Journal of Agricultural Sciences (JAS) 32(4):1165-1187
DOI: 10.15832/ankutbd.1946462

J Agr Sci-Tarim Bili
e-ISSN: 2148-9297
jas.ankara.edu.tr

From Handcrafted Descriptors to Deep Embeddings: A Comparative Study for Regional Hazelnut Image Classification

Özgür Tomak[a]*

[a]*Department of Computer Engineering, Faculty of Engineering, Giresun University, Giresun, TÜRKİYE*

ARTICLE INFO
Research Article
Corresponding Author: Özgür Tomak, E-mail: ozgur.tomak@giresun.edu.tr
Received: 07 May 2026 / Revised: 09 June 2026 / Accepted: 23 June 2026 / Online: 12 October 2026

Cite this article
Tomak Ö (2026). From Handcrafted Descriptors to Deep Embeddings: A Comparative Study for Regional Hazelnut Image Classification. *Journal of Agricultural Sciences (Tarim Bilimleri Dergisi),* 32(4):1165-1187. DOI: 10.15832/ankutbd.1946462

ABSTRACT
Automated regional hazelnut image classification is relevant for agricultural traceability, origin aware sorting, and decision support in food processing systems. This study presents a unified comparative benchmark for three class hazelnut image classification using Van, Ordu, and Giresun categories. The experimental framework integrates classical handcrafted descriptors, deep transfer learning models, and reduced space hybrid classical and quantum compatible classifiers. In the classical branch, HOG and RGB color histogram features were combined and evaluated using KNN and SVM based classifiers. In the deep learning branch, ResNet50 and EfficientNetB0 were trained through a two stage transfer learning strategy, and Grad-CAM was employed to provide qualitative interpretability. In the hybrid branch, 1280 dimensional embeddings extracted from the GAP layer of the best EfficientNetB0 model were compressed into low dimensional spaces for reduced space SVC and QSVC experiments. Under the present image level split protocol, the strongest handcrafted baseline, PCA + SVM, achieved 98.90% accuracy and 0.9888 macro-F1, while ResNet50 reached 99.45% accuracy and 0.9945 macro-F1. EfficientNetB0 achieved 100% test accuracy, while the reduced space classical SVC trained on compressed EfficientNetB0 embeddings attained 99.63% accuracy and 0.9961 macro-F1 on the full test set. Under the fair subset protocol, classical SVC achieved 98.33% accuracy, whereas the best QSVC configuration reached 90.83% accuracy. This result should be interpreted as a strong within dataset benchmark rather than definitive evidence of object level or deployment level generalization. Grad-CAM visualizations further suggested that the deep models focused primarily on semantically meaningful hazelnut regions rather than on background patterns. The findings indicate that representation quality is a central factor in this task, while the reported near ceiling performance should be interpreted in light of the image level partitioning strategy.

Keywords: Hazelnut Image Classification, Transfer Learning, Grad-CAM, Reduced Space Classification, Quantum Support Vector Classifier

1. Introduction

Automated image based inspection has become an important component of modern agricultural and food processing systems because product quality, traceability, and process consistency increasingly depend on fast, repeatable, and objective visual evaluation. Manual inspection is still widely used in many agri-food workflows, however, it is limited by subjectivity, operator fatigue, low repeatability, and restricted scalability. For this reason, computer vision and machine learning methods have become practical tools for non destructive agricultural inspection, grading, and decision support (Xiao et al. 2022; Liao et al. 2026). Hazelnut image analysis is a relevant example of this broader problem because shell appearance, surface texture, color variation, and local structural patterns may provide discriminative visual cues for classification and quality related assessment (Silvestri et al. 2021; Hossen et al. 2025).

In this study, the classification problem is formulated as a three class regional hazelnut image recognition task using the dataset defined categories Van, Ordu, and Giresun (Köklü 2024). This setting provides a controlled benchmark for comparing different representation and classification strategies under the same agricultural image analysis problem. The inclusion of these three categories follows the structure of the available dataset and enables a multi class regional comparison. Therefore, the present study should not be interpreted as a complete geographical origin authentication system in the strict chemical, genetic, or supply chain sense. Instead, it evaluates whether image patterns associated with the predefined regional categories can be discriminated under a controlled machine learning protocol.

A central motivation of the study is to examine how feature representation affects classification performance. Classical handcrafted descriptors, such as Histogram of Oriented Gradients and RGB color histograms, remain useful because they are interpretable, computationally accessible, and suitable for establishing transparent baseline models (Dalal & Triggs 2005; Swain & Ballard 1991). In contrast, pretrained convolutional neural networks can learn more abstract and hierarchical visual representations through transfer learning, which is especially useful in agricultural tasks where dataset sizes are often moderate

but discriminative visual patterns may be subtle (Georgiou et al. 2020; Hossen et al. 2025). Accordingly, this study compares handcrafted feature based classifiers with pretrained deep transfer learning models, including ResNet50 and EfficientNetB0 (He et al. 2016; Tan & Le 2019). In addition, Grad-CAM is incorporated to qualitatively examine whether deep models focus on meaningful hazelnut regions rather than irrelevant background patterns (Selvaraju et al. 2017; Geirhos et al. 2020).

The study further extends the comparison to reduced space hybrid classification. Deep embeddings extracted from the strongest pretrained model are compressed into low dimensional spaces and evaluated using both classical SVC and quantum compatible QSVC models. This design is motivated by the current practical constraints of quantum machine learning for image classification: raw image data are too high dimensional for direct quantum processing in realistic settings, so quantum compatible models generally require compact and informative feature representations (Havlíček et al. 2019; Schuld & Killoran 2019; Senokosov et al. 2024). In this framework, QSVC is not treated as a replacement for deep learning, but as a downstream reduced space classifier applied to compressed deep embeddings.

The main contribution of this study is a unified comparative benchmark that evaluates handcrafted descriptors, pretrained deep transfer learning, Grad-CAM based interpretability, reduced deep embeddings, and quantum compatible classification under the same regional hazelnut image classification protocol. Unlike previous hazelnut image studies that mostly focus on a single task type or one methodological family, such as variety recognition, defect detection, kernel sorting, or transfer learning based classification (Giraudo et al. 2018; Taner et al. 2021; Ünal & Aktaş 2023), the present study investigates the interaction between representation quality, classifier choice, dimensionality reduction, and interpretability in a single controlled framework. A bootstrap based robustness analysis and a descriptive representation level correlation analysis are included to support the stability of the comparative findings and to examine whether representation quality plays a central role in model performance.

Accordingly, the study is guided by the following research questions: (i) How do handcrafted HOG + RGB histogram descriptors perform compared with pretrained deep transfer learning models for regional hazelnut image classification? (ii) To what extent do EfficientNetB0 deep embeddings preserve discriminative class information after severe dimensionality reduction? (iii) How does QSVC perform against a classical SVC baseline under the same reduced space and balanced subset conditions? (iv) Do Grad-CAM visualizations suggest that deep models focus on semantically meaningful hazelnut regions rather than on background patterns?

2. Related Work

2.1. Literature search and selection strategy

To improve transparency and reproducibility, the related studies were selected through a structured literature search focusing on hazelnut image analysis, agricultural machine vision, transfer learning in agri-food inspection, explainable convolutional neural networks, and quantum or quantum compatible machine learning for image classification. The literature search considered peer reviewed journal articles, conference papers, and recent review studies indexed in major scientific databases, including Scopus, Web of Science, ScienceDirect, SpringerLink, IEEE Xplore, and Google Scholar.

The search terms included combinations of the following keywords: “hazelnut classification”, “hazelnut image analysis”, “hazelnut defect detection”, “hazelnut kernel classification”, “agricultural image classification”, “machine vision in food inspection”, “transfer learning in agriculture”, “Grad-CAM agricultural vision”, “quantum machine learning image classification”, and “quantum support vector classifier”. Studies were included if they met at least one of the following criteria: (i) they addressed hazelnut or closely related agricultural product classification using image data, (ii) they used handcrafted visual descriptors or classical machine learning classifiers, (iii) they applied pretrained convolutional neural networks or transfer learning to agricultural image classification, (iv) they incorporated explainability techniques such as Grad-CAM, or (v) they provided methodological relevance to reduced space or quantum compatible image classification.

Studies were excluded if they did not involve image based classification, did not report sufficient methodological detail, were unrelated to agricultural or food image analysis, or focused only on non visual chemical, genetic, or supply chain origin authentication. Additional relevant studies exist in broader agricultural vision and quantum machine learning, however, the final comparison prioritized representative works that were methodologically closest to the present study. This selection strategy was used to position the proposed benchmark against classical handcrafted descriptors, deep transfer learning, explainable visual analysis, and hybrid quantum compatible classification.

2.2. Thematic overview of related studies

Computer vision has become an important component of agri-food inspection because it enables non destructive, repeatable, and scalable analysis of product appearance, surface quality, and sorting related attributes. In food and agricultural applications, image based systems are used not only to improve inspection speed, but also to reduce subjectivity and increase consistency in quality evaluation. Recent studies in agricultural vision indicate a gradual shift from handcrafted image analysis toward deep learning based pipelines, especially when visual differences are complex, nonlinear, or difficult to describe manually (Xiao et

al. 2022; Hossen et al. 2025). This transition is directly relevant to hazelnut image analysis because hazelnut samples may contain discriminative information in shell morphology, color distribution, surface texture, local defects, and regional visual patterns.

Hazelnut related image studies can be grouped into several partially overlapping task categories. One group focuses on cultivar or variety recognition, where CNN based models are used to distinguish different hazelnut varieties or visually similar groups (Taner et al. 2021). Another group addresses industrial inspection and postharvest sorting problems, including the separation of whole kernels from damaged kernels, shells, undersized samples, or defective hazelnuts (Giraudo et al. 2018; Ünal & Aktaş 2023). These studies demonstrate that hazelnut image analysis is not a single homogeneous problem, but a family of related classification tasks. Therefore, results obtained for defect detection, kernel sorting, cultivar recognition, and regional image classification should not be interpreted as directly interchangeable. This distinction is important for the present study because its objective is not general defect detection or kernel quality sorting, but dataset defined regional image classification using the Van, Ordu, and Giresun categories.

Classical machine vision approaches remain important in this literature because they provide interpretable and computationally transparent baselines. Earlier visual recognition methods commonly relied on descriptors that summarize shape, color, gradient, or local structure, such as moment based features, color histograms, and local feature descriptors (Hu 1962; Swain & Ballard 1991; Lowe 2004). Histogram of Oriented Gradients (HOG) is especially relevant because it encodes local gradient orientation patterns and has remained a strong baseline for structured image classification tasks (Dalal & Triggs 2005). In hazelnut imagery, HOG and color based features are meaningful because class related differences may appear through shell boundaries, surface texture, visible cracks, and global chromatic variation. However, handcrafted descriptors are also limited because their success depends strongly on the quality of manually designed features and may decrease when class boundaries are visually subtle or spatially heterogeneous.

Deep learning approaches address this limitation by learning hierarchical visual representations directly from data. Transfer learning is particularly useful in agricultural image classification because domain specific datasets are often moderate in size, while pretrained convolutional networks provide strong generic visual priors learned from large scale image collections (Georgiou et al. 2020; Hossen et al. 2025). Architectures such as ResNet and EfficientNet represent two influential directions in this context. ResNet introduced residual learning to support the training of deeper networks, while EfficientNet proposed compound scaling to balance accuracy and computational efficiency (He et al. 2016; Tan & Le 2019). In hazelnut and agri-food inspection tasks, these architectures are relevant because they can learn both global appearance cues and fine local details that may not be fully captured by handcrafted descriptors. Nevertheless, previous hazelnut deep learning studies have generally focused on specific tasks such as variety recognition, kernel classification, or defect oriented sorting rather than a unified comparison of handcrafted, deep, and reduced space hybrid models.

Interpretability is another important theme in applied agricultural vision. High classification accuracy alone does not guarantee that a model is using meaningful object regions; a model may also exploit shortcut cues such as background regularities, lighting patterns, or acquisition artifacts (Geirhos et al. 2020). Grad-CAM provides a practical way to generate class discriminative visual explanations for CNN predictions and has become widely used to inspect whether deep models attend to plausible regions of the input image (Selvaraju et al. 2017). In the context of hazelnut classification, Grad-CAM is useful because it can help determine whether predictions are supported by attention on the hazelnut body, shell surface, and texture regions rather than irrelevant background areas. Therefore, explainability should be considered a complementary component of model evaluation rather than a replacement for quantitative performance metrics.

More recently, quantum machine learning has been explored as an alternative framework for nonlinear classification through quantum feature maps, quantum kernels, and variational circuits. Foundational studies suggest that quantum feature spaces may provide different similarity structures from conventional classical kernels, which motivates QSVC style classification experiments (Havlíček et al. 2019; Schuld & Killoran 2019). However, raw image data are too high dimensional for direct quantum encoding under current practical constraints. For this reason, image based quantum machine learning is more realistically studied through hybrid pipelines in which classical feature extraction or deep embedding generation is followed by dimensionality reduction and quantum compatible classification (Mari et al. 2020; Alam et al. 2021; Senokosov et al. 2024). This point is central for the present study because QSVC is not used as a direct image classifier, but as a downstream classifier applied to low dimensional PCA compressed EfficientNetB0 embeddings.

Taken together, the literature shows that handcrafted descriptors, pretrained CNNs, explainable deep learning, and quantum compatible classification each provide useful but different perspectives for agricultural image classification. However, previous hazelnut studies have usually focused on a single methodological family or a specific task type, such as variety classification, defect detection, or kernel sorting. Fewer studies evaluate handcrafted descriptors, deep transfer learning, Grad-CAM analysis, reduced deep embeddings, and QSVC under the same fixed dataset partition. The present study addresses this gap by positioning these methodological families within a single controlled benchmark for regional hazelnut image classification.

2.3. Structured comparison of previous studies

A structured comparison of representative studies related to hazelnut image analysis, agricultural vision, explainable deep learning, and quantum compatible image classification is presented in Table 1.

Table 1 shows that previous studies provide important but mostly separate contributions to hazelnut image analysis, agricultural vision, explainable deep learning, and quantum machine learning. Classical hazelnut studies demonstrate the usefulness of handcrafted visual descriptors, whereas deep learning studies show the strength of CNN based feature learning for variety recognition, defect detection, and kernel sorting. However, these studies generally focus on a single methodological family or a specific task type. In contrast, the present study combines handcrafted descriptors, pretrained deep models, Grad-CAM based interpretability, reduced deep embeddings, and QSVC within a single controlled regional hazelnut image classification benchmark.

Table 1- Structured comparison of representative studies related to the present work

Study	Method type	No. of classes	Task type	Dataset / domain	Feature representation	Deep learning	Main relevance or limitation
Hu (1962)	Handcrafted descriptor	N/A	Shape representation	General image analysis	Moment invariants	No	Foundational shape descriptor
Swain & Ballard (1991)	Handcrafted descriptor	N/A	Color based recognition	General computer vision	Color histogram / color indexing	No	Foundational color representation, useful for baseline comparison
Lowe (2004)	Handcrafted descriptor	N/A	Local feature matching	General computer vision	SIFT local descriptors	No	Important local descriptor, not designed for agricultural inspection
Dalal & Triggs (2005)	Handcrafted descriptor	2	Object detection	General image analysis	HOG	No	Establishes HOG as a strong structural descriptor
Giraudo et al. (2018)	Classical machine vision	2 / defect oriented	Defective hazelnut identification	Hazelnut images	RGB image analysis / colourgrams	No	Hazelnut specific classical baseline, not a deep benchmark
Taner et al. (2021)	Deep learning	Multiple varieties	Hazelnut variety classification	Hazelnut variety images	CNN learned features	Yes	Hazelnut specific deep learning, focused on variety recognition
Xiao et al. (2022)	Review	N/A	Food inspection	Food detection studies	Machine vision features	Mixed	Shows broader relevance of machine vision in food inspection
Ünal & Aktaş (2023)	Deep learning	Multiple classes	Hazelnut kernel classification	Hazelnut kernel images	CNN learned features	Yes	Focuses on kernel/shell/defect sorting rather than regional classification
Tomak (2025)	Deep learning	Hazelnut classes	Hazelnut classification	Hazelnut images	CNN learned features	Yes	Relevant pretrained/deep hazelnut classification study
Liao et al. (2026)	Review	N/A	Food science and technology	Food image studies	Deep representations	Yes	Shows recent shift toward deep learning in food vision
He et al. (2016)	Deep learning architecture	1000 ImageNet classes	Generic image classification	ImageNet	Residual CNN features	Yes	Provides ResNet backbone used for transfer learning
Tan & Le (2019)	Deep learning architecture	1000 ImageNet classes	Generic image classification	ImageNet	EfficientNet features	Yes	Provides EfficientNet backbone used for transfer learning
Selvaraju et al. (2017)	Explainable AI	N/A	Visual explanation	CNN predictions	Grad-CAM heatmaps	Yes	Provides interpretability method used in this study
Geirhos et al. (2020)	Deep learning analysis	N/A	Shortcut learning	General deep learning	Model behavior analysis	Yes	Motivates checking whether models rely on shortcut cues
Havlíček et al. (2019)	Quantum machine learning	Synthetic tasks	Quantum kernel classification	Quantum feature spaces	Quantum feature map	No classical DL	Foundational QSVC motivation, not image specific
Schuld & Killoran (2019)	Quantum machine learning	N/A	Quantum feature spaces	QML theory	Quantum embeddings / kernels	No	Provides theoretical basis for quantum feature space learning
Mari et al. (2020)	Hybrid quantum classical learning	Benchmark tasks	Transfer learning / QML	Hybrid ML datasets	Classical features + quantum model	Partial	Shows feasibility of hybrid quantum classical design
Alam et al. (2021)	Quantum / hybrid learning	Benchmark tasks	Quantum neural classification	Image like or benchmark datasets	Reduced features / quantum circuits	Partial	Relevant to reduced space QML but not hazelnut specific
Senokosov et al. (2024)	Quantum image classification	Image classification tasks	QML for images	Image datasets	Quantum compatible representations	Partial	Highlights constraints and potential of QML for images
Present study	Classical + deep + explainable + QSVC	3	Regional hazelnut image classification	Van, Ordu, Giresun hazelnut images	HOG + RGB, CNN embeddings, PCA reduced embeddings	Yes	Unified benchmark comparing handcrafted, deep, interpretable, and quantum compatible branches

2.4. Research gap and positioning of the present study

Existing hazelnut image studies have typically focused on cultivar recognition, kernel sorting, defect detection, or single family deep learning models. Although these studies are valuable, they do not fully clarify whether classification performance is mainly driven by handcrafted visual descriptors, deep representation learning, dimensionality reduction, or classifier choice. Moreover, quantum compatible models have rarely been evaluated together with classical and deep learning baselines in a controlled hazelnut image classification setting. The present study addresses this gap by comparing handcrafted HOG + RGB descriptors, pretrained CNN models, Grad-CAM visual explanations, reduced EfficientNetB0 embeddings, classical SVC, and QSVC under the same fixed dataset partition.

3. Material and Methods

3.1. Dataset and classification task

The study addressed a three class regional hazelnut image classification problem using the publicly available Hazelnut Images Dataset provided by Köklü (2024). The dataset is organized into three predefined regional categories: Van, Ordu, and Giresun. These categories were retained because they represent the original class structure of the dataset and allow the construction of a controlled three class visual benchmark. Ordu and Giresun are directly relevant because they are strongly associated with hazelnut production in Türkiye, while Van provides a visually and regionally distinct third category that increases the difficulty of the multi class classification task. Therefore, the inclusion of Van should be interpreted as part of the dataset defined regional category structure rather than as a claim of comprehensive geographical origin authentication.

All experiments were conducted under a fixed train/validation/test partition consisting of 2538 training images, 544 validation images, and 545 test images. The same split was preserved throughout the classical, deep learning, and hybrid classification branches in order to ensure a fully controlled comparison across methodological families. This design was adopted because the main objective of the study was not only to identify the best performing model, but also to compare how different feature representations and classifier families behave under the same experimental protocol. Representative images from the three dataset defined regional classes are shown in Figure 1.



Figure 1- Representative hazelnut images from the Van, Ordu, and Giresun classes used in the regional classification task

Figure 1 illustrates representative samples from the three dataset defined regional classes. The images show that the classification task is based on visual differences in hazelnut appearance, including shell color, surface texture, shape, and local structural patterns. Since the images were acquired under a controlled dataset structure, the figure should be interpreted as an example of the visual input used for model comparison rather than as evidence of complete geographical origin authentication.
The class distribution across the training, validation, and test subsets is summarized in Table 2.

Table 2- Class distribution across the training, validation, and test partitions used in the study

Class	***Train***	***Validation***	***Test***	***Total***
Van	796	171	171	1138
Ordu	927	198	199	1324
Giresun	815	175	175	1165
Total	2538	544	545	3627

Table 2 shows that the dataset is moderately balanced across the three regional classes. Ordu has the largest number of images, followed by Giresun and Van, however, the difference between class sizes is not large enough to create a severely imbalanced classification problem. The fixed split also preserves class representation across the training, validation, and test subsets. This structure supports a controlled comparison of classical, deep, and hybrid model families under the same data distribution.

An important limitation of the dataset partitioning is that a strict object level split could not be reconstructed from the available file naming structure. The image filenames, such as Giresun 1165.jpg, Ordu 1324.jpg, and Van 1138.jpg, identify the class and image index but do not provide a reliable physical object identifier. Therefore, the present partition should be interpreted as an image level split. This protocol supports a controlled within dataset comparison across model families, but it may overestimate deployment level generalization if multiple visually similar images or multiple views of the same physical hazelnut are distributed across training and test subsets.

Regarding dataset traceability, a DOI indexed version of the Hazelnut Images Dataset could not be identified. Therefore, the dataset was cited using its original publicly available source URL provided by Köklü (2024). In the present study, the class labels were inherited from the original dataset organization, where images were provided in class specific folders. Before the experiments, the dataset structure was checked to confirm the presence of the three target classes and to construct a fixed train/validation/test partition. Since the present study used the dataset as publicly released, the reliability of the empirical findings should be interpreted together with the limitations of the original dataset documentation.

The suitability of this dataset for controlled image based hazelnut classification is also supported by its use in recent peer reviewed studies. Gencturk et al. (2024) used the same Giresun, Ordu, and Van hazelnut image dataset and reported 100% classification accuracy using an InceptionV3 + ResNet50 data fusion model. Similarly, Dönmez et al. (2024) evaluated BigTransfer models on the same open access dataset and reported 99.49% accuracy with the BiT-M R152 × 4 model. These independent applications indicate that the dataset contains visually discriminative patterns that can be learned by modern image classification models under controlled experimental settings. However, such results should be interpreted as indirect evidence of dataset utility rather than as definitive proof of complete dataset validity, because the original source does not provide exhaustive documentation of the data collection, object level grouping, or annotation protocol.

Let the dataset be denoted by Equation (1)

$$D = \{(I_i, y_i)\}_{i=1}^{N} \tag{1}$$

Where: I_i is the i-th hazelnut image and $y_i \in \{1,2,3\}$ denotes its regional class label corresponding to Van, Ordu, or Giresun. The full dataset is divided into training, validation, and test subsets, denoted by D_{tr}, D_{val}, and D_{te}, respectively. Equation (1) formalizes the dataset as a collection of image label pairs. Here, each image is assigned to one of the three regional class labels corresponding to Van, Ordu, or Giresun. The same notation is used throughout the study to describe the training, validation, and test subsets. This formalization clarifies that the experimental unit in the present benchmark is the image record rather than a verified physical hazelnut object. The overall experimental workflow is illustrated in Figure 2.

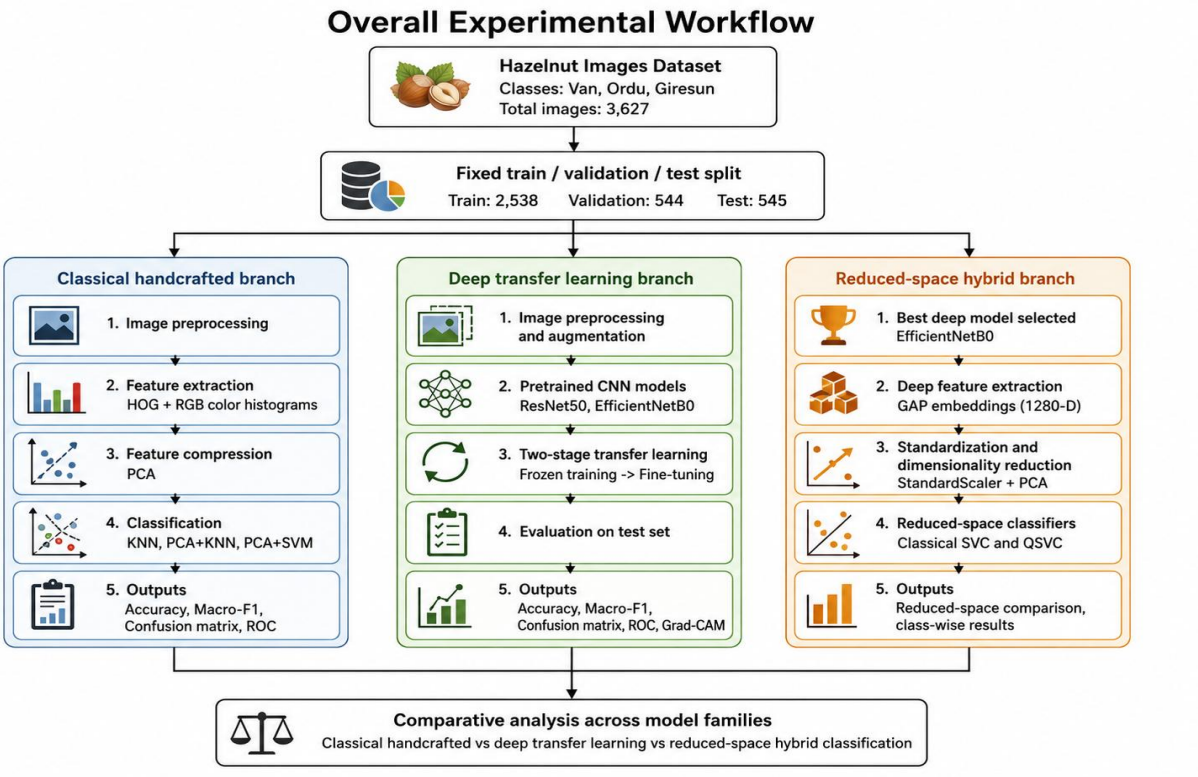


Figure 2- Overview of the classical, deep learning, and reduced space hybrid experimental branches used in the study

Figure 2 summarizes the complete experimental design of the study. The workflow begins with a fixed train/validation/test split and then separates the comparison into three methodological branches: handcrafted classical classification, deep transfer learning, and reduced space hybrid classification. This structure makes it possible to compare different representation and classifier families under the same dataset partition. It also highlights that the quantum compatible branch was not applied directly to raw images, but to compressed deep embeddings extracted from the best performing pretrained model.

3.2. Hyperparameter selection and validation strategy

Hyperparameter selection was performed using the validation set, while the test set was kept strictly for final reporting. Macro-F1 was used as the primary model selection metric because the study aimed to evaluate balanced performance across the three regional classes. After the best configuration was selected on the validation set, the final classical and reduced space models were refitted using the combined training and validation partitions and then evaluated on the held out test set. The training subset was used for fitting candidate models, the validation subset was used for hyperparameter selection, and the test subset was reserved only for final performance reporting. Macro-F1 was used as the main selection metric for classical and reduced space classifiers because it provides a balanced evaluation across the three regional classes. The hyperparameter search space, validation strategy, and model selection criteria used across all experimental branches are summarized in Table 3.

Table 3- Hyperparameter search space, validation strategy, and model selection metrics used in the experimental branches

Experimental branch	***Model/ method***	***Candidate hyperparameters***	***Validation strategy***	***Selection metric***	***Final testing protocol***
Classical handcrafted branch	Raw KNN	k = 3, 5, 7, 9	Models were fitted on the training set and compared on the validation set	Validation macro-F1	Best model was refitted on train + validation and evaluated on the test set
Classical handcrafted branch	PCA + KNN	PCA = 95% variance, 100 components, 200 components; k = 3, 5, 7, 9	PCA and KNN configurations were selected using the validation set	Validation macro-F1	Best configuration was refitted on train + validation and evaluated on the test set
Classical handcrafted branch	PCA + SVM	Kernel = linear/RBF; C = 0.1, 1, 10; gamma = scale	SVM configurations were selected using the validation set	Validation macro-F1	Best configuration was refitted on train + validation and evaluated on the test set
Deep learning branch	ResNet50	Two stage transfer learning; frozen backbone followed by fine tuning of the last 30 layers	Best checkpoint was selected using validation performance	Validation loss and validation accuracy	Best restored model was evaluated on the held out test set
Deep learning branch	EfficientNetB0	Two stage transfer learning; frozen backbone followed by fine tuning of the last 30 layers	Best checkpoint was selected using validation performance	Validation loss and validation accuracy	Best restored model was evaluated on the held out test set
Reduced space hybrid branch	SVC on EfficientNetB0 embeddings	PCA dimensions = 4, 6, 8; kernel/C/gamma settings	Candidate reduced space configurations were selected using the validation set	Validation macro-F1	Best model was evaluated on the full test set and/or balanced test subset
Reduced space quantum-compatible branch	QSVC	PCA dimensions = 4, 6, 8; ZZFeatureMap; reps = 1, 2	Candidate quantum compatible configurations were compared on the balanced validation subset	Balanced validation macro-F1	Best QSVC configuration was evaluated on the balanced test subset

Table 3 shows that hyperparameter selection was performed before test set evaluation in all experimental branches. For the classical handcrafted branch, KNN, PCA, and SVM configurations were selected according to validation macro-F1. For the deep learning branch, the best model checkpoints were selected according to validation performance during the two stage transfer learning process, which included a frozen backbone stage and a fine tuning stage. Data augmentation was applied only to the training subset, while validation and test images were evaluated without augmentation. For the reduced space SVC and QSVC experiments, candidate PCA dimensions and classifier configurations were selected using validation performance before final testing. This protocol ensures that the test set was not used during hyperparameter optimization.

3.3. Handcrafted image representation

To establish a conventional computer vision baseline, each input image was resized to 128 × 128 pixels prior to handcrafted feature extraction. This resolution was selected as a compromise between computational efficiency and preservation of sufficient

structural detail for classical descriptors (Dalal & Triggs 2005). The handcrafted representation combined two complementary feature families: HOG, to encode local edge and contour structure, and RGB color histograms, to capture global chromatic distribution. PCA was used to reduce redundancy in the high dimensional HOG + RGB histogram representation. Candidate PCA settings included 95% explained variance retention, 100 components, and 200 components. The best performing handcrafted configuration was selected according to validation macro-F1. For the SVM classifier, both linear and RBF kernels were evaluated with C values of 0.1, 1, and 10. The final PCA + SVM model used PCA with 200 retained components and a linear SVM with C = 0.1.

The HOG descriptor was computed on grayscale images using 9 orientation bins, 8 × 8 pixels per cell, and 2 × 2 cells per block. In parallel, a normalized RGB histogram was extracted using 32 bins per channel. The final descriptor for each image was formed by concatenating the HOG vector with the RGB histogram vector, yielding a 40868 dimensional handcrafted feature representation. The dimensionality of the handcrafted descriptor resulted from the concatenation of the HOG feature vector and the three channel RGB histogram vector extracted from each resized image. This design was chosen deliberately to couple local structural cues with global appearance information, thereby providing an interpretable and reasonably expressive baseline for regional hazelnut discrimination. In the handcrafted branch, each input image I_i is represented by the concatenation of a HOG descriptor and an RGB color histogram. $\phi_{HOG}(.)$, as given in Equation (2), extracts local gradient orientation structure, while $\phi_{HOG}(.)$, as given in Equation (3), produces the normalized color histogram representation. The final handcrafted descriptor x_i, as given in Equation (4), is obtained by concatenating the two feature vectors.

$$x_i^{HOG} = \phi_{HOG}(I_i) \tag{2}$$

$$x_i^{RGB} = \phi_{RGB}(I_i) \tag{3}$$

$$x_i = \left[x_i^{HOG};\ x_i^{RGB}\right] \in R^{40868} \tag{4}$$

Equations (2)-(4) formalize the handcrafted representation used in the classical branch. Equation (2) defines the HOG based structural descriptor, which captures local edge and gradient orientation patterns from the input image. Equation (3) defines the RGB color histogram descriptor, which summarizes the global chromatic distribution of the same image. Equation (4) combines these two complementary descriptors into a single 40,868 dimensional handcrafted feature vector, allowing the classical classifiers to use both local structural information and global color information for regional hazelnut classification.

3.4. Classical machine learning baselines

Three classical baselines were evaluated on the handcrafted feature space. The first baseline was a raw KNN classifier trained directly on the full handcrafted descriptor. The neighborhood size k was selected from the candidate set {3, 5, 7, 9} according to validation performance. This model was intended to assess whether the original handcrafted space was already sufficiently discriminative without explicit compression.

The second baseline was Principal Component Analysis (PCA) compressed KNN, introduced to reduce redundancy and improve geometric compactness in the handcrafted feature space. Given the very high original dimensionality of the descriptor, PCA was used prior to classification. Validation experiments were carried out using both a variance preserving setting and fixed reduced dimensions. This branch allowed the study to test whether a compact projection of the handcrafted space would improve generalization relative to raw KNN.

The third and strongest classical baseline was PCA + SVM. In this branch, handcrafted features were first standardized using z-score normalization, then projected into a lower dimensional PCA space, and finally classified using SVM. Both linear and radial basis function (RBF) kernels were considered during model selection. This design was motivated by the expectation that a margin based classifier would be better suited than KNN to a compact but still high information handcrafted space. In the final rerun workflow, model selection was always performed on the validation set, and the chosen configuration was then retrained on the combined development set before final testing.

Since the handcrafted HOG and RGB histogram features may have different numerical ranges and variances, feature wise standardization was applied before PCA and kernel based classification. This step prevents features with larger numerical scales from dominating the projection and distance based or margin based decision geometry. For each feature dimension j, the standardized value of the i-th sample was computed as given in Equation (5).

$$\tilde{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{5}$$

Where: x_{ij} denotes the original value of feature j for sample i, while μ_j and σ_j represent the mean and standard deviation of that feature estimated from the training data. The resulting standardized vector $\tilde{x}_i$ was then used as input to PCA and subsequent

KNN or SVM classification. To avoid information leakage, the same training derived standardization parameters were applied to the validation and test sets.

Equation (5) defines the standardization procedure applied to each feature dimension before PCA and classifier training. For each feature j, the mean μ_j and standard deviation σ_j were estimated only from the training data and then used to transform the training, validation, and test samples. This procedure ensures that all feature dimensions are placed on a comparable scale and prevents features with larger numerical ranges from dominating the PCA projection or the distance and margin based classifiers. Estimating the scaling parameters only from the training set also avoids information leakage from the validation or test subsets. PCA was used as the principal dimensionality reduction technique in the classical branch. Let $\tilde{x}_i$ denote a standardized handcrafted feature vector and $W_k \in R^{dxk}$ denote the matrix whose columns are the first k principal directions. The reduced representation was then obtained as given in Equation (6).

$$z_i = W_k^T \tilde{x}_i \tag{6}$$

The resulting compact representation was subsequently used as input either to KNN or to SVM. For the SVM case, multiclass classification was handled using the library's built in multiclass strategy.

Equation (6) defines the PCA based dimensionality reduction step used after feature standardization. In this equation, $\tilde{x}_i$ represents the standardized feature vector of sample i, while W_k^T denotes the projection matrix formed by the first k principal components. The resulting vector z_i is the reduced representation of the original feature vector in a lower dimensional PCA space. This transformation reduces redundancy and computational complexity while preserving the most informative variance structure of the handcrafted or deep embedding representation.

After PCA based dimensionality reduction, the compact handcrafted representation was classified using SVM. SVM aims to construct a decision boundary that maximizes the margin between classes in the selected feature space. Since the regional hazelnut categories may not be linearly separable in the reduced handcrafted space, both linear and RBF kernels were evaluated during validation based model selection. The RBF kernel measures the similarity between two reduced feature vectors as an exponentially decreasing function of their squared Euclidean distance as given in Equation (7).

$$K(z_i, z_j) = exp\left(-\gamma \|z_i - z_j\|^2\right) \tag{7}$$

Where: z_i and z_j denote PCA reduced feature vectors, and γ controls the width of the RBF kernel. A smaller γ produces a smoother decision boundary, whereas a larger γ makes the classifier more sensitive to local variations in the feature space. Equation (7) defines the RBF kernel used in the SVC candidate configurations. The kernel assigns higher similarity values to samples that are close to each other in the reduced feature space and lower similarity values to samples that are farther apart. The parameter γ controls how rapidly this similarity decreases as the distance between two samples increases. Therefore, the RBF kernel allows SVC to model nonlinear class boundaries in the PCA reduced representation when such boundaries are present. The regularization parameter controls the trade off between maximizing the margin and penalizing misclassified training samples. In the present multiclass problem, SVM classification was handled using the library level multiclass strategy.

The KNN classifier assigns a test sample to the class most frequently represented among its nearest training samples in the selected feature space. Depending on the experimental branch, this feature space corresponds either to the original handcrafted descriptor or to its PCA compressed representation. Let v_i denote the feature vector of the i-th sample and let $N_K(v_i)$ represent the set of K nearest training samples according to the Euclidean distance. The predicted class label was computed by majority voting among the K nearest neighbors, as given in Equation (8).

$$\hat{y}_i = \arg\ max_{c\in\{1,2,3\}} \sum_{j\in N_K(v_i)} 1\left(y_j = c\right) \tag{8}$$

where: $N_K(v_i)$ denotes the set of K nearest neighbors of sample v_i,y_i is the class label of neighbor j, c represents a candidate class label, and $1(y_j = c)$ is an indicator function that equals 1 when the neighbor belongs to class c and 0 otherwise. The neighborhood size K was selected from the candidate set {3, 5, 7, 9} according to validation performance. Equation (8) defines the majority voting rule used by the KNN classifier. For each test sample, the algorithm identifies the K closest training samples in the selected feature space and counts how many neighbors belong to each class. The class receiving the highest number of votes is assigned as the predicted label. This rule allows KNN to classify hazelnut images based on local similarity patterns in the handcrafted or PCA reduced feature representation.

3.5. Deep transfer learning models

The deep learning branch was designed around pretrained convolutional neural networks operating on 224 × 224 RGB inputs. Two backbones were evaluated, ResNet50 and EfficientNetB0. Both were implemented through transfer learning, with the

ImageNet pretrained convolutional base retained as the initial visual encoder and a task specific classification head appended on top. The goal was to compare two strong but architecturally different pretrained models under the same hazelnut classification task.

For ResNet50, the input pipeline used the model specific preprocessing expected by the TensorFlow/Keras ResNet implementation. A lightweight classification head was added to the frozen backbone, consisting of a Global Average Pooling (GAP) layer, a fully connected layer with ReLU activation, dropout, and a final three unit softmax output layer. For EfficientNetB0, the native TensorFlow/Keras implementation was used, and images were supplied according to the model's expected internal input scaling behavior. A corresponding task specific classification head was constructed using GAP, a compact fully connected layer, dropout, and softmax classification.

Training followed a two stage transfer learning schedule in both deep branches. In Stage 1, the pretrained backbone was frozen and only the newly added classifier head was optimized. In Stage 2, the final portion of the backbone was unfrozen and finetuned with a lower learning rate, while batch normalization layers remained frozen to improve training stability. In the proposed methodology, the deep models were optimized for 10 epochs in Stage 1 and 8 epochs in Stage 2, with learning rates of 10^{-4} and 10^{-5}, respectively. Fine tuning was limited to the final 30 backbone layers, which provided a compromise between adaptation capacity and stability.

Data augmentation was used during training to improve robustness and reduce overfitting. The augmentation pipeline included random rotations, width and height shifts, zooming, shear transformations, and horizontal flips. Model checkpointing, early stopping, adaptive learning-rate reduction, and training history logging were used in both branches to ensure stable optimization and reproducible experiment tracking. This training protocol allowed both deep models to be evaluated under a common and restart safe pipeline suitable for iterative experimentation in the Colab environment. Data augmentation was applied only to the training subset in order to improve robustness to minor visual variations. No augmentation was applied to the validation or test subsets, these subsets were evaluated using deterministic resizing and model specific preprocessing only. Early stopping was used to reduce overfitting, with patience = 5, and the best model weights were restored according to validation performance. The configurations of the deep learning models are presented in Table 4 and fine tuning and augmentation settings of the deep transfer learning models are presented in Table 5. Both deep models were trained using categorical cross entropy loss and the Adam optimizer. The best model checkpoint was selected according to validation performance. Early stopping and adaptive learning rate reduction were used to reduce overfitting and stabilize fine tuning. During fine tuning, batch normalization layers were kept frozen to avoid destabilizing the pretrained feature statistics. In the fine tuning stage, only the last 30 layers of the pretrained backbone were unfrozen. The number of unfrozen layers was used as a predefined conservative fine tuning setting rather than as a test set optimized hyperparameter. This value was adopted as a standard conservative fine tuning depth commonly used in transfer learning studies, where unfreezing the final portion of the backbone is recommended to adapt higher level visual features while preserving the more general lower level representations (He et al. 2016; Tan & Le 2019).This setting was chosen to allow limited adaptation of highlevel visual filters while preserving the more general low level features learned during ImageNet pretraining. It also helped reduce the risk of overfitting under the available dataset size.

Table 4- Main configuration settings of the deep transfer learning models

Model	*Input Size*	*Stage 1 Epochs*	*Stage 2 Epochs*	*Stage 1 LR*	*Stage 2 LR*
ResNet50	224 × 224	10	8	10^{-4}	10^{-5}
EfficientNetB0	224 × 224	10	8	10^{-4}	10^{-5}

Table 4 summarizes the main training configuration of the deep transfer learning models. Both ResNet50 and EfficientNetB0 were trained using 224 × 224 RGB inputs and the same two stage transfer learning protocol. In the first stage, the pretrained backbone was kept frozen and only the classification head was trained. In the second stage, the last 30 backbone layers were fine tuned with a lower learning rate. This common configuration enabled a controlled comparison between the two pretrained CNN backbones.

Table 5- Fine tuning and augmentation settings of the deep transfer learning models

Model	*Batch Size*	*Fine tuned Layers*	*Augmentation*	*Output Classes*
ResNet50	32	Last 30 layers	Rotation, shift, zoom, shear, horizontal flip	3
EfficientNetB0	32	Last 30 layers	Rotation, shift, zoom, shear, horizontal flip	3

Table 5 presents the fine tuning and augmentation settings used for the deep transfer learning models. Both ResNet50 and EfficientNetB0 used a batch size of 32, fine tuning was restricted to the last 30 layers, and the same augmentation operations were applied only to the training subset. The validation and test subsets were evaluated without augmentation. This setting supports a fair comparison between the two deep models while reducing the risk of overfitting.

3.6. Grad-CAM based interpretability analysis

To qualitatively inspect whether the deep models relied on meaningful visual evidence, Grad-CAM was applied as a post hoc interpretability method. Grad-CAM highlights regions of the input image that contribute most strongly to the predicted class by computing gradients with respect to the final convolutional representation. In the present study, Grad-CAM was used to verify whether the models attended primarily to visually meaningful hazelnut regions such as the body, surface texture, and local structural variation.

In the ResNet50 branch, Grad-CAM was computed using the final convolutional block of the network. In the EfficientNetB0 branch, Grad-CAM was computed using the top_conv layer, which was the final convolutional feature map used for heatmap generation. The resulting heatmaps were overlaid on representative test images for qualitative interpretation. These visualizations were interpreted as qualitative plausibility checks for model attention behavior rather than as causal evidence of model reasoning.

3.7. Deep embedding extraction

Because direct quantum processing of raw image pixels is impractical in the present setting, the hybrid branch was built on deep embeddings rather than on images themselves. The final EfficientNetB0 model, which achieved the strongest overall end to end performance, was repurposed as a fixed feature extractor. Embeddings were taken from the GAP layer, producing a 1280 dimensional representation for each sample. Separate embedding sets were exported for the training, validation, test, and train+validation partitions in order to support downstream reduced space experiments.

This design choice served two purposes. First, it allowed the hybrid branch to operate on semantically rich visual representations rather than on low level image features. Second, it created a unified latent feature space from which both classical and quantum compatible reduced space classifiers could be derived. In this sense, the hybrid experiments were structured not as fully independent learners from raw images, but as downstream classifiers operating on a common high quality deep representation. Images were first processed by the final EfficientNetB0 model, GAP embeddings were extracted, and the resulting 1280 dimensional feature vectors were transformed for full data reduced space SVC and fair subset SVC/QSVC experiments. The reduced space classical and quantum compatible classification pipeline is illustrated in Figure 3.

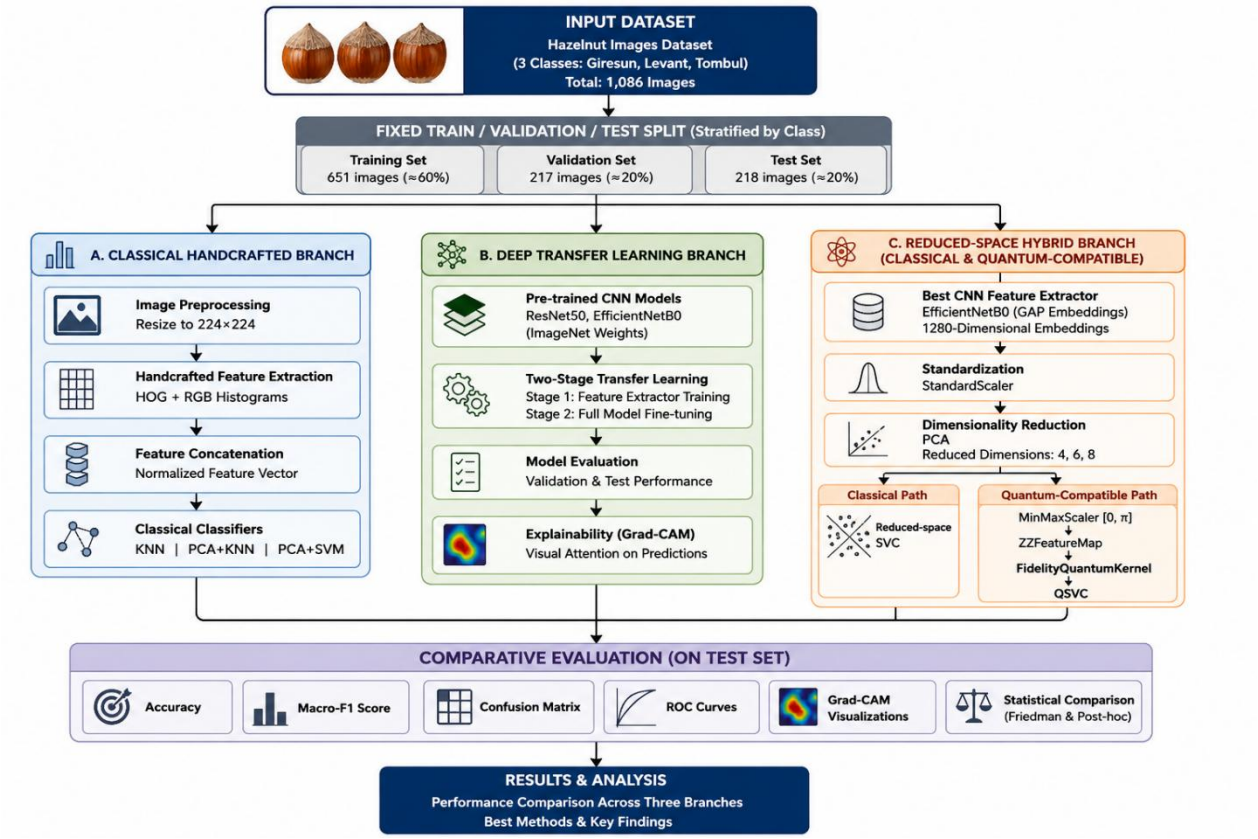


Figure 3- Reduced space classical and quantum compatible classification pipeline. EfficientNetB0 GAP embeddings were standardized, compressed into low dimensional PCA spaces, and evaluated using both classical SVC and QSVC under the same reduced space experimental setting

Figure 3 illustrates the reduced space hybrid classification branch used in the study. In this pipeline, the best performing EfficientNetB0 model was first used as a feature extractor, and 1280 dimensional GAP embeddings were obtained from the hazelnut images. These embeddings were then standardized and compressed into low dimensional PCA spaces in order to evaluate whether the learned deep representation retained discriminative class information after dimensionality reduction. The same reduced representation was used for classical SVC and QSVC, which enabled a controlled comparison between a classical margin based classifier and a quantum compatible kernel classifier. For the QSVC branch, the reduced features were additionally scaled to the interval $[0, \pi]$ before quantum feature mapping, making them suitable for angle based quantum encoding.

3.8. Reduced space classical and quantum compatible classifiers

The extracted EfficientNetB0 embeddings were transformed through a staged reduction pipeline. First, embeddings were standardized using StandardScaler. Second, PCA was applied to compress the embeddings into candidate reduced spaces of 4, 6, and 8 dimensions. For quantum compatible experiments, the reduced vectors were then mapped to the interval $[0, \pi]$ using MinMaxScaler, thereby producing a quantum compatible angle encoded representation for angle based encoding. The reduced dimensions of 4, 6, and 8 were selected to represent progressively richer but still computationally feasible quantum compatible feature spaces. The lower bound of four dimensions was selected because two dimensional encodings were considered too restrictive to represent three regional classes with visually overlapping shell texture and color cues. The upper bound of eight dimensions was selected because larger settings such as 10 or 12 dimensions would increase the number of qubits and the size of the quantum kernel computation, making the simulator based QSVC comparison substantially more expensive. Therefore, the 4-8 range was used as a practical compromise between representational capacity and computational feasibility. When PCA was fitted on the standardized EfficientNetB0 training embeddings, the first 4, 6, and 8 components explained 31.32%, 37.61%, and 42.68% of the cumulative variance, respectively. Although explained variance reflects the unsupervised variance structure rather than class separability, these values clarify how much of the 1280-dimensional EfficientNetB0 embedding variance was retained by the selected reduced spaces. In the QSVC setting, each reduced feature dimension corresponds to one qubit, therefore increasing the number of retained PCA components also increases the quantum circuit width and kernel computation cost. The MinMaxScaler transformation to the interval $[0, \pi]$ was applied because the reduced features were used as angle encoded inputs in the ZZFeatureMap. In this setting, each feature value parameterizes a rotation angle in the quantum circuit. Mapping all features to a bounded angular interval avoids unbounded or scale dependent rotations and ensures that all PCA components contribute within a comparable phase range. The upper bound π was selected as a standard half period angular range that provides sufficient phase variation for encoding while avoiding unnecessary wrapping effects associated with larger angular intervals. The complete transformation chain for the quantum compatible branch was therefore: EfficientNetB0 GAP embedding $\rightarrow$ StandardScaler $\rightarrow$ PCA $\rightarrow$ MinMaxScaler$(0, \pi)$ $\rightarrow$ QSVC. This configuration was explicitly recorded in the experiment metadata for the hybrid branch. Two related evaluation settings were used in the reduced space branch. In the first setting, a full data reduced space classical SVC was trained on the combined train+validation embeddings and evaluated on the full test set. This branch was intended to measure how much discriminative information remained in the deep embedding after severe compression, independent of quantum resource constraints. Hyperparameters for the reduced space classical SVC were selected by validation search over the reduced dimensions and standard SVC parameters.

In the second setting, a fair subset protocol was introduced to make quantum kernel evaluation computationally feasible while preserving class balance. Under this protocol, the hybrid comparison used 100 training, 40 validation, and 40 test samples per class. The same subset was used for both the classical SVC and the Quantum Support Vector Classifier (QSVC), ensuring that both models operated on identical training and testing conditions. This protocol yielded subset sizes of 300/120/120 for train, validation, and test, respectively. The fair subset protocol was used only for the direct SVC versus QSVC comparison. This restriction was necessary because quantum kernel computation scales with the number of samples and becomes substantially more expensive as the kernel matrix size increases. Using the same balanced subset for both SVC and QSVC ensured that the performance difference reflected classifier behavior under identical reduced space conditions rather than differences in sample composition or class imbalance.

For the classical fair subset baseline, an RBF-SVC was trained on the same quantum ready reduced representation used for the quantum branch. For the quantum classifier, the final search space considered ZZFeatureMap circuits with reduced dimensions of 4, 6, and 8, and repetition depths (reps) of 1 and 2. ZZFeatureMap was selected because it is a standard entangling feature map commonly used in QSVC style quantum kernel experiments and provides a controlled baseline for low dimensional angle encoded inputs. Its use allowed the study to keep the quantum branch computationally feasible while preserving a nonlinear quantum kernel structure. However, this choice also represents a methodological limitation. Alternative feature maps, such as PauliFeatureMap, may induce different kernel geometries and could affect classification performance. Therefore, comparison of ZZFeatureMap with alternative quantum feature maps should be considered in future work. The kernel was constructed using FidelityQuantumKernel, and the resulting kernel matrix was used within QSVC. Model selection was carried out on the validation subset, and the selected configuration was retrained on the combined train+validation subset before final evaluation on the held out fair test subset. In the final experimental design, the quantum comparison was restricted to kernel based QSVC, variational quantum classifiers were not included in the reported benchmark, so the fair quantum comparison reported here focused on the SVC versus QSVC contrast. All QSVC experiments were implemented using Qiskit Machine Learning and were executed in a simulator based environment rather than on real quantum hardware. In the implemented workflow, the quantum

kernel was constructed using FidelityQuantumKernel with ZZFeatureMap, and no finite shot backend, sampler, or explicit shots parameter was configured. Therefore, a hardware shot count is not applicable to the reported QSVC results. The results should be interpreted as simulator based quantum kernel classification results rather than hardware executed quantum experiments.

For QSVC, the number of qubits was equal to the reduced feature dimension. ZZFeatureMap was used with candidate dimensions of 4, 6, and 8 and repetition depths of 1 and 2. The quantum kernel matrix was computed using FidelityQuantumKernel and supplied to QSVC for classification. Model selection was performed on the balanced validation subset, and the selected configuration was refitted using the combined training and validation subset before final testing. Representation and preprocessing settings for reduced space classical and quantum compatible classifiers are detailed in Table 6. Training protocols and hyperparameter settings for reduced space classical and quantum compatible classifiers are detailed in Table 7.

Table 6- Representation and preprocessing settings for reduced space classical and quantum compatible classifiers

Branch	*Representation*	*Initial Dimension*	*Preprocessing*	*Reduced Dimensions*
Full data reduced space SVC	EfficientNetB0 GAP embedding	1280	StandardScaler + PCA	4 / 6 / 8
Fair subset SVC	EfficientNetB0 GAP embedding	1280	StandardScaler + PCA + MinMaxScaler(0, π)	4 / 6 / 8
QSVC	EfficientNetB0 GAP embedding	1280	StandardScaler + PCA + MinMaxScaler(0, π)	4 / 6 / 8

Table 6 summarizes the representation and preprocessing settings used for the reduced space classical and quantum compatible classifiers. The full data reduced space SVC branch used EfficientNetB0 GAP embeddings as the input representation and applied StandardScaler followed by PCA to obtain compact feature spaces for classical classification. In contrast, the fair subset SVC and QSVC branches were evaluated under the same reduced space protocol to ensure a controlled comparison between classical and quantum compatible classifiers. For the QSVC branch, the PCA reduced features were additionally transformed into the interval [0, π] using MinMaxScaler, since the quantum feature map requires angle encoded inputs. Therefore, the table clarifies that QSVC was not applied directly to raw images, but to standardized, PCA compressed, and quantum ready EfficientNetB0 embeddings.

Table 7- Training protocols and hyperparameter settings for reduced space classical and quantum compatible classifiers

Branch	*Training Protocol*	*Classifier*	*Key Hyperparameters*	*Evaluation Scope*
Full data reduced space SVC	Train+Validation → Test	SVC	kernel, C, gamma	Full test set
Fair subset SVC	100 train / 40 val / 40 test per class	SVC	C, gamma	Balanced fair subset
QSVC	100 train / 40 val / 40 test per class	QSVC	reps = 1 / 2, quantum kernel	Balanced fair subset

Table 7 presents the training and hyperparameter settings used in the reduced space SVC and QSVC experiments. The full data SVC experiment was performed to evaluate the discriminative strength of PCA reduced EfficientNetB0 embeddings on the complete test set, whereas the fair subset SVC and QSVC experiments were conducted on identical balanced subsets to ensure a fair classical versus quantum compatible comparison. For the classical SVC models, candidate configurations were selected according to validation performance using predefined C, gamma, and PCA dimension settings. For QSVC, the experiments considered PCA reduced feature dimensions of 4, 6, and 8 together with ZZFeatureMap repetition depths of 1 and 2. Since QSVC was implemented in a simulator based Qiskit workflow without an explicit finite shot backend, no hardware shot count was applicable. Overall, the table clarifies that the reduced space comparison was controlled by using the same embedding source, comparable dimensionality reduction settings, and validation guided model selection.

3.9. Evaluation criteria and reporting

Model performance was assessed using accuracy, macro precision, macro recall, macro-F1, and weighted F1. For the deep and classical main branches, confusion matrices and ROC based summaries were also generated to inspect classwise discrimination behavior. In the hybrid branch, confusion matrices and classwise classification reports were used to compare the classical and quantum compatible models under both the full data reduced space setting and the fair subset protocol. Aggregated result tables were then constructed in order to compare the strongest model from each methodological family under a common reporting framework.

The experiments were implemented in Google Colab using Python. Deep learning models were developed with TensorFlow/Keras (Abadi et al. 2016) and PyTorch (Paszke et al. 2019), classical baselines were implemented with scikit learn (Pedregosa et al. 2011), and quantum compatible classifiers were implemented with Qiskit Machine Learning (Sahin et al. 2025).

To improve reproducibility, a fixed random seed of 42 was used for data partitioning, fair subset sampling, and model related procedures wherever applicable. This seed value was used consistently in the main experimental workflow to support reproducible train/validation/test splitting, subsampling, and evaluation. Intermediate artifacts, including dataset splits, handcrafted features, deep embeddings, trained model checkpoints, and evaluation outputs, were stored in Google Drive so that the workflow could be resumed reliably under Colab runtime interruptions. The figures used in this manuscript were generated using Python, primarily with the Matplotlib library (Hunter 2007), in the Google Colab environment. Since the experiments were conducted using a single fixed train/validation/test split, the reported metrics should be interpreted as split specific benchmark results.

The experiments were executed in Google Colab using an NVIDIA L4 GPU environment with approximately 53 GB of system RAM. The software environment included Python, TensorFlow/Keras, scikit-learn, Matplotlib, and Qiskit Machine Learning. Because Colab resources may vary across sessions, the reported computational setting should be interpreted as the runtime environment used for the final rerun experiments.

4. Results and Discussion

4.1. Overall comparative performance

The aggregated benchmark reveals a clear hierarchy across the evaluated model families. Under the common full test protocol, EfficientNetB0 produced the strongest overall result, achieving 1.0000 accuracy and 1.0000 macro-F1. This result should be interpreted within the fixed image level split protocol and should not be read as definitive evidence of object level generalization. The next best result was obtained by the reduced space classical SVC trained on PCA compressed EfficientNetB0 embeddings, which reached 0.9963 accuracy and 0.9961 macro-F1. ResNet50 followed closely with 0.9945 accuracy and 0.9945 macro-F1, while the strongest handcrafted baseline, PCA + SVM, achieved 0.9890 accuracy and 0.9888 macro-F1. This ranking indicates that all three methodological branches were highly competitive under the present protocol, but that learned deep representations provided the clearest advantage in terms of separability and robustness.

From a comparative standpoint, the main result is not merely that deep learning performed best, but that the distance between the best deep and the best non deep alternatives became progressively smaller as representation quality improved. The classical handcrafted branch remained strong after compression, and the reduced space SVC nearly matched the end to end deep models once it operated on high quality EfficientNetB0 embeddings. This suggests that classifier performance was strongly influenced by the separability of the feature space used by each model family.

4.2. Classical handcrafted baselines

The handcrafted branch provides a useful reference for understanding how much information can be extracted without deep representation learning. The baseline Raw KNN model achieved 0.9560 accuracy and 0.9535 macro-F1, indicating that the concatenated HOG + RGB histogram descriptor captured substantial regional class structure. After dimensionality reduction, PCA + KNN improved performance to 0.9817 accuracy and 0.9808 macro-F1. The strongest classical configuration, PCA + SVM, further improved the result to 0.9890 accuracy and 0.9888 macro-F1.

This progressive improvement is methodologically important. It shows that the handcrafted feature space was not inherently weak, but rather redundant and geometrically inefficient in its raw form. Once redundancy was suppressed through PCA and the decision boundary was learned by a stronger classifier, the handcrafted representation became substantially more competitive. In other words, the classical branch demonstrates that regional hazelnut categories can be separated to a large extent using explicit structure and color descriptors, but that their full discriminative potential emerges only after representation compression and appropriate classifier selection.

A second implication is that handcrafted methods remain scientifically relevant in this problem setting. Their performance is lower than that of the best deep models, but still high enough to show that the task is not dependent exclusively on end to end deep learning. This strengthens the comparative value of the study, since the deep branches are not being contrasted against trivially weak baselines, but against a carefully optimized classical pipeline. The performance results of the three handcrafted baselines are presented in Table 8.

Table 8- Performance comparison of handcrafted classical baselines

Model	***Accuracy***	***Macro-F1***	***Main interpretation***
Raw KNN	0.9560	0.9535	Handcrafted space is informative but redundant
PCA + KNN	0.9817	0.9808	PCA improves compactness
PCA + SVM	0.9890	0.9888	Best handcrafted baseline

Table 8 summarizes the performance of the handcrafted classical classification branch. The Raw KNN baseline achieved competitive results, indicating that the combined HOG and RGB histogram representation already contained useful class

discriminative information. However, the PCA based configurations improved performance, suggesting that dimensionality reduction helped remove redundant or noisy components from the high dimensional handcrafted feature vector. The best classical result was obtained with PCA + SVM, which achieved 0.9890 accuracy and 0.9888 macro-F1, outperforming both Raw KNN and PCA + KNN. This finding shows that handcrafted descriptors remain valuable for hazelnut image classification when combined with appropriate feature compression and classifier selection.

4.3. Deep transfer learning performance

Both deep transfer learning models produced near ceiling performance, confirming that the regional hazelnut classes are highly separable in a modern pretrained feature space. ResNet50 achieved 0.9945 accuracy, 0.9947 macro precision, 0.9942 macro recall, and 0.9945 macro-F1 on the test set. The classwise AUC values were also extremely high, approximately 0.9994 for Van, 1.0000 for Ordu, and 0.99995 for Giresun. These values indicate that the model not only performed accurately at the decision threshold, but also maintained highly reliable ranking behavior across classes.

The strongest end to end result was obtained by EfficientNetB0, which achieved perfect classification on the held out test set under the present image level split protocol, including 1.0000 accuracy, 1.0000 macro-F1, and 1.0000 AUC for all three classes. Under the present evaluation setting, this places EfficientNetB0 at the top of the model hierarchy. The result is especially notable because it was obtained after a two stage transfer learning procedure rather than through a large bespoke architecture or a highly customized training schedule. This suggests that the pretrained EfficientNetB0 backbone provides a particularly suitable latent representation for the regional hazelnut categories used in this study.

The training curves of both deep models further support the reliability of the learned representations. In both branches, Stage 1 rapidly achieved very high validation performance, and Stage 2 fine tuning produced either further improvement or stable near optimal behavior. In the ResNet50 branch, the fine tuning stage converged to a best validation accuracy of approximately 0.9963, whereas the EfficientNetB0 branch also reached approximately 0.9963 validation accuracy before producing perfect test classification. These training dynamics suggest that both models benefited from controlled fine tuning without destabilizing their pretrained visual structure.

4.4. Interpretability analysis of deep models

The Grad-CAM visualizations provide qualitative support for the plausibility of the deep learning results. In both the ResNet50 and EfficientNetB0 branches, the heatmaps are concentrated primarily on the hazelnut body and visually distinctive surface regions rather than only on the empty black background. This suggests that the models are not predominantly driven by visually obvious background regions in the inspected examples, but are instead using semantically relevant object evidence. The EfficientNetB0 Grad-CAM example, in particular, suggests a sharper focus on localized surface structure and boundary regions, which is consistent with its perfect classification result under the present protocol.

From a methodological standpoint, these results strengthen the interpretability of the deep branch, but they should still be read with appropriate caution. Grad-CAM is a post hoc explanatory tool and does not prove causal feature usage in a strict sense. Nevertheless, in agricultural image analysis, such visualizations remain valuable because they provide qualitative reassurance that very high predictive accuracy is not being achieved through obviously implausible attention patterns.

Grad-CAM visualizations are presented in Figure 4. The heatmaps suggest that both ResNet50 and EfficientNetB0 focus primarily on the hazelnut body, including surface texture, boundary regions, and visually distinctive local structures. In the inspected examples, the models do not appear to be driven predominantly by the empty background. This supports the plausibility of the deep learning results by showing that the highlighted regions correspond to semantically meaningful parts of the hazelnut. For each class, the original test image is shown together with the Grad-CAM heatmaps generated by ResNet50 and EfficientNetB0. The visualizations illustrate the image regions that contributed most strongly to the class predictions of the deep learning models.

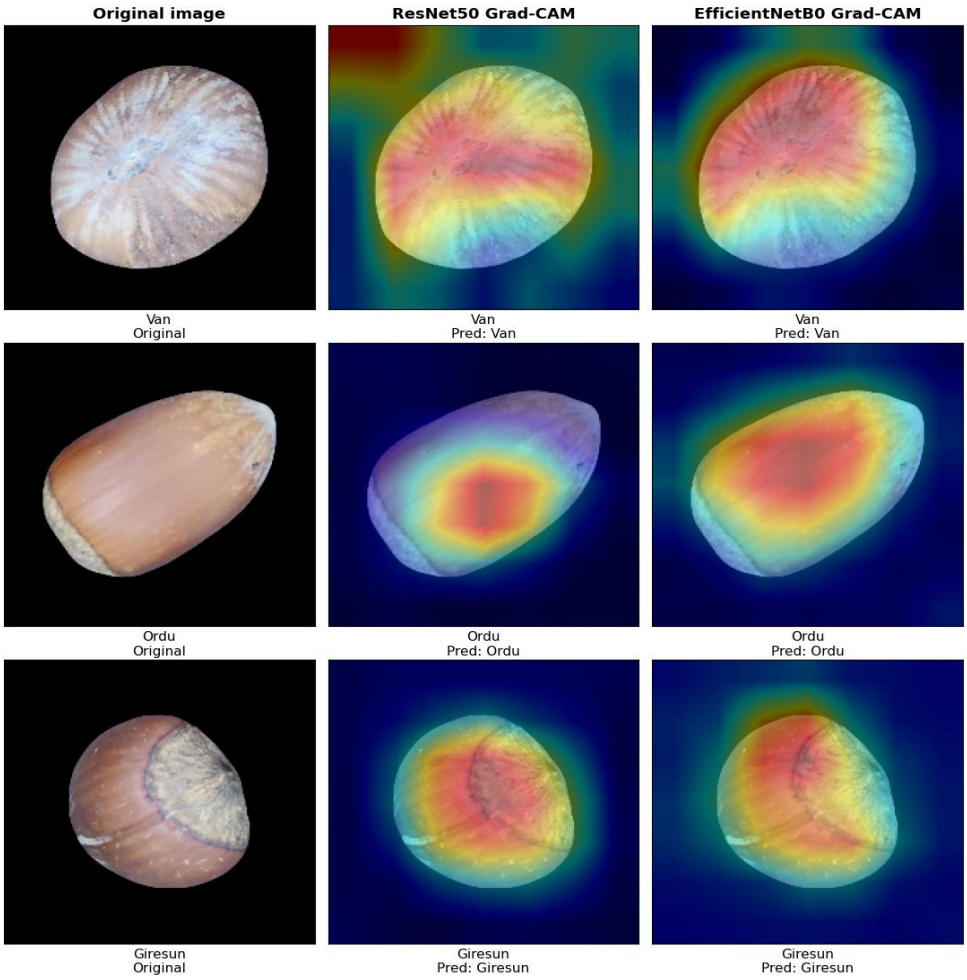


Figure 4- Grad-CAM visualizations for ResNet50 and EfficientNetB0 across test images from the Van, Ordu, and Giresun classes. The displayed samples were selected from the test set using a fixed random seed and were confirmed to be correctly classified by both models. Therefore, the figure is intended to illustrate representative attention behavior in correctly classified cases rather than to provide an exhaustive analysis of all possible model attention patterns

Figure 4 presents Grad-CAM visualizations for selected correctly classified test samples from the Van, Ordu, and Giresun classes. For each class, the original image is shown together with the attention maps generated by ResNet50 and EfficientNetB0. Across the three categories, both models concentrated their activations primarily on the hazelnut body and visually distinctive surface regions, rather than relying only on the background. This suggests that the predictions were supported by object relevant cues such as shell texture, local structure, and surface appearance. In several cases, EfficientNetB0 produced more localized and concentrated activation patterns, whereas ResNet50 showed broader attention over the hazelnut region. These visual explanations provide qualitative support for the plausibility of the deep learning predictions, although Grad-CAM should be interpreted as an explanatory aid rather than definitive causal evidence of model reasoning.

The heatmaps indicate that both models mainly attend to the hazelnut body and visually distinctive surface regions. The visualizations are provided as qualitative plausibility checks rather than causal explanations.

4.5. Reduced space classification on deep embeddings

The reduced space experiments based on EfficientNetB0 embeddings yielded one of the most informative outcomes of the study. The final embedding source was the GAP layer of the best EfficientNetB0 model, producing 1280 dimensional feature vectors. These embeddings were then standardized and compressed through PCA before classification. On the full data setting, the strongest reduced space classical model was an RBF-SVC trained on a 6 dimensional PCA compressed embedding space, with C=1and γ="scale". This model achieved 0.9963 accuracy and 0.9961 macro-F1, thereby nearly matching the strongest end to end deep models.

This result has two important implications. First, it shows that the EfficientNetB0 latent space remains highly informative even after extreme dimensional compression. Reducing a 1280 dimensional embedding to only six retained components while preserving near ceiling classification performance indicates that the learned representation is not only powerful, but also compact in a discriminative sense. Second, it suggests that the principal performance gain of the deep branch arises from representation

quality more than from the final softmax classifier itself. Once the EfficientNetB0 latent structure has been learned, even a relatively simple kernel classifier can perform exceptionally well in the resulting compressed space.

The full data validation search also provides a useful technical insight. Across reduced dimensions, γ="scale" consistently outperformed γ="auto", in some cases by a large margin. This suggests that data adaptive kernel width estimation was considerably better aligned with the geometry of the compressed deep latent space than the more naive fixed feature count scaling implied by auto. In other words, the compressed embedding retained class structure that was best exploited when the kernel bandwidth adapted to the actual variance of the projected data.

4.6. Fair subset comparison between SVC and QSVC

To make quantum evaluation computationally feasible while preserving a fair comparison, the study introduced a balanced subset protocol in which both the classical and quantum models were trained and evaluated on exactly the same reduced space data. In this setting, the original EfficientNetB0 embeddings were transformed through the sequence StandardScaler → PCA → MinMaxScaler(0, π), and balanced subsets of 100 training, 40 validation, and 40 test samples per class were used. This yielded train, validation, and test subset sizes of 300, 120, and 120, respectively, with the final refit performed on the combined train+validation subset.

Under this fair subset protocol, the best classical baseline was SVC with 4 retained dimensions, C=0.1, and γ="scale". This model achieved 0.9833 accuracy and 0.9832 macro-F1. The strongest quantum result was obtained by QSVC with 4 dimensions and reps = 1, which achieved 0.9083 accuracy and 0.9082 macro-F1.

The classwise results indicate that the remaining errors in the fair subset experiments were not uniformly distributed across classes. In particular, the fair subset SVC achieved classwise F1 scores of 0.99 for Van, 0.99 for Ordu, and 0.97 for Giresun, while QSVC achieved 0.93, 0.92, and 0.88 for Van, Ordu, and Giresun, respectively. Therefore, the interpretation that Giresun related boundaries were more sensitive in the reduced space setting is supported by the classwise F1 scores and the corresponding confusion matrices, rather than being based only on visual speculation.

A qualitative inspection of the representative images also suggests that the exposed cap region, shell tone, and local surface texture may overlap visually across some samples, which can make the reduced space boundaries more sensitive. This pattern is consistent with previous observations in agricultural image analysis where surface texture variations within a single cultivar can create overlapping feature distributions after aggressive dimensionality reduction (Taner et al. 2021). However, this interpretation should be regarded as a qualitative explanation of the observed error pattern rather than as an independent biological or geographical claim.

This comparison allows a more nuanced interpretation than a simple “classical versus quantum” ranking. In the same reduced space and balanced subset setting, classical SVC was more effective and more stable than QSVC. However, the QSVC result remains informative because it shows how a quantum compatible kernel classifier behaves when applied to severely compressed deep embeddings. The lower QSVC performance suggests that, in the present setting, the quantum compatible kernel did not provide an advantage over the classical SVC baseline. This outcome may be related to the restricted fair subset size, the low dimensional angle encoded representation, and the use of a single ZZFeatureMap configuration. The classification performance of the models is further analyzed using confusion matrices (Figures 5-7). A fair comparison between SVC and QSVC is presented in Figures 8 and 9.

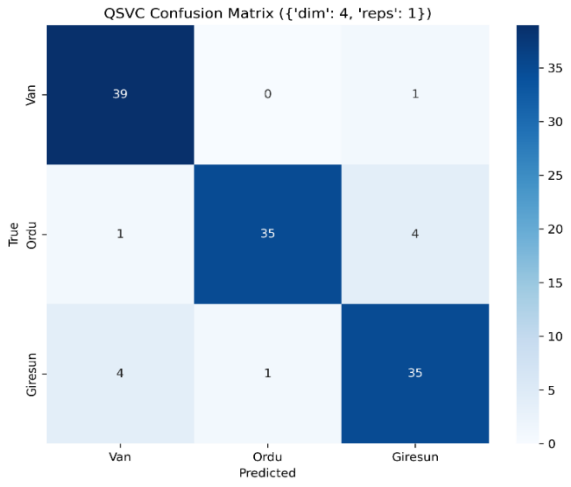


Figure 5- Confusion matrix of the best QSVC model under the fair subset protocol

Figure 5 shows the classwise prediction behavior of the best QSVC model on the balanced fair test subset. The model correctly classified 39 of 40 Van samples, indicating strong separability for this class in the reduced quantum compatible feature space. For Ordu and Giresun, 35 of 40 samples were correctly classified in each class, while the main errors occurred between Ordu and Giresun and between Giresun and Van. Overall, the diagonal dominance of the matrix indicates that QSVC learned a meaningful class structure, but the off diagonal errors also show that the reduced space quantum compatible classifier was more sensitive to visually overlapping regional patterns than the classical SVC baseline.

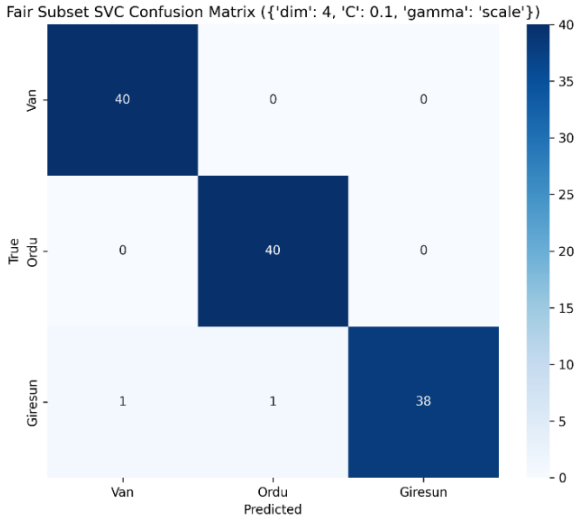


Figure 6- Confusion matrix of the fair subset SVC baseline under the same reduced space protocol used for QSVC comparison

Figure 6 shows the classwise performance of the classical SVC baseline under the fair subset protocol. The model correctly classified all Van and Ordu samples, with 40 out of 40 correct predictions for each class. For the Giresun class, 38 of 40 samples were correctly classified, while one sample was misclassified as Van and one as Ordu. This nearly perfect diagonal structure indicates that the PCA reduced EfficientNetB0 embeddings retained highly discriminative class information even in a very low dimensional space. Compared with the QSVC confusion matrix, the SVC baseline produced fewer classwise errors, suggesting that the classical margin based classifier was more robust than QSVC under the same reduced space and balanced subset conditions.

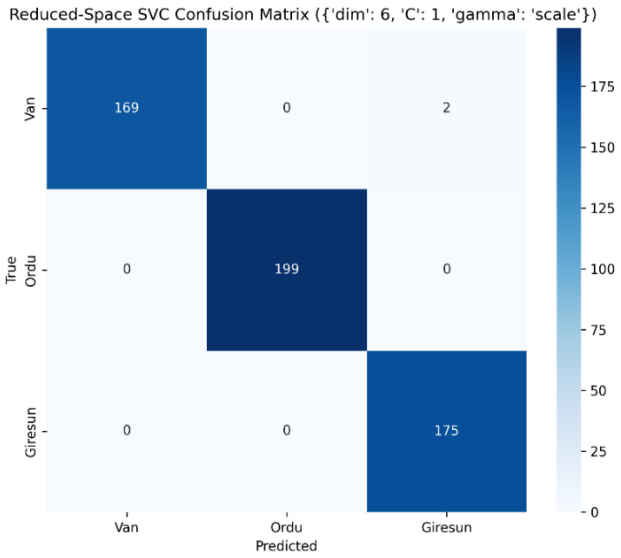


Figure 7- Confusion matrix of the full data reduced space SVC model

Figure 7 shows the classwise prediction behavior of the reduced space SVC model evaluated on the full subset protocol. The model correctly classified all Ordu and Giresun samples, with 199/199 and 175/175 correct predictions, respectively. For the Van class, 169 of 171 samples were correctly classified, while only two Van samples were misclassified as Giresun. This near perfect diagonal structure indicates that the 6 dimensional PCA reduced EfficientNetB0 embeddings preserved highly

discriminative class information despite substantial dimensionality reduction. The result supports the interpretation that the learned deep representation, rather than classifier complexity alone, played a central role in the strong classification performance.

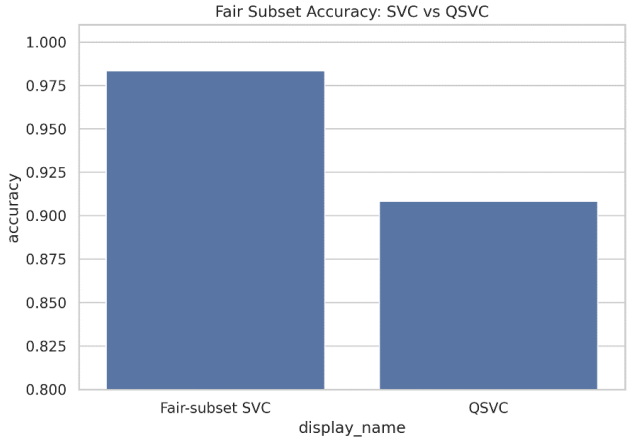


Figure 8- Accuracy comparison between fair subset SVC and QSVC under the same reduced space protocol

Figure 8 compares the test accuracy of the classical SVC baseline and the QSVC model under the same fair subset protocol. The fair subset SVC achieved higher accuracy than QSVC, indicating that the classical margin based classifier was more effective in the reduced EfficientNetB0 embedding space. This result suggests that although QSVC was able to learn meaningful class structure, its performance remained below that of the classical SVC baseline under the current simulator based quantum compatible setting. Therefore, the advantage in this experiment appears to come primarily from the discriminative quality of the deep embedding representation rather than from the quantum compatible classifier itself.

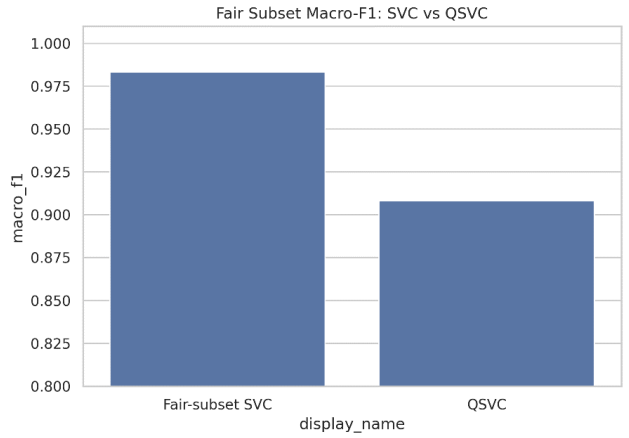


Figure 9- Macro-F1 comparison between fair subset SVC and QSVC under the same reduced space protocol

Figure 9 compares the macro-F1 scores of the fair subset SVC and QSVC models. Macro-F1 was used to evaluate class balanced performance because it gives equal importance to each regional class regardless of class frequency. The fair subset SVC achieved a higher macro-F1 score than QSVC, indicating that it provided more consistent classwise performance under the same reduced space embedding conditions. Although QSVC produced a meaningful classification result, the lower macro-F1 score suggests that it was more affected by classwise confusion, particularly between visually overlapping regional categories. This finding supports the conclusion that, in the present setting, PCA compressed EfficientNetB0 embeddings were more effectively exploited by the classical SVC than by the simulator based QSVC model.

4.7. Statistical robustness and representation quality analysis

To assess whether the observed performance differences were robust to sampling variability, a bootstrap analysis was conducted using the test set predictions. Accuracy and macro-F1 were resampled over 1000 bootstrap iterations, and 95% confidence intervals were computed for the evaluated models. This analysis provides an estimate of performance stability under the fixed test split and complements the point estimates reported in the main comparison tables. The resulting bootstrap confidence intervals are reported in Table 9.

Table 9- Bootstrap confidence intervals for accuracy and macro-F1 across the evaluated models

Model	*Test protocol*	*n*	*Accuracy*	*95% CI for accuracy*	*Macro-F1*	*95% CI for macro-F1*
EfficientNetB0	Full test set	545	1.0000	1.0000–1.0000	1.0000	1.0000–1.0000
Reduced space SVC	Full test set	545	0.9963	0.9908–1.0000	0.9961	0.9904–1.0000
ResNet50	Full test set	545	0.9945	0.9872–1.0000	0.9945	0.9870–1.0000
PCA + SVM	Full test set	545	0.9890	0.9798–0.9963	0.9888	0.9797–0.9963
PCA + KNN	Full test set	545	0.9817	0.9688–0.9927	0.9808	0.9673–0.9920
Raw KNN	Full test set	545	0.9560	0.9358–0.9725	0.9535	0.9329–0.9704
Fair subset SVC	Balanced fair subset	120	0.9833	0.9583–1.0000	0.9832	0.9557–1.0000
QSVC	Balanced fair subset	120	0.9083	0.8581–0.9502	0.9082	0.8518–0.9518

Note: Full test results were obtained under an image level split. Because physical object identifiers were not available, the possibility that visually similar or near duplicate images of the same hazelnut object appear in both training and test subsets cannot be excluded. Therefore, near ceiling performance should be interpreted as within dataset benchmark performance rather than definitive object level generalization.

Table 9 reports the 95% bootstrap confidence intervals obtained from 1000 resampling iterations on the test predictions. The full test models were evaluated on 545 test images, whereas fair subset SVC and QSVC were evaluated on the balanced fair test subset consisting of 120 samples. EfficientNetB0, reduced space SVC, and ResNet50 produced very narrow confidence intervals, indicating stable performance under the fixed test split. The handcrafted baselines showed wider intervals, especially Raw KNN, suggesting greater sensitivity to sample variation. In the fair subset comparison, classical SVC achieved higher and more stable performance than QSVC, while the wider QSVC interval reflects its lower accuracy and greater classwise sensitivity under the reduced space quantum compatible setting. Representation type, effective dimensionality, and performance across the main model families are presented in Table 10.

Table 10- Representation type, effective dimensionality, and performance across the main model families

Model	*Representation type*	*Rep. level*	*Classifier*	*Effective dimension*	*Accuracy*	*Macro-F1*
Raw KNN	Raw handcrafted HOG + RGB	1	KNN	40868	0.9560	0.9535
PCA + KNN	PCA compressed handcrafted HOG + RGB	2	KNN	200	0.9817	0.9808
PCA + SVM	PCA compressed handcrafted HOG + RGB	2	SVM	200	0.9890	0.9888
ResNet50	End to end deep representation	3	CNN softmax head	2048	0.9945	0.9945
EfficientNetB0	End to end deep representation	3	CNN softmax head	1280	1.0000	1.0000
Reduced space SVC	PCA compressed EfficientNetB0 embeddings	4	SVC	6	0.9963	0.9961

Table 10 summarizes the relationship between representation type, effective dimensionality, and classification performance. The results show a consistent improvement from raw handcrafted descriptors to PCA compressed handcrafted features and then to deep learned representations. The reduced space SVC result is particularly important because it achieved 0.9963 accuracy and 0.9961 macro-F1 using only six PCA components extracted from EfficientNetB0 embeddings. This indicates that the EfficientNetB0 embedding space retained highly separable class structure even after severe dimensionality reduction. Therefore, the findings support the interpretation that representation quality and compressibility were major determinants of performance in the present classification task. The Spearman rank correlation results are presented in Table 11.

Table 11- Descriptive rank based correlation between representation level and model performance

Metric	*Spearman's ρ*	*p-value*
Accuracy	0.8827	0.0198
Macro-F1	0.8827	0.0198

Table 11 presents the descriptive rank based correlation between representation level and model performance. Spearman's correlation showed a positive association between representation level and both accuracy and macro-F1. This supports the interpretation that stronger and more compact representations were associated with higher classification performance. However,

because the number of compared model families was limited, this correlation should be interpreted descriptively rather than as a large sample inferential test.

Overall, the bootstrap and representation level analyses strengthen the main comparative interpretation of the study. The bootstrap confidence intervals show that the strongest full test models remained stable under resampling, while the representation analysis shows that performance increased as the feature representation became more informative and compact. These findings provide additional support for the comparative interpretation of the study, while remaining limited to the fixed image level split protocol.

4.8. General interpretation and study limitations

4.8.1. Main comparative findings

The overall comparison shows that all methodological branches achieved strong performance under the fixed image level split protocol, but their strengths differed. EfficientNetB0 produced the highest full test result, followed closely by reduced space SVC and ResNet50. The handcrafted PCA + SVM baseline also remained competitive, indicating that explicit HOG + RGB descriptors retained substantial class discriminative information when combined with dimensionality reduction and a margin based classifier. Overall, the comparison indicates that deep and reduced space representations provided the strongest class separation, while optimized handcrafted descriptors remained valuable baselines.

4.8.2. Why EfficientNetB0 and reduced space SVC performed strongly

The strong performance of EfficientNetB0 can be attributed to its ability to learn compact and highly discriminative visual representations from hazelnut appearance cues such as shell texture, surface structure, shape, and color distribution. The reduced space SVC result further supports this interpretation. Although the original EfficientNetB0 GAP embedding had 1280 dimensions, the reduced space SVC achieved near ceiling performance using only six PCA components. This indicates that the EfficientNetB0 embedding space preserved class separable structure even after severe dimensionality reduction. In practical terms, the final classifier did not need to be highly complex once the learned representation was sufficiently informative.

4.8.3. Why QSVC remained behind classical SVC

The fair subset comparison showed that QSVC remained below the classical SVC baseline under the same quantum ready reduced space protocol. This does not necessarily imply that quantum learning is unsuitable for agricultural image classification. Rather, it suggests that the current QSVC setting was constrained by the low dimensional encoding, limited fair subset size, and simulator based quantum kernel evaluation. The classical SVC was able to exploit the compressed EfficientNetB0 embedding space more effectively, whereas QSVC appeared more sensitive to classwise overlap, especially around the Giresun related decision boundaries. Therefore, the quantum compatible branch should be interpreted as an exploratory reduced space comparison rather than as a practical replacement for the strongest classical or deep models.

4.8.4. Limitations

Several limitations should be considered when interpreting the results. First, the experiments were conducted under an image level split rather than a strict object level split because the available filename structure did not provide reliable physical object identifiers. Most importantly, because the dataset does not provide physical object identifiers, visually similar or near duplicate images of the same hazelnut object may have been distributed across the training and test subsets. This possibility cannot be ruled out and may have partially contributed to the near ceiling performance observed for EfficientNetB0 and reduced space SVC. Therefore, the reported near ceiling performance values should be interpreted as strong within dataset benchmark results rather than definitive evidence of object level or deployment level generalization. Second, the benchmark was based on a single fixed train/validation/test partition. Although this design ensured controlled comparison across model families, repeated multi seed evaluation or cross validation would provide a more robust estimate of split level variability. Third, the Van category was retained because it belongs to the original dataset defined three class structure, however, the study does not claim that these three classes fully represent the geographical diversity of hazelnut production. Although the Van category was retained because it is part of the original dataset defined class structure, future studies should apply stricter dataset selection criteria and include regions that are directly aligned with the intended geographical or production related research question. Fourth, the QSVC experiments were simulator based and did not use real quantum hardware or an explicit finite shot backend. Future studies should therefore validate the findings using object level partitioning, independently collected samples, additional regions, repeated splits, and alternative quantum feature maps.

5. Conclusions

This study presented a controlled comparative benchmark for regional hazelnut image classification using three dataset defined categories corresponding to Van, Ordu, and Giresun. By evaluating handcrafted descriptors, deep transfer learning models,

reduced space classical classifiers, and quantum compatible methods under a unified protocol, the study examined how feature representation, classifier family, and dimensional compression interact in an agricultural image classification setting.

Among the evaluated full test models, EfficientNetB0 achieved the strongest performance, reaching 100% accuracy and 1.0000 macro-F1 under the present image level split protocol. ResNet50 also produced near ceiling performance, while the strongest handcrafted baseline, PCA + SVM, remained highly competitive. The reduced space SVC result was especially informative because a classical SVC trained on only six PCA compressed EfficientNetB0 embedding components still achieved 99.63% accuracy and 0.9961 macro-F1. The bootstrap confidence intervals and descriptive representation level correlation analysis further supported the stability of the main comparative ranking and the interpretation that representation quality was a central determinant of performance.

The quantum compatible branch led to a more cautious conclusion. Under the fair subset protocol, classical SVC achieved 98.33% accuracy, whereas the best QSVC configuration reached 90.83% accuracy. This indicates that, in the current simulator based and low dimensional encoding setting, QSVC did not outperform the classical SVC baseline. Nevertheless, the QSVC experiment remains useful as an exploratory comparison that helps clarify the current limitations of quantum-compatible classification when applied to compressed agricultural image embeddings.

The findings should be interpreted within the limitations of the study design. The experiments used an image level split, a single fixed partition, dataset defined regional classes, and simulator based QSVC experiments. Therefore, the reported near ceiling performance should be viewed as a strong within dataset benchmark rather than definitive evidence of object level or deployment level generalization. Future work should evaluate independently collected samples, stricter object level splits, additional regions, repeated multi seed protocols, and alternative quantum feature maps or hardware aware quantum learning settings.

Author Declarations

Authors' contributions: Study conception and design: ÖT; data collection: ÖT; analysis and interpretation of results: ÖT; draft manuscript preparation: ÖT. The author reviewed and approved the final version of the manuscript.

Conflict of interest/Competing interests: The author declares no conflict of interest.

Ethics approval/declarations: No need/ Not applicable.

Funding: This research received no external funding.

Consent to participate: No need/ Not applicable.

Consent for publication: No need/ Not applicable.

Data availability: Data are available on request due to privacy or other restrictions.

References

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J & Zheng X (2016). TensorFlow: A system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) (pp. 265-283). USENIX Association.

Alam M, Kundu S, Topaloglu R O & Ghosh S (2021). Quantum-classical hybrid machine learning for image classification (ICCAD special session paper). In 2021 IEEE/ACM International Conference on Computer-Aided Design (ICCAD) (pp. 1-7). IEEE. https://doi.org/10.1109/ICCAD51958.2021.9643516

Dalal N & Triggs B (2005). Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE. https://doi.org/10.1109/CVPR.2005.177

Dönmez E, Kılıçarslan S & Diker A (2024). Classification of hazelnut varieties based on BigTransfer deep learning model. European Food Research and Technology 250: 1433-1442. https://doi.org/10.1007/s00217-024-04468-1

Geirhos R, Jacobsen J H, Michaelis C, Zemel R, Brendel W, Bethge M & Wichmann F A (2020). Shortcut learning in deep neural networks. Nature Machine Intelligence 2(11): 665-673. https://doi.org/10.1038/s42256-020-00257-z

Gencturk B, Arsoy S, Taspinar Y S, Cinar I, Kursun R, Yasin E T & Koklu M (2024). Detection of hazelnut varieties and development of mobile application with CNN data fusion feature reduction-based models. European Food Research and Technology 250: 97-110. https://doi.org/10.1007/s00217-023-04369-9

Georgiou T, Liu Y, Chen W & Lew M S (2020). A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision. International Journal of Multimedia Information Retrieval 9(3): 135-170. https://doi.org/10.1007/s13735-019-00183-w

Giraudo A, Calvini R, Orlandi G, Ulrici A, Geobaldo F & Savorani F (2018). Development of an automated method for the identification of defective hazelnuts based on RGB image analysis and colourgrams. Food Control 94: 233-240. https://doi.org/10.1016/j.foodcont.2018.07.018

Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M & Gambetta J M (2019). Supervised learning with quantum-enhanced feature spaces. Nature 567(7747): 209-212. https://doi.org/10.1038/s41586-019-0980-2

He K, Zhang X, Ren S & Sun J (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

Hossen M I, Awrangjeb M, Pan S & Mamun A A (2025). Transfer learning in agriculture: A review. Artificial Intelligence Review 58: Article 97. https://doi.org/10.1007/s10462-024-11081-x

Hu M K (1962). Visual pattern recognition by moment invariants. IRE Transactions on Information Theory 8(2): 179-187. https://doi.org/10.1109/TIT.1962.1057692

Hunter J D (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering 9(3): 90-95. https://doi.org/10.1109/MCSE.2007.55

Köklü M (2024). Hazelnut Images Dataset [online]. Website https://www.muratkoklu.com/datasets/Hazelnut_Images_Dataset.zip [accessed 5 March 2026].

Liao B, Wang Y, Li X, Cheng Y, Zhao G & Zhou Y (2026). Deep learning-based image recognition for food science and technology: End-to-end workflows and domain-specific solutions. Comprehensive Reviews in Food Science and Food Safety 25(1): Article e70388. https://doi.org/10.1111/1541-4337.70388

Lowe D G (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2): 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Mari A, Bromley T R, Izaac J, Schuld M & Killoran N (2020). Transfer learning in hybrid classical-quantum neural networks. Quantum 4: Article 340. https://doi.org/10.22331/q-2020-10-09-340

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, ... & Chintala S (2019). PyTorch: An imperative style, high-performance deep learning library. In H Wallach, H Larochelle, A Beygelzimer, F d'Alché-Buc, E Fox & R Garnett (Eds.), Advances in Neural Information Processing Systems 32 (pp. 8024-8035). Curran Associates, Inc.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M & Duchesnay É (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12: 2825-2830.

Sahin M E, Altamura E, Wallis O, Wood S P, Dekusar A, Millar D A, Imamichi T, Matsuo A & Mensa S (2025). Qiskit Machine Learning: An open-source library for quantum machine learning tasks at scale on quantum hardware and classical simulators. arXiv. https://doi.org/10.48550/arXiv.2505.17756

Schuld M & Killoran N (2019). Quantum machine learning in feature Hilbert spaces. Physical Review Letters 122(4): Article 040504. https://doi.org/10.1103/PhysRevLett.122.040504

Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D & Batra D (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 618-626). IEEE. https://doi.org/10.1109/ICCV.2017.74

Senokosov A, Sedykh A, Sagingalieva A, Kyriacou B & Melnikov A (2024). Quantum machine learning for image classification. Machine Learning: Science and Technology 5(1): Article 015040. https://doi.org/10.1088/2632-2153/ad2aef

Silvestri C, Bacchetta L, Bellincontro A & Cristofori V (2021). Advances in cultivar choice, hazelnut orchard management, and nut storage to enhance product quality and safety: An overview. Journal of the Science of Food and Agriculture 101(1): 27-43. https://doi.org/10.1002/jsfa.10557

Swain M J & Ballard D H (1991). Color indexing. International Journal of Computer Vision 7(1): 11-32. https://doi.org/10.1007/BF00130487

Tan M & Le Q V (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In K Chaudhuri & R Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning (Vol. 97, pp. 6105-6114). PMLR. https://proceedings.mlr.press/v97/tan19a.html

Taner A, Öztekin Y B & Duran H (2021). Performance analysis of deep learning CNN models for variety classification in hazelnut. Sustainability 13(12): Article 6527. https://doi.org/10.3390/su13126527

Tomak Ö (2025). Comparison of the performance of ShuffleNet and DenseNet deep learning algorithms in the classification of hazelnut fruit. Journal of Hendese Technical Sciences and Engineering 2(2): 96-102 (In Turkish). https://doi.org/10.5281/zenodo.17474580

Ünal Z & Aktaş H (2023). Classification of hazelnut kernels with deep learning. Postharvest Biology and Technology 197: Article 112225. https://doi.org/10.1016/j.postharvbio.2022.112225

Xiao Z, Wang J, Han L, Guo S & Cui Q (2022). Application of machine vision system in food detection. Frontiers in Nutrition 9: Article 888245. https://doi.org/10.3389/fnut.2022.888245

Copyright 2026 The Author(s). This is an open-access article published by Faculty of Agriculture, Ankara University under the terms of the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is properly cited.