets. Finally, Random Forest, evaluated by (Cinar & Koklu, 2022), offers robust performance with 91% accuracy, being an intermediate option between complexity, speed, and accuracy. This analysis suggests that the choice of algorithm depends on the required balance between prediction quality and available computational resources, with CNN being the most accurate, K-means the fastest, and SVM/Random Forest the most balanced.

**Figure 2.** Comparison of the performance of computer vision algorithms in agribusiness

**Source:** authors' development based on literature.

### Feasibility of Implementation in Peru

The integration of algorithms such as CNN, SVM, K-means, and Random Forest in the Peruvian agro-export sector allows for the automation of quality control and the optimization of grain selection for international markets. These systems guarantee homogeneous batches, free of impurities and accurately classified by size and color, reducing returns and increasing the added value of products such as coffee, cocoa, quinoa, and rice. Studies such as those by (Avila-George et al., 2023) and (Yang et al., 2023) show that, with accuracies close to 98%, computer vision not only improves product traceability but also strengthens Peru's competitiveness in demanding markets such as Europe and Asia (Girma et al., 2021; Salas & Salas, 2022; Sokudlor et al., 2023).

Each computer vision algorithm has inherent limitations that must be evaluated before adoption. Convolutional Neural Networks (CNN) require extensive, labeled datasets, standardized capture conditions, and a computing infrastructure with high processing capacity (GPU) (Choudhary et al., 2022; 张 et al., 2023). The Support Vector Machine (SVM) method has difficulties scaling in contexts with a large number of attributes or massive data volumes ( ), which can slow down processing. The K-means algorithm, being unsupervised, critically depends on the correct choice of the value of K and lacks supervised predictive capabilities, limiting its application in complex classification (Laub et al., 2022; Zhu et al., 2023). Random Forest, on the other hand, offers robustness and accuracy, but can become opaque in its interpretation and require high memory consumption during the inference phase. These limitations must be aligned with the availability of resources, the type of product, and the operating conditions of each agro-industrial plant (Breiman, 2001; Swathi et al., 2025).

### Feasibility of implementation in Peru

The integration of these algorithms into Peruvian agro-industrial plants requires a well-defined minimum infrastructure. This includes machine vision cameras with an optical resolution of at least 5 megapixels, lighting controlled by LED systems to minimize shadows and color variations, and adequate processing units such as industrial computers with GPUs or optimized microcontrollers (NVIDIA Jetson Nano, Raspberry Pi 4 with TPU accelerators) (Bordas, Bordas, et al., 2022; Don et al., 2024; Zerrouki et al., 2024). In addition, it is essential to implement simple and ergonomic user interfaces for operators, ensuring intuitive system operation (Daud et al., 2024; X. Liu et al., 2024). This range of investment can be competitive when considering the increase in efficiency, reduction in waste, and improvement in the quality of batches for export.

## Emerging technological trends

Among the innovations with the greatest potential to transform grain classification in Peru are hyperspectral imaging and transfer learning (from C. Motta et al., 2024; Oscar et al., 2024). Hyperspectral imaging, thanks to its ability to capture information across a wide range of wavelengths, allows for the identification of internal defects, fungal infections, or moisture levels that are imperceptible to the human eye (Asif et al., 2025; Bordas, Zhang, et al., 2022). On the other hand, transfer learning makes it possible to reuse neural networks previously trained on large international databases, adapting them to local conditions and products, which significantly reduces training times and the need to collect extensive datasets (Capajaña & Capajaña, 2020; Y. Liu et al., 2024). The strategic incorporation of these technologies is emerging as a central axis in the consolidation of a smart, competitive, and sustainable agro-export model for Peru.

## FINAL REMARKS

The synthesized evidence shows that there is no "universal algorithm" for visual grain classification; the choice must respond to the analytical objective, data quality, and operational constraints. CNNs offer maximum accuracy for defect detection and fine classification when standardized images and computing power are available; SVM and Random Forest provide a robust balance between accuracy and computational cost in scenarios with medium data and greater variability; K-means is valuable in exploratory stages and for stratification by size, although it does not replace supervised approaches. In the Peruvian context, feasibility depends on integrating controlled capture ( $\geq 5$  MP, LED lighting), a reproducible preprocessing channel, and accessible hardware (lightweight GPU or accelerators), complemented by simple interfaces for operators. Trends such as hyperspectral vision and transfer learning reduce dependence on large datasets and extend detection to non-visible attributes (moisture, internal damage), enabling systems that are more sensitive and adaptable to national agricultural heterogeneity.

Overall, a hybrid strategy is recommended: rapid preprocessing and clustering for triage, followed by supervised classifiers (CNN or RF) when traceability and export quality requirements so dictate. This approach maximizes performance, scalability, and real adoption in the field.

## REFERENCES

- Adrah, K., Adrah, K., Ananey-Obiri, D., Ananey-Obiri, D., Tahergorabi, R., & Tahergorabi, R. (2021). Sweet potato starch and a protein-based edible coating minimize the fat-uptake in deep-fat fried chicken. *Cyta-Journal of Food*. <https://doi.org/10.1080/19476337.2021.1914736>
- Alshahrani, H. M., Saeed, M. K., Alotaibi, S. S., Mohamed, A., Assiri, M., & Ibrahim, S. S. (2023). Quantum-Inspired Moth Flame Optimizer Enhanced Deep Learning for Automated Rice Variety Classification. *IEEE Access*. <https://doi.org/10.1109/access.2023.3330918>
- Ardeshirifar, R. (2024). *Automated Classification of Dry Bean Varieties Using XGBoost and SVM Models*. <http://arxiv.org/abs/2408.01244>
- Asif, M. J., Khan, H., Tehseen, R., Tahir, S., Rizvi, H., Asad, M., Saqib, S., & Ahmad, R. F. (2025). *Exploring Convolutional Neural Networks for Rice Grain Classification: An Explainable AI Approach*. <https://doi.org/10.3390/xxxxx>
- Avila-George, H., De-La-Torre, M., Sánchez-Garcés, J., Quispe, J. J. C., Prieto, J. M., & Castro, W. (2023). Discrimination of foreign bodies in quinoa (*Chenopodium quinoa* Willd.) grains using convolutional neural networks with a transfer learning approach. *PeerJ*, 11. <https://doi.org/10.7717/peerj.14808>
- Bordas, A., Bordas, A., Zhang, J., Zhang, J., Nino, J. C., & Nino, J. C. (2022). Application of Deep Learning Workflow for Autonomous Grain Size Analysis. *Molecules*. <https://doi.org/10.3390/molecules27154826>
- Bordas, A., Zhang, J., & Nino, J. C. (2022). Application of Deep Learning Workflow for Autonomous Grain Size Analysis. *Molecules*, 27(15). <https://doi.org/10.3390/molecules27154826>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Busi, R. A., Prasanna, V., Bhavana, S. S. L., Kishore, T. Y., Santhi, V., & Sainath, Y. V. (2024). Rice Grains Detection, Classification, and Quality Prediction Using Deep Learning. *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2024.59321>
- Capajaña, M. Q., & Capajaña, M. Q. (2020). Estudio de calidad física y calidad de taza de café (*Coffea arabica* L.) de dos zonas de cultivo de la provincia de Sandia cosecha 2019. *Revista Científica I+D Aswan Science*. <https://doi.org/10.51392/rcidas.v1i1.2>
- Chen, Y., Bao, J., Chen, R., Li, B., Yang, Y., Renteria, L., Delgado, D., Forbes, B., Goldman, A. E., Simhan, M., Barnes, M., Laan, M., McKeever, S. A., Hou, Z. J., Chen, X., Scheibe, T. D., & Stegen, J. (2024). Quantifying Streambed Grain Size, Uncertainty, and Hydrobiogeochemical Parameters Using Machine Learning Model YOLO. *Water Resources Research*. <https://doi.org/10.1029/2023wr036456>
- Choi, J., Kim, S., Jin, J., Kim, J., Chang, S., & Kim, I.-S. (2025). Estimation of Sediment Grain Size Distribution Using Optical Image-Based Spatial Feature Representation Learning with Data Augmentation. *Journal of Marine Science and Engineering*. <https://doi.org/10.3390/jmse13061108>
- Choudhary, A. K., Choudhary, A. K., Jansche, A., Jansche, A., Grubesa, T., Grubesa, T., Trier, F., Trier, F., Goll, D., Goll, D., Bernthaler, T., Bernthaler, T., Schneider, G., & Schneider, G. (2022). Grain size analysis in permanent magnets from Kerr microscopy images using machine learning techniques. *Materials Characterization*. <https://doi.org/10.1016/j.matchar.2022.111790>
- Cinar, I., & Koklu, M. (2022). Identification of Rice Varieties Using Machine Learning Algorithms. *Tarim Bilimleri Dergisi*, 28(2), 307–325. <https://doi.org/10.15832/ankutbd.862482>

- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Daud, M. S. H., Tabatabai, S. M., & Wong, F. K. (2024). Cascaded Machine Learning in NMR: Unveiling a Continuous Grain Size Distribution Approach for Tackling Sand Production Challenges. *SPWLA 65th Annual Symposium Transactions*. <https://doi.org/10.30632/spwla-2024-0077>
- de C. Motta, I. V., Vuillerme, N., Pham, H.-H., & Pereira de Figueiredo, F. A. (2024). *Machine Learning Techniques for Coffee Classification: A Comprehensive Review of Scientific Research*. <https://doi.org/10.20944/preprints202406.1462.v1>
- Dhawal, K., Sivaramakrishnan, U., Zhang, X., Belay, K., Oakes, J., Wei, X., & Li, S. (2023). Machine Learning Analysis of Hyperspectral Images of Damaged Wheat Kernels. *Sensors*, 23(7). <https://doi.org/10.3390/s23073523>
- Don, H. J., Boustani, S., Yang, C., & Shanks, D. R. (2024). A grain of truth in the grain size effect: Retrieval practice is more effective when interspersed during learning. *Journal of Experimental Psychology. Learning, Memory and Cognition*. <https://doi.org/10.1037/xlm0001382>
- El Akrouchi, M., Mhada, M., Gracia, D. R., Hawkesford, M. J., & Gérard, B. (2025a). Optimizing Mask R-CNN for enhanced quinoa panicle detection and segmentation in precision agriculture. *Frontiers in Plant Science*, 16. <https://doi.org/10.3389/fpls.2025.1472688>
- El Akrouchi, M., Mhada, M., Gracia, D. R., Hawkesford, M. J., & Gérard, B. (2025b). Optimizing Mask R-CNN for enhanced quinoa panicle detection and segmentation in precision agriculture. *Frontiers in Plant Science*, 16. <https://doi.org/10.3389/fpls.2025.1472688>
- García, L., Castro, F. A., Hernández-Amasifuen, A. D., Corazon-Guivin, M. A., Vásquez, J. A., Guerrero-Abad, J. C., Arellanos, E., Veneros, J., Rojas B., N. B., Quintana, S. C., & Oliva, M. (2021). Global studies of cadmium in relation to Theobroma cacao: A bibliometric analysis from Scopus (1996 -2020). In *Scientia Agropecuaria* (Vol. 12, Issue 4, pp. 611–623). Universidad Nacional de Trujillo. <https://doi.org/10.17268/sci.agropecu.2021.065>
- Girma, A., Amare, G., Kebede, B., Kebede, B., Korji, D., & Korji, D. (2021). Pre-extension Demonstration of Improved Faba Bean Varieties in Highlands of Guji Zone, Southern Oromia, Ethiopia. *International Journal of Applied Agricultural Sciences*. <https://doi.org/10.11648/j.ijaas.20210706.11>
- Ibrahim, S., Zulkifli, N. A., Sabri, N., Shari, A. A., & Noordin, M. R. M. (2019). Rice grain classification using multi-class support vector machine (SVM). *IAES International Journal of Artificial Intelligence*, 8(3), 215–220. <https://doi.org/10.11591/ijai.v8.i3.pp215-220>
- King, D. A., King, D. A., Hunt, M., Hunt, M. C., Barbut, S., Barbut, S., Claus, J. (Jim), Claus, J., Cornforth, D., Cornforth, D., Joseph, P., Joseph, P., Kim, B., Kim, Y. H. B., Kim, Y. H. B., Lindahl, G., Lindahl, G., Mancini, R. A., Mancini, R. A., ... Weber, M. J. (2022). AMSA Meat Color Measurement Guidelines. *Meat and Muscle Biology*. <https://doi.org/10.22175/mmb.12473>
- Laub, M., Bachmann, B.-I., Detemple, E., Scherff, F., Staudt, T., Müller, M., Britz, D., Mücklich, F., & Motz, C. (2022). Determination of grain size distribution of prior austenite grains through a combination of a modified contrasting method and machine learning. *Praktische Metallographie*. <https://doi.org/10.1515/pm-2022-1025>
- Laudari, S., Laudari, S., Marks, B., Marks, B., Rognon, P., & Rognon, P. (2022). Classifying grains using behaviour-informed machine learning. *Scientific Reports*. <https://doi.org/10.1038/s41598-022-18250-4>
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- Liu, X., Sha, Z., & Liang, J. (2024). Multibranch Block-Based Grain Size Classification Of Hybrid Disk Using Ultrasonic Scattering: A Deep Learning Method. *Materials Evaluation*. <https://doi.org/10.32548/2024.me-04388>
- Liu, Y., Wang, T., Wen, T., Zhang, J., Liu, B., Li, Y., Zhang, H., Rong, X., Ma, L., Guo, F., Liu, X., & Sun, Y. (2024). Deep learning-based grain-size decomposition model: A feasible solution for dealing with methodological uncertainty. *Sedimentology*. <https://doi.org/10.1111/sed.13195>
- López, M. S. L., & Jimenez, K. G. S. (n.d.). *Efectos de la implementación de medidas sanitarias y fitosanitarias en la exportación del grano de cacao a Países Bajos en el marco del Tratado de Libre Comercio entre Perú y la Unión Europea, durante el periodo 2009 y 2019*. <https://doi.org/10.19083/tesis/653376>
- MacQueen, J. B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281–297. <https://projecteuclid.org/euclid.bsmmsp/1200512992>
- María, A. L. P. S., Ninaquispe, J. C. M., Capuñay, L. M. P., Salazar, P. O. M., Díaz, R. A., & Tocto, R. I. M. (2024). Dinámica de la producción y exportación de quinua (*Chenopodium quinoa*) en Perú. *Revista Alfa*. <https://doi.org/10.33996/revistaalfa.v8i22.250>
- Miranda, L., Huilca, J., & Marques-Pérez, I. (2024). El cultivo reciente de Kiwicha (*Amaranthus caudatus* L.) en el Perú: expansión de producción y comercialización. *Agroalimentaria*. <https://doi.org/10.53766/agroalim/2024.30.58.09>
- Naik, N. K., Subbarao, M. V., Sethy, P. K., Behera, S. K., & Panigrahi, G. R. (2024). Machine learning with analysis-of-variance-based method for identifying rice varieties. *Journal of Agriculture and Food Research*, 18. <https://doi.org/10.1016/j.jafr.2024.101397>
- Oscar, S. T., Etasmo, S. S., & Magomedov, I. A. (2024). Convolutional Neural Networks Applied to the Performance of a Coffee Tree. *E3S Web of Conferences*, 537. <https://doi.org/10.1051/e3sconf/202453708015>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pastor, C. S. S. C., Vidal, L. A. G., Luk, M. C., & Garcia, A. M. R. (2022). Relación y causalidad de la competitividad de la exportación y el precio del grano de cacao en Perú. *Natura@economía*. <https://doi.org/10.21704/ne.v7i2.2204>
- Paz, P., Paz, P., Andrea, D., & Andrea, D. (2018). *Estudio comercial y diseño de una guía para la exportación de barras de cereal a base de quinua a Brasil*. <https://www.semanticscholar.org/paper/8782ee1b42332b2b4e311efaa9ff90a6a1f72cc1>
- Ramdhani, Y., & Alamsyah, D. P. (2023). Enhancing Sustainable Rice Grain Quality Analysis with Efficient SVM Optimization Using Genetic Algorithm. *E3S Web of Conferences*, 426. <https://doi.org/10.1051/e3sconf/202342601035>
- Reza, M. N., Na, I. S., Baek, S. W., & Lee, K. H. (2019). Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images. *Biosystems Engineering*, 177, 109–121. <https://doi.org/10.1016/j.biosystemseng.2018.09.014>
- Salas, M. F. R., & Salas, M. F. R. (2022). *Plan de exportación de café tostado en grano al mercado de Seúl-Corea del Sur*. <https://doi.org/10.21142/tb.2022.2223>
- Saxena, P., Priya, K., Goel, S., Aggarwal, K., Sinha, A., & Jain, P. (n.d.). RICE VARIETIES CLASSIFICATION USING MACHINE LEARNING ALGORITHMS. *Journal of Pharmaceutical Negative Results*, 13, 2022. <https://doi.org/10.47750/pnr.2022.13.S07.479>
- Sharon Ayme Aguirre Matos. (n.d.). *Factores que influyen en la exportación del cacao en el Perú en los años 2020 a 2022*. <https://doi.org/10.19083/tesis/673844>
- Shobana, G., Bushra, S. N., Maheswari, K. U., & Subramanian, N. (2022). Multivariate Classification of Dry Beans using Pipelined Dimensionality Reduction Technique. *International Conference on Signals and Electronic Systems*. <https://doi.org/10.1109/icses55317.2022.9914079>
- Sokudlor, N., Laloon, K., Junsiri, C., & Sudajan, S. (2023). Enhancing milled rice qualitative classification with machine learning techniques using morphological features of binary images. *International Journal of Food Properties*. <https://doi.org/10.1080/10942912.2023.2264533>

- Swathi, T., Karinki Mani Kishore, Jujjavarapu Usha Sri, Ramasetti Chinni Krishna, & Bhavani, A. D. (2025). A Survey on Comprehensive Rice Grain Quality Analysis using Deep Learning. *INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*. <https://doi.org/10.55041/ijrem43693>
- Taspinar, Y. S., Taspinar, Y. S., Dogan, M., Dogan, M., Cinar, I., Çinar, İ., Kursun, R., Kursun, R., Ozkan, I. A., Özkan, İ. A., Koklu, M., & Köklü, M. (2022). Computer vision classification of dry beans (*Phaseolus vulgaris* L.) based on deep transfer learning techniques. *European Food Research and Technology*. <https://doi.org/10.1007/s00217-022-04080-1>
- Tello, A. M. V., Tello, R. J. O., Galarza, M. O. E., & Fuentes, L. P. C. (2024). Estrategias Competitivas para Incrementar la Exportación de Cacao en Grano Ecuatoriano hacia Corea del Sur en el 2024. *Ciencia Latina Revista Científica Multidisciplinar*. [https://doi.org/10.37811/cl\\_rcm.v8i4.12180](https://doi.org/10.37811/cl_rcm.v8i4.12180)
- Verma, D., Bhatt, U. M., & Vidyarthi, A. (2023). Optimal Machine Learning Model for the Relationship Between Grain Size, Channel Thickness, and Grain Boundary Trap Density in 3D NAND Strings. *2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM)*. <https://doi.org/10.1109/elexcom58812.2023.10370202>
- Yang, Y., Xiao, Y., Chen, Z., Tang, D.-S., Li, Z., & Li, Z. (2023). FCBTYOLO: A Lightweight and High-Performance Fine Grain Detection Strategy for Rice Pests. *IEEE Access*. <https://doi.org/10.1109/access.2023.3314697>
- Yasar, A. (2024). Analysis of selected deep features with CNN-SVM-based for bread wheat seed classification. *European Food Research and Technology*, 250(6), 1551–1561. <https://doi.org/10.1007/s00217-024-04488-x>
- Zerrouki, N., Ait-Djafer, A. Z., Harrou, F., Lafane, S., Abdelli-Messaci, S., & Sun, Y. (2024). Image-driven machine learning for automatic characterization of grain size and distribution in smart vanadium dioxide thin films. *Measurement*. <https://doi.org/10.1016/j.measurement.2024.114791>
- Zhu, L., Ma, G., Zhang, Y., Wang, J., & Kan, X. (2023). Reconstruction of Snow Cover in Kaidu River Basin via Snow Grain Size Gap-filling Based on Machine Learning. *Water*. <https://doi.org/10.3390/w15213726>
- 张 Z. 朝 Z., Li, Y. 李, Yang, G. 杨, Zeng, Q. 曾, 李 X. 小龙 L. I., Chen, L. 陈, Qian, D. 钱, Sun, D. 孙, Su, M. 苏, Yang, L. 杨, Zhang, S. 张, & Ma, X. 马. (2023). Estimating the grain size of microgranular material using laser-induced breakdown spectroscopy combined with machine learning algorithms. *Plasma Science and Technology*. <https://doi.org/10.1088/2058-6272/ad1792>

### Contribution of each author to the manuscript:

Task	% of contribution of each author			
	A1	A2	A3	A4
A. theoretical and conceptual foundations and problematization:	25%	25%	25%	25%
B. data research and statistical analysis:	25%	25%	25%	25%
C. elaboration of figures and tables:	25%	25%	25%	25%
D. drafting, reviewing and writing of the text:	25%	25%	25%	25%
E. selection of bibliographical references	25%	25%	25%	25%
F. discussion of results and concepts	25%	25%	25%	25%

### Indication of conflict of interest

There is no conflict of interest

### Source of funding

There is no source of funding

### Acknowledgment

There is no acknowledgment