# Multi-Task Deep Learning for Automated Tobacco Leaf Grading in a Controlled Environment

Charlie S. Marzan [1,2,*], Conrado Ruiz Jr. [2,3], and Oya Aran [4]

[1]College of Computer Science, Don Mariano Marcos Memorial State University—South La Union Campus, La Union, Philippines
[2]College of Computer Studies, De La Salle University, Manila, Philippines
[3]HER-La Salle-Universitat Ramon Llull, Barcelona, Spain
[4]Telecommunications and Informatics Technologies Research Center, Bogazici University, Istanbul, Turkey
Email: cmarzan@dmmmsu.edu.ph (C.S.M.); conrado.ruiz@salle.url.edu (C.R.J.); aranoya@gmail.com (O.A.)
*Corresponding author

Manuscript received October 31, 2025; revised November 28, 2025; accepted January 30, 2026; published April 17, 2026

***Abstract*—Grading tobacco leaves is crucial for ensuring fair pricing and quality control, however, the process is still largely carried out manually, resulting in a slow, subjective, and often inconsistent outcome. In this work, we present a multi-task deep learning approach designed to automate the grading of air-cured Burley tobacco leaves in controlled settings. The model is constructed with shared convolutional layers and separate task-specific branches, allowing it to predict stalk group, quality, and color at the same time, in line with the hierarchical grading system. To improve consistency, images were preprocessed using coin-based size normalization, rotation alignment, and segmentation. In our experiments, the multi-task model with EfficientNetB0 achieved an accuracy of 94.82% and significantly outperformed the multi-class and single-task baselines, while reducing both training time and inference delay. These findings suggest that multi-task learning can be a valuable and robust method for automated tobacco grading, showing gains in accuracy, speed, and scalability compared to other algorithms.**

***Keywords*—air-cured Burley tobacco, image preprocessing, multi-task deep learning, tobacco leaf grading**

## I. Introduction

Tobacco grading plays a critical role in agricultural production systems, pricing structures, and quality assurance across global supply chains. As a high-value commercial crop, tobacco leaves are evaluated based on visual characteristics, including size, texture, maturity, structural integrity, and color. These characteristics collectively determine the economic worth of each leaf and guide decisions during processing and manufacturing. As a result, grading accuracy directly affects farmer compensation, buyer assessments, and the consistency of tobacco products delivered to domestic and international markets.

In current industry settings, tobacco grading is carried out almost entirely by human graders. While trained graders can develop expertise through experience, manual grading remains inherently subjective, labor-intensive, and sensitive to fatigue or environmental conditions. These limitations often result in inconsistencies, pricing disparities, and disagreements between farmers and purchasing stations. With the increasing demand for standardization and fairness, automated grading has become an essential area of research, and several studies have explored computer vision-based approaches to support or replace manual grading.

Early research attempted to classify tobacco leaves using handcrafted visual features combined with traditional machine learning methods such as Support Vector Machines (SVM), fuzzy systems, symbolic classifiers, and Generalized Regression Neural Networks (GRNN) [1–4]. More recent advances in deep learning demonstrated that Convolutional Neural Networks (CNNs) can achieve strong performance in grading air-cured and flue-cured tobacco leaves by eliminating the need for manual feature engineering [5–6]. Despite these advances, quantitative gaps remain, prior works have been handled anywhere from 3 to 41 grades, yet those evaluating 41 grades used only 1588 [7] and 1498 images [8], limiting their capacity for robust generalization. Conversely, one study used 66,966 images, but these covered only six grades [9], highlighting the lack of balanced, large-scale datasets that can capture the full complexity of tobacco grading.

Although these developments have improved automated grading, several important gaps remain unaddressed. First, most existing systems treat tobacco grading as a flat multi-class classification task, collapsing stalk group, quality, and color into a single label. This simplifies the modeling process but ignores the inherent hierarchical and interdependent nature of the official grading system. As previous studies have shown, these characteristics are closely related, where stalk group influences leaf size and structure, quality reflects maturity and finish, and color is strongly affected by curing conditions [10, 11]. When merged into a single category, these relationships are lost, making it harder for models to generalize and distinguish among visually similar grades.

Second, although deep learning methods such as Alex Krizhevsky Network (AlexNet) [12], Residual Network (ResNet) [13], Visual Geometry Group 16-layer Network (VGG16) [8], Squeeze-and-Excitation Network variants (SENet) [14], Densely Connected Convolutional Network (DenseNet) [10, 15], and hybrid CNN-Transformer architectures [16–19] have shown substantial progress, they still do not explicitly model the dependencies among grading components, resulting in inefficiencies and reduced accuracy in operational scenarios.

To address these challenges, this study proposes a Multi-Task Learning (MTL) framework designed to model the hierarchical structure of tobacco grading. Instead of collapsing the stalk group, quality, and color into a single output, the MTL architecture learns them simultaneously through a shared feature extractor and task-specific prediction branches. This design reduces computational redundancy, captures natural task dependencies, and improves both accuracy and inference speed. To support this framework, we also developed a large-scale dataset comprising 10,137 high-resolution images of air-cured Burley tobacco leaves, collected under controlled lighting and background conditions commonly used in grading centers.

The key contributions of this study are as follows:

1) Development of a comprehensive, expert-annotated dataset of air-cured Burley tobacco leaves containing labels for stalk group, quality, and color. The dataset reflects real grading-room conditions and represents one of the largest publicly documented collections for this domain.
2) Systematic comparison of three learning approaches: multi-class classification, single-task learning, and the proposed multi-task learning framework, using both raw and pre-processed images to evaluate the impact of image normalization and feature consistency.
3) Introduction of a unified multi-task learning framework that jointly predicts stalk group, quality, and color through shared representations. This approach improves accuracy, reduces training time, and significantly increases inference speed, making it more suitable for practical implementation in grading facilities.

## II. Literature Review

This section reviews existing work on automated tobacco leaf grading, with emphasis on how methods have evolved from handcrafted features and traditional machine-learning models to deep convolutional and transformer-based architectures.

### *A. Early Machine Vision Approaches with Handcrafted Features*

Early automated tobacco grading methods relied on handcrafted features, including color, texture, shape, maturity, and injury tolerance, paired with traditional machine-learning algorithms. Support Vector Machines were used to classify stalk positions based on selected visual attributes [1], while texture- and shape-based descriptors supported automatic harvesting applications [2]. Fuzzy comprehensive evaluation was introduced to handle the uncertainty in quality assessment [3], and a GRNN was applied for image-based grading [4]. Rule-based strategies such as the barrel-theory decision algorithm achieved moderate accuracy but showed limited generalization across conditions [20]. Collectively, these early approaches demonstrated the feasibility of automated grading but remained constrained by the need for manual feature design and sensitivity to environmental variation.

### *B. Feature Selection and Optimization-Based Methods*

As researchers sought to improve handcrafted-feature pipelines, several studies incorporated optimization techniques to enhance discriminative feature selection. Particle Swarm Optimization (PSO) was used to identify the most relevant tobacco leaf attributes and eliminate redundant descriptors [21]. Similarly, feature screening integrated with a genetic algorithm and clustering improved robustness by refining feature sets before classification [7]. Multi-feature integration using Fisher's discriminant analysis further demonstrated that combining texture, color, and structural cues could improve separability among grades [22]. While these methods provided performance gains over earlier handcrafted approaches, they still relied heavily on manual feature design and lacked adaptability to complex visual variation.

### *C. Deep Learning for Tobacco Leaf Grading*

The introduction of deep learning significantly advanced tobacco grading by allowing models to learn discriminative features directly from images. Early CNN-based approaches demonstrated clear improvements over handcrafted methods. Deep CNNs were first applied to classify flue-cured tobacco leaves, showing stronger performance in capturing texture and structural details [5]. A similar direction was taken for air-cured Burley leaves, where CNN-based grading outperformed traditional pipelines in real-world settings [6]. Transfer learning further enhanced performance, with pre-trained models improving recognition of hard samples and visually challenging grades [12].

More advanced CNN architectures were later explored. Transfer-learning schemes using ResNet-based models provided more robust feature extraction for large-scale datasets [8], while multi-scale fusion techniques combined features at different resolutions to improve fine-grained grading [13]. Residual networks [23] and deeper CNN variants [24] also demonstrated strong improvements. Collectively, these studies established CNNs as the dominant approach for tobacco grading. However, they typically treated the problem as a single-task or flat multi-class classification, ignoring the hierarchical structure of actual grading systems.

CNN-based approaches have also been applied to plant leaf disease detection, demonstrating the effectiveness of deep feature learning for agricultural image analysis. Recent studies used CNN, ResNet, and AlexNet architectures for coffee leaf disease classification [25, 26] and deep transfer learning with quantum-behaved particle swarm optimization for sugarcane leaf disease detection, achieving promising results under controlled conditions [27]. Although these works focus on disease recognition rather than grading, they support the suitability of CNN-based pipelines and motivate the proposed multi-task grading framework.

### *D. Attention Mechanisms, Dense Connectivity, and Large-Scale Architectures*

As tobacco grading tasks grew more complex, researchers began adopting attention-enhanced and densely connected architectures to capture finer visual distinctions. Large-scale grading frameworks using DenseNet demonstrated that dense feature reuse greatly improves recognition of subtle texture and structural variations [10]. Attention mechanisms were also integrated into CNNs to highlight discriminative leaf regions. Improved bilinear CNNs with pyramid and attention modules enhanced fine-grained feature extraction for multi-category grading [9].

More recent works incorporated transformer-inspired components. Dual-encoder aggregation networks combined deep features with reinforcement modules to strengthen representation learning [28], while shifted-window self-attention mechanisms provided more effective global-local context modeling [29]. Multi-scale separable dilated CNNs further improved performance by capturing variations at multiple spatial resolutions with attention-guided refinement [30]. These advances collectively show a clear trend toward deeper, more expressive, and fine-tuned architectures capable of handling the high variability present in tobacco leaf datasets.

### E. Multimodal and Cross-Modal Tobacco Grading

Recent research has explored multimodal and cross-modal approaches to enhance grading accuracy beyond what RGB images alone can provide. Early multimodal work combined reflectance and transmittance images, demonstrating that integrating complementary optical properties yields better characterization of internal leaf attributes than single-modality inputs [15]. More advanced models incorporated planting metadata and additional imaging signals into unified deep-learning systems. The Cross-Modal Enhancement Network (CMENet) fused reflection, transmission, and contextual metadata to strengthen feature representation across modalities [11].

Other approaches focused on synchronizing multi-aspect grading through specialized architectures. Multi-Dimensional Convolutional Network (MDCNet) enabled simultaneous recognition of multiple tobacco attributes by integrating correlated features across channels [31]. These multimodal and cross-modal strategies consistently improved classification performance, though at the cost of greater hardware complexity and more demanding data acquisition requirements. As such, they provide important insights into feature complementarity but remain difficult to deploy widely in real-world grading facilities.

### F. Lightweight and Real-Time Architectures

To support deployment in grading rooms and purchasing stations, recent studies have focused on lightweight and real-time architectures capable of fast inference with limited computational resources. MobileViT-based models demonstrated that combining convolutional features with transformer-like components can deliver strong performance while maintaining low model complexity [16]. High-efficiency grading networks optimized for speed and stability were also developed to reduce training time and computational load without sacrificing accuracy [17].

Further improvements were introduced through multi-view strategies, such as a six-channel ResNet designed to process both the front and back images of leaves, enabling richer feature extraction within a compact architecture [32]. These lightweight models show promising potential for real-time grading applications. However, they still primarily rely on single-output classification, limiting their ability to capture hierarchical relationships among stalk group, quality, and color.

### G. Hybrid Architectures and Expert-Guided Approaches

Recent studies have explored hybrid models that integrate domain knowledge with advanced deep-learning architectures to enhance fine-grained tobacco grading. Some works used expert-driven visual cues, such as venation patterns, and incorporated vein-based feature fusion to improve discrimination among similar leaf grades [18]. Other researchers proposed integrated architectures combining classical and modern deep-learning components, such as a VGG16-DenseNet hybrid, to leverage both strong hierarchical feature extraction and dense connectivity [33].

Transformers have also begun to influence the domain, with expert-guided Vision Transformer (ViT) models embedding grading knowledge to better align feature representations with human grading criteria [19]. These hybrid and expert-enhanced systems reflect a growing shift toward incorporating agricultural expertise into model design. However, they still operate under flat or single-task formulations that overlook the hierarchical nature of tobacco grading.

### H. Multi-Task Learning in Vision and ITS Relevance to Tobacco Grading

While most tobacco grading studies rely on single-task or flat multi-class formulations, Multi-Task Learning (MTL) has shown strong potential in other computer-vision domains. MTL frameworks use shared feature extractors with task-specific branches to learn related outputs simultaneously, improving generalization and reducing computational overhead. For example, deep MTL frameworks have been successfully applied to malware image analysis, where jointly learning correlated tasks improved performance and training efficiency [34].

For tobacco grading, the primary components, such as stalk group, quality, and color, are inherently interrelated and form a structured hierarchy. However, existing deep-learning approaches typically treat these components in two limiting ways: (1) by collapsing them into a single multi-class label, which obscures their natural dependencies, or (2) by training separate models for each task, which increases computational redundancy and inference time.

Despite substantial advancements in CNNs, attention mechanisms, multimodal models, and hybrid architectures, no published work to date has formulated tobacco grading explicitly as a multi-task learning problem. The absence of such an approach highlights a key gap in the literature and presents a strong rationale for exploring MTL as a means to capture task interdependencies, reduce redundancy, and improve overall grading performance.

## III. Materials and Methods

The proposed method comprises four stages: leaves acquisition, image acquisition, preprocessing, and classification.

### A. Leaves Acquisition

This study examined air-cured Burley-type tobacco leaves obtained from a subsidiary of Universal Corporation between 2019 and 2024. The samples were collected across different tobacco-growing regions of the Philippines and graded using a standardized system with 58 grade marks, determined by stalk position, quality, and color. Stalk position was divided into four groups: Flyings (X), Cutters (C), Leaf (B), and Tips (T), which influence characteristics such as leaf size, texture, and nicotine content. Quality was rated on a five-point scale, from Choice (1) to Low (5), taking into account maturity, body, structure, finish, and tolerance to injury. Color was coded as L (buff), F (tan), R (reddish), K (variegated), V (running green), or G (green), reflecting differences from fermentation and leaf type. Leaves that did not meet grading standards were designated as Nondescript (e.g., Tips Nondescript (TND), Leaf Nondescript (BND)). Each grade combined stalk position, quality, and color. For instance, X1F refers to a Tan Flyings leaf rated as Choice quality.

## *B. Image Acquisition*

Tobacco images were collected under controlled conditions using a Nikon D3400 DSLR camera. The setup was arranged in a grading room equipped with 22 LED lights, with leaves placed on either a whiteboard table or a flour sack. Photographs were taken at a shutter speed of 1/125 with auto white balance to closely replicate the environment used in manual grading. To ensure proper camera settings and image clarity, an expert grader validated the initial samples. Each image was labeled and verified by grading specialists, with a coin included in the frame as a size reference. The images were stored as high-quality JPEGs (300 dpi) at 4000×6000 pixels. In total, 10,137 images representing 35 grades (out of 58 possible) were collected. The dataset is imbalanced, with class sizes ranging from 140 to 542 images. Examples of the dataset are presented in Fig. 1, while distribution details are provided in Table 1.

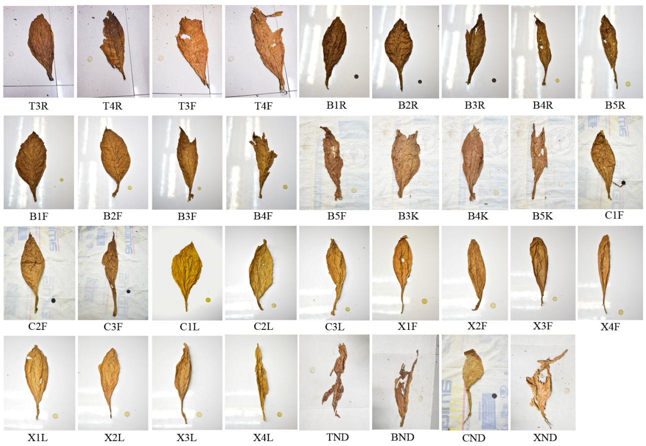


Fig. 1. Representative images of air-cured Burley tobacco leaves captured under controlled imaging conditions, showing natural variations in grade, color, texture, and physical structure across 35 different classes.

Table 1. Air-cured burley tobacco leaf dataset taken under controlled conditions

| Grade | Quantity |
|---|---|
| T3R | 180 |
| T4R | 188 |
| T3F | 192 |
| T4F | 185 |
| B1R | 471 |
| B2R | 458 |
| B3R | 228 |
| B4R | 256 |
| B5R | 205 |
| B1F | 482 |
| B2F | 311 |
| B3F | 430 |
| B4F | 540 |
| B5F | 359 |
| B3K | 220 |
| B4K | 219 |
| B5K | 212 |
| C1F | 472 |
| C2F | 519 |
| C3F | 202 |
| C1L | 200 |
| C2L | 213 |
| C3L | 208 |
| X1F | 382 |
| X2F | 326 |
| X3F | 261 |
| X4F | 201 |
| X1L | 204 |
| X2L | 200 |
| X3L | 190 |
| X4L | 233 |
| TND | 140 |
| BND | 263 |
| CND | 245 |
| XND | 542 |
| Total | 10,137 |

## *C. Preprocessing*

Image preprocessing was performed to ensure that each leaf was consistently scaled, properly oriented, and accurately isolated from the background, while also preserving size information through a reference coin. This is also a crucial step, as emphasized in prior leaf-classification studies that improved accuracy by removing background and extracting Region-of-Interest (ROI) regions [35]. The Hough Circle Transform (HCT) was used to detect the coin in each image, enabling scale adjustment via bilinear interpolation. To standardize orientation, RGB images were first converted to grayscale (using the blue channel), followed by Otsu's thresholding to produce a binary image. Boundary points were then identified, and the leaf was rotated upright using a line drawn between its two farthest points. To reduce noise, Gaussian blurring was applied, and segmentation retained only the largest connected component (i.e., the leaf), resulting in a clean binary mask.

As illustrated in Fig. 2, the coin was detected (Fig. 2(b)) and used for scaling (Fig. 2(c)). A line between the farthest leaf pixels (Fig. 2(d)) guided rotation, aligning the leaf vertically. The resulting mask (Fig. (f)) was then multiplied by the original image (Fig. 2(e)) to isolate the leaf (Fig. 2(g)). Finally, images were resized to 256×256 pixels with black padding to maintain consistency for CNN input (Fig. 2(h)). This preprocessing step was crucial to preserve the actual size of the tobacco leaf, since its distance from the camera's focal plane varied slightly across samples.

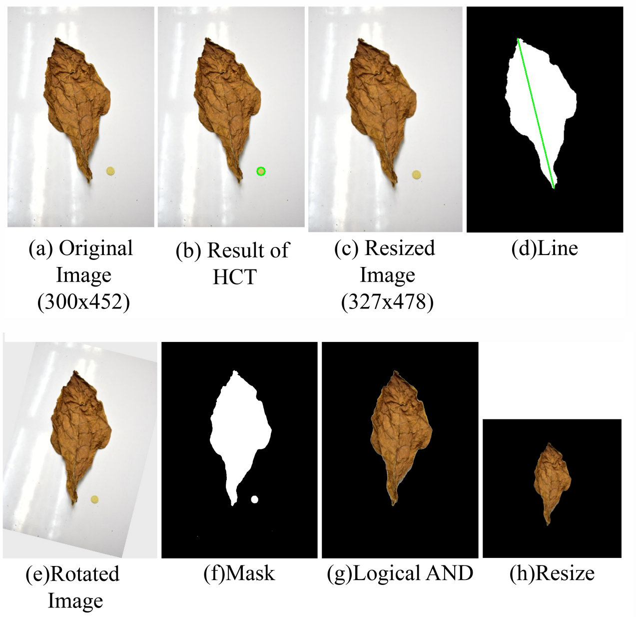


Fig. 2. Illustration of the complete preprocessing sequence used to standardize tobacco leaf images. The procedure includes coin-based scale estimation, rotation correction, segmentation, and resizing, ensuring consistent leaf orientation and dimensions before the images are input to the learning model.

## *D. Classification*

The final stage of the proposed method focuses on classification, where the model predicts the stalk group, quality, and color of each tobacco leaf. Fig. 3 illustrates the Multi-Task Learning (MTL) architecture designed for this purpose. Unlike flat multi-class or isolated single-task approaches, the MTL framework allows a single network to learn all three grading components simultaneously. Because stalk group, quality, and color share many visual cues, such as texture, maturity, and leaf structure, learning them together helps the model capture their natural relationships more effectively [33].

At the core of the model is a shared deep learning backbone that extracts a unified feature representation from each input image. This shared backbone learns a sequence of visual features useful across tasks, reducing redundancy and improving generalization. From this shared representation, the model branches into three task-specific heads that predict stalk group, quality, and color. The stalk group and color branches are relatively shallow, while the quality branch is deeper to capture the more subtle variations needed for quality assessment.

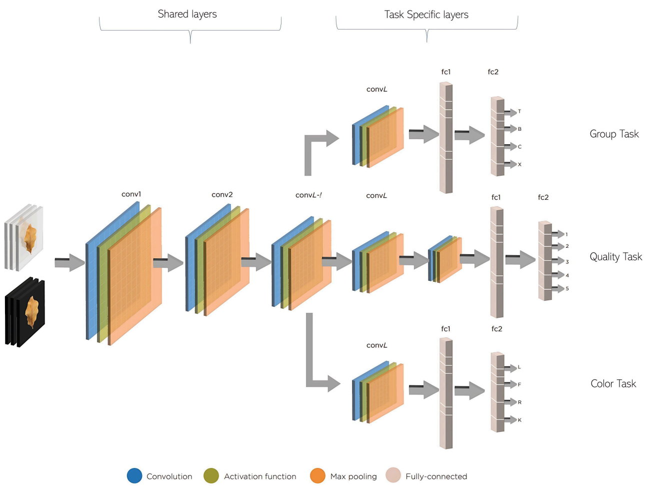


Fig. 3. Overview of the proposed CNN-based multi-task learning framework. A shared feature extractor feeds three task-specific branches responsible for predicting stalk group, quality, and color, enabling joint optimization of all grading components.

The procedure for training the MTL model is summarized in Algorithm 1, which presents the algorithm for optimizing the shared backbone and the three task-specific branches. As shown in the algorithm, each mini-batch first passes through the shared backbone to extract feature representations. These features are then fed into the three branches, producing parallel predictions for stalk group, quality, and color. A separate cross-entropy loss is calculated for each task, and the final training objective is computed as a weighted sum of these losses, as shown in Eq. (1).

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{group} + \beta\mathcal{L}_{quality} + \gamma\mathcal{L}_{color} \quad (1)$$

where $\mathcal{L}_{group}$, $\mathcal{L}_{quality}$, and $\mathcal{L}_{color}$ represent the loss functions for group, quality, and color classification tasks, respectively, and $\alpha$, $\beta$, $\gamma$ are scalar task-specific weighting coefficients that control each task's contribution to the total loss. In this study, the stalk group and color tasks were each assigned a weight of 0.3 ($\alpha = 0.3$, $\gamma = 0.3$), while the quality task was assigned a higher weight of 0.4 ($\beta = 0.4$) to reflect its greater classification difficulty. This joint optimization strategy encourages the model to learn shared representations across tasks while allowing increased emphasis on the more challenging quality classification.

**Algorithm 1: Training algorithm of the proposed multi-task learning framework**

```
Input: D = {(x, y_i^stalk, y_i^qual, y_i^color)}
        α, β, γ
Output: θ*, φ_stalk*, φ_qual*, φ_color*
        // trained backbone and task heads
1: // Model initialization
2: Initialize shared backbone f_θ (EfficientNetB0)
3: Initialize task heads:
   g_stalk(; φ_stalk)
   g_qual(; φ_qual)
   g_color(; φ_color)
4: Configure optimizer (Adam), learning-rate scheduler,
   dropout, and L2 regularization
5: Enable early stopping (patience = 10)
6: // Training loop
7: Repeat for each epoch until early stopping:
8:     For each mini-batch B from D:
9:         h = f_θ(B.x)
10:        ŷ_stalk = g_stalk(h)
11:        ŷ_qual  = g_qual(h)
12:        ŷ_color = g_color(h)
13:        L_stalk = CE(B.y_stalk, ŷ_stalk)
14:        L_qual  = CE(B.y_qual, ŷ_qual)
15:        L_color = CE(B.y_color, ŷ_color)
16:        L_total = α·L_stalk + β·L_qual + γ·L_color
17:        Backpropagate L_total and update all parameters
18:    Evaluate on validation set and update best model
19: Return θ*, φ_stalk*, φ_qual*, φ_color*
```

The total loss is backpropagated to jointly update all parameters, and early stopping is used to retain the best model based on validation performance. Overall, the architecture and algorithm reflect the core idea of the proposed system: a unified network that efficiently learns multiple correlated grading tasks. Certain leaves, such as TND, BND, Cutters Nondescript (CND), and Flyings Nondescript (XND), fall outside the official grading categories and are considered nondescript. These samples exhibit irregular characteristics that do not align cleanly with the stalk group, quality, or color labels. Rather than forcing them into the multi-task framework, a separate binary CNN classifier was trained exclusively to detect nondescript leaves. This ensures that such samples are accurately identified without disrupting the main MTL model's structure.

To provide meaningful points of comparison, two baseline approaches were also implemented: multi-class classification and single-task learning.

- Multi-class classification treats each of the 35 grades as a single label. While straightforward, this approach collapses the hierarchical grading structure and struggles with subtle visual overlaps between related grades.
- Single-task learning trains three separate models, one each for stalk group, quality, and color. Although this specialization can improve accuracy, it comes at the cost of increased training time, higher computational demand, and slower inference because every image must pass through three independent networks.

The proposed MTL framework addresses these limitations by combining specialization and efficiency within a single architecture, offering improved accuracy, faster inference, and a grading process that remains faithful to the hierarchical structure of tobacco classification.

### E. Experimental Setup and Performance Evaluation

Tables 2 and 3 present the experimental conditions used throughout this study. Table 2 outlines how the dataset was prepared and the configuration of the training process. The setup emphasizes consistency, fairness, and reproducibility. By establishing a fixed test set and balancing the remaining data before training through class-aware random undersampling of majority classes and controlled data augmentation of minority classes, the evaluation becomes more reliable and resistant to class bias. Classes containing more than 300 images were randomly undersampled, while classes with fewer than 300 images were augmented to reach a uniform size of 300 samples using horizontal flipping and translation applied equally across these classes. Random undersampling was selected over class-weighted loss to maintain training stability and avoid synthetic feature distortion, as preliminary experiments, discussed in Section IV, indicated unstable optimization with class-weighted loss and increased overfitting when aggressive oversampling was applied. The training strategy, supported by augmentation, regularization, learning-rate scheduling, and early stopping, was designed to optimize performance while preventing overfitting. These choices ensure that the resulting model performance reflects genuine learning rather than artifacts of the dataset or training procedure.

Table 2. Summary of experimental setup

| Component | Description |
|---|---|
| Dataset Type | Image-controlled tobacco leaf dataset (35 classes) |
| Test Set Construction | 50 images per class (1750 total), fixed and separated from training/validation |
| Training & Validation Dataset | 300 images per class (balanced via undersampling and augmentation) |
| Dataset Split | 80% training (8400 images), 20% validation (2100 images) |
| Data Augmentation | Horizontal flip, translation |
| Preprocessing Operations | Segmentation, rotation alignment, size normalization |
| Optimizer | Adam optimizer |
| Initial Learning Rate | 0.0001 |
| Batch Size | 8 |
| Learning Rate Scheduler | Reduce LR by factor of 0.1 if validation loss plateaus for 5 epochs |
| Regularization Techniques | L2 weight decay ($\lambda$ = 0.0005), dropout (rate = 0.5) |
| Early Stopping | Patience = 10 epochs |
| Evaluation Metrics | Accuracy, training time (hours), inference speed (ms/frame) |

On the other hand, Table 3 summarizes the hardware and software environment in which all experiments were conducted. All experiments were performed on a dedicated local machine to ensure consistency, stability, and full reproducibility of the results. The experimental workflow was arranged so that all models (e.g., multi-class, single-task, and multi-task) were trained and tested on the same hardware under identical conditions, ensuring a fair and unbiased comparison across all approaches. The GPU served as the primary accelerator for model training and inference, while the CPU handled preprocessing, data loading, and augmentation tasks. This standardized arrangement ensured a fair comparison across all evaluated techniques.

Table 3. Hardware and software configurations

| Component | Description |
|---|---|
| Operating System | Windows 10 (64-bit) |
| CPU | Intel(R) Core(TM) i7-7700HQ @ 2.4 GHz |
| RAM | 32 GB |
| GPU | NVIDIA GeForce GTX 1050 (4 GB VRAM) |
| Programming Language | Python |
| Deep Learning Framework | Keras with TensorFlow backend |
| Additional Libraries | NumPy, OpenCV, scikit-learn, Matplotlib |

## IV. Results and Discussion

This study presents a comprehensive evaluation of three approaches: multi-class classification, single-task learning, and the proposed multi-task classification, tested across various deep learning architectures and image settings (pre-processed, raw, and mixed inputs).

### A. Comparison Performance of Multi-Class, Single-Task, and Multi-Task Learning Approaches

Table 4 presents the consolidated performance results of multi-class classification, single-task learning, and the proposed Multi-Task Learning (MTL) framework.

Table 4. Performance comparison of different learning strategies across various CNN architectures under preprocessed, raw, and mixed input settings

| Learning Strategies | Models | Using Pre-processed Images (%) | Using Raw Images (%) | Mixed Inputs (%) |
|---|---|---|---|---|
| Multi-Class Classification | 3-Conv | 79.91 | 79.57 | 79.56 |
| | 7-Conv | 83.86 | 83.19 | 83.41 |
| | AlexNet | 84.10 | 83.81 | 83.95 |
| | GoogleNet | 85.14 | 84.00 | 86.02 |
| | ResNet50 | 88.67 | 85.00 | 87.25 |
| | MobileNetV2 | 88.67 | 88.00 | 89.25 |
| | EfficientNetB0 | 89.94 | 88.04 | 89.60 |
| Single-Task Classification | 3-Conv | 83.00 | 82.67 | 81.78 |
| | 7-Conv | 84.29 | 83.95 | 84.00 |
| | AlexNet | 85.97 | 84.44 | 85.38 |
| | GoogleNet | 87.65 | 85.46 | 87.64 |
| | ResNet50 | 89.03 | 86.62 | 89.51 |
| | MobileNetV2 | 91.05 | 89.00 | 89.86 |
| | EfficientNetB0 | 92.25 | 90.00 | 90.25 |
| Multi-Task Classification | 3-Conv | 83.99 | 84.10 | 80.84 |
| | 7-Conv | 85.97 | 85.63 | 83.46 |
| | AlexNet | 86.42 | 85.82 | 85.03 |
| | GoogleNet | 88.98 | 86.37 | 85.92 |
| | ResNet50 | 90.41 | 87.32 | 89.62 |
| | MobileNetV2 | 92.90 | 91.68 | 92.25 |
| | EfficientNetB0 | 94.82 | 92.07 | 94.10 |

In the multi-class formulation, each image is assigned directly to one of the 35 tobacco grades, treating the grading system as a flat classification problem and ignoring the natural separation between stalk group, quality, and color. The results show that model performance improves as the network architecture becomes deeper and more expressive. Lightweight models such as 3-Conv and 7-Conv achieve modest accuracies, while classical pre-trained networks such as AlexNet and GoogleNet perform substantially better. The strongest results are obtained from more modern architectures, including MobileNetV2, ResNet50, and EfficientNetB0, with EfficientNetB0 achieving the highest multi-class accuracy of 89.94% using pre-processed images. A consistent trend appears across all architectures: pre-processed images outperform raw and mixed inputs, while mixed inputs lead to the lowest performance, suggesting that inconsistent visual distributions confuse the classifier. Although multi-class classification performs reasonably well, its main limitation becomes clear: compressing all three grading components into a single label makes it difficult for the model to distinguish visually similar grades, particularly when differences exist only in quality or color.

When the problem is reformulated into single-task classification, where separate models are trained independently for stalk group, quality, and color, accuracies increase across nearly all architectures. This improvement demonstrates the benefit of reducing task complexity, as each model specializes in learning a single grading dimension. For example, EfficientNetB0 achieves 92.25% accuracy with pre-processed images, surpassing its multi-class counterpart by more than two percentage points. Pre-processed images again yield the highest performance, confirming the importance of clean and standardized visual inputs, while mixed inputs produce the lowest results across most models. However, despite the accuracy improvement, single-task learning introduces a practical drawback. Because three independent networks must be trained, validated, and executed, the approach significantly increases training time and slows inference, as every image must pass sequentially through three models. These findings indicate that although single-task learning improves predictive accuracy compared with multi-class classification, it is not computationally efficient.

The proposed multi-task learning framework further improves performance by using a shared backbone to extract common visual features, followed by dedicated branches for stalk group, quality, and color prediction. This approach consistently outperforms both baseline methods across all evaluated architectures. The best-performing configuration, EfficientNetB0 with pre-processed images, achieves 94.82% accuracy, representing the highest result among all experiments. The observed performance gain highlights the advantage of learning stalk group, quality, and color jointly within a unified architecture. Because these attributes are visually interrelated, shared representation learning provides richer supervisory signals and enables the model to capture their dependencies more effectively than either flattened multi-class classification or isolated single-task models. The results also demonstrate that MTL remains stable across different image types, although pre-processed inputs consistently produce the best outcomes.

Overall, the comparative accuracy results show that the multi-task framework achieves superior classification performance while better preserving the hierarchical structure of tobacco grading.

### B. *Computational Efficiency and Real-Time Performance Evaluation*

Table 5 summarizes the comparative performance and computational efficiency of three tobacco leaf grading strategies: multi-class classification, single-task learning, and the proposed Multi-Task Learning (MTL) framework. The multi-class and single-task models are treated as baseline configurations to establish a clear reference for evaluating the accuracy, training time, and inference speed of the proposed approach.

Table 5. Comparative performance and computational efficiency of multi-class, single-task, and multi-task learning approaches using EfficientNetB0 with pre-processed images

| Techniques | Training Time (h) | Accuracy (%) | Inference Speed (ms/frame) |
|---|---|---|---|
| Multi-Class (EfficientNetB0 + Preprocessed Images) | 5.35 | 89.94 | 35 |
| Single-Task Learning (+ Preprocessed Images) | 13.56 | 92.25 | 84 |
| Multi-Task Learning (+ Preprocessed Images) – Our proposed method | 4.20 | 94.82 | 18 |

The multi-class approach is the fastest to train (5.35 h) but achieves only 89.94% accuracy. Single-task learning improves accuracy to 92.25%, but at the cost of dramatically longer training (13.56 h) and slow inference (84 ms/frame) due to multiple networks. This baseline indicates that a deep network can effectively distinguish among grade categories when treated as a single classification task. However, even with a balanced dataset, this approach struggled with the high visual similarity between specific grades. For example, leaves that differ only slightly in quality or color are particularly difficult to distinguish under a flattened classification formulation. More importantly, flattening the grading system ignores its hierarchical nature, where stalk group, quality, and color interact to determine the final grade. As a result, the multi-class baseline was less effective in modeling these interdependencies.

When the grading tasks were separated into single-task models, with a dedicated EfficientNetB0 trained individually for stalk group, quality, and color, accuracy improved to 92.25%. This improvement reflects the advantage of reducing task complexity, since each model could specialize in a single grading dimension. The drawback, however, was efficiency as maintaining three independent models extended the training time to 13.56 h and slowed inference to 84 ms per frame, since each image had to pass through three networks sequentially.

In contrast, the proposed MTL framework (EfficientNetB0) delivers the strongest overall performance, with an accuracy of 94.82%. It also required only 4.20 h of training, which was faster than both baselines. This efficiency comes from using a shared backbone with three task-specific branches, eliminating redundant computations across separate models. In addition, MTL offered the lowest inference latency of 18 ms per frame, which corresponds to approximately 55 Frames Per Second (FPS). Since real-time vision systems typically require only 25–30 FPS [36] for practical deployment, this processing speed is more than sufficient for live grading setups.

Overall, these results demonstrate that the MTL framework not only improves classification accuracy but also provides substantial benefits in training efficiency and inference speed. By explicitly modeling the interdependencies among stalk group, quality, and color in a single unified architecture, MTL emerges as a more accurate, scalable, and practical solution for real-world automated tobacco grading. The improvements are quantitatively meaningful. The MTL framework delivers a +4.88% accuracy gain over the multi-class baseline and a +2.57% gain over the single-task model, while reducing training time by up to 69% and improving inference speed by a factor of 4.6. These consistent gains across all core metrics demonstrate that the proposed method not only enhances predictive accuracy but also provides substantial computational benefits, reinforcing its suitability for real-world, real-time deployment.

### *C. Influence of Preprocessing on Classification Accuracy*

Across all evaluated models and learning strategies, a consistent pattern emerges in which pre-processed images outperform raw and mixed inputs. The structured preprocessing pipeline, which is composed of segmentation, rotation alignment, and size normalization, removes irrelevant variations, enabling the models to focus on meaningful leaf characteristics. This consistency across architectures and learning paradigms indicates that preprocessing contributes not merely to incremental gains, but to systematic stabilization of feature learning.

In contrast, mixing raw and pre-processed images leads to heterogeneous visual distributions that destabilize learning. This decline in performance is visible across all architectures and classification techniques.

These findings reinforce the importance of using controlled images for fine-grained agricultural grading, where slight differences in color, texture, and structure can change the assigned grade.

### *D. Ablation Study on Dataset Balancing Strategies*

Table 6 presents the ablation results comparing different dataset balancing strategies using EfficientNetB0 under identical experimental conditions. Random undersampling combined with controlled data augmentation achieved the highest accuracy (94.82%) and the most stable training behavior, indicating that reducing majority-class dominance while modestly augmenting minority classes supports more effective feature learning.

Table 6. Performance comparison of dataset balancing strategies

| Strategy | Overall Accuracy (%) | Training Stability |
|---|---|---|
| Random undersampling + augmentation (horizontal flip + translation) | 94.82 | Stable |
| Oversampling via data augmentation only | 93.67 | Moderate |
| Class-weighted loss (no resampling) | 92.41 | Unstable |

In contrast, oversampling via data augmentation only yielded lower accuracy (93.67%) and moderate stability, suggesting that repeated augmented samples can introduce redundancy and overfitting, particularly for visually similar quality grades. The class-weighted loss approach produced the lowest accuracy (92.41%) and unstable convergence, likely due to dominant gradient updates from rare classes.

These results confirm that random undersampling combined with controlled augmentation provides the most stable optimization behavior and best generalization performance, justifying its selection in the final multi-task framework.

### *E. Error Analysis and Misclassified Samples*

Fig. 4 presents the per-task confusion matrices for stalk group, quality, and color classification using the proposed multi-task learning model. Across all three tasks, the matrices exhibit strong diagonal dominance, indicating that the model correctly classifies most samples and captures the key visual characteristics relevant to each grading dimension.

For the stalk group task, misclassifications are limited and primarily occur between adjacent stalk positions, particularly between B (Leaf) and C (Cutter). These errors are expected, as neighboring stalk groups often share similar size and structural features, making their boundaries visually subtle even for expert graders. The X (Flyings) group shows minimal confusion with other classes, reflecting its more distinct visual appearance.

In the quality classification task, the confusion matrix reveals a higher level of inter-class confusion compared to the stalk group and color. Most errors occur between neighboring quality levels, where differences are mainly due to gradual changes in maturity, surface smoothness, dryness, and minor defects. These attributes are inherently fine-grained and difficult to separate based solely on visual cues, explaining why quality classification emerges as the most challenging task among the three.

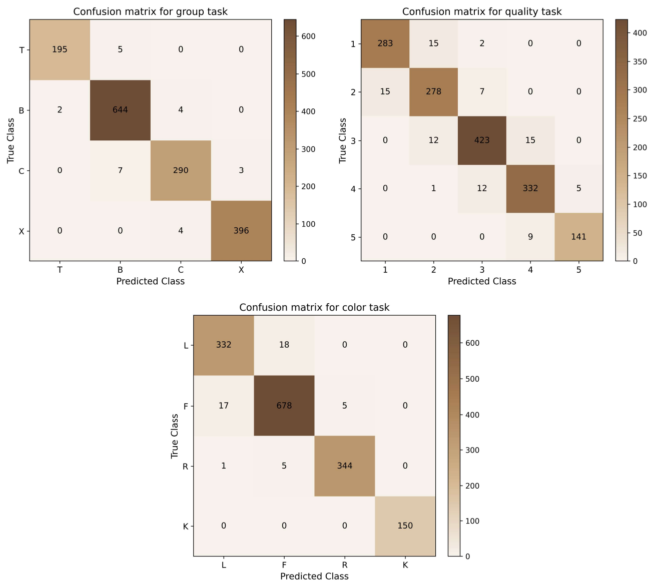


Fig. 4. Per-task confusion matrices for the proposed multi-task learning model showing performance across the stalk group, quality, and color classification tasks. Rows denote true classes and columns denote predicted classes.

For the color task, the confusion matrix shows very high accuracy, with limited misclassification between visually similar color categories such as L (buff) and F (tan). This indicates that the model effectively learns chromatic features under controlled imaging conditions, with residual errors likely caused by subtle lighting variations and transitional color tones.

The error patterns observed in Fig. 4 suggest that most misclassifications occur in borderline cases between visually similar classes, rather than from systematic model failures. To further illustrate these error sources, Fig. 5 presents representative examples of misclassified tobacco leaves. As shown, most misclassified samples involve quality-related errors, where slight variations in maturity, dryness, surface texture, or localized defects result in confusion between adjacent quality grades. These cases highlight the inherent difficulty of quality assessment and are consistent with the confusion patterns observed in the quality confusion matrix.

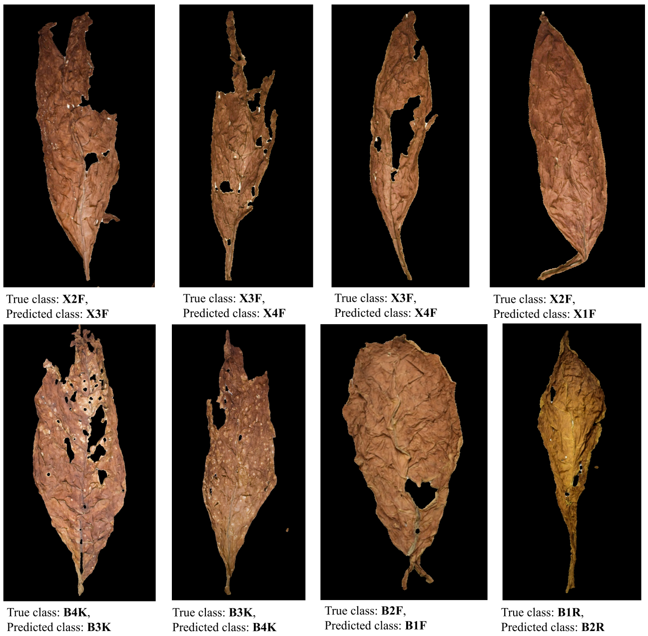


Fig. 5. Sample misclassified images from the multi-task EfficientNetB0 model, showing cases where minimal differences in texture, dryness, or color intensity confused closely related quality classes.

Despite this complexity, assigning a higher loss weight to the quality branch helped the model focus on these fine-grained cues. As a result, the multi-task network produced fewer quality-related mistakes and exhibited a cleaner diagonal structure in the quality confusion matrix, indicating more consistent predictions.

The improvement is mainly due to the model's ability to use contextual signals, such as how color and stalk position correlate with quality, to better resolve borderline cases. In summary, the multi-task framework demonstrates robust performance across all grading dimensions, while the remaining errors reflect intrinsic challenges in fine-grained tobacco leaf grading. Importantly, the absence of large off-diagonal error clusters indicates that the model does not exhibit systematic bias toward any specific class, further supporting the robustness of the proposed framework.

## V. Conclusion and Future Work

This study presented a Multi-Task Learning (MTL) framework to modernize the grading of Burley tobacco leaves. By modeling stalk group, quality, and color together, rather than treating them as separate predictions or combining them into a single flat label, the framework mirrors the way human graders evaluate leaves, but with the added benefits of consistency and speed. This shift from conventional single-task and multi-class models represents a meaningful engineering improvement in automated agricultural assessment. The preprocessing pipeline, particularly coin-based size normalization and orientation correction, also strengthened the consistency and reproducibility of feature extraction, supporting more stable model performance.

The results demonstrate the value of this approach. EfficientNetB0, used as the shared backbone, achieved 94.82%, outperforming both the multi-class baseline (89.94%) and the single-task approach (92.25%), while also reducing training time and delivering the fastest inference at 18 ms/frame. These improvements translate directly into practical advantages in real-world grading environments, where decisions must be both reliable and efficient.

Compared with existing studies, the proposed framework achieves substantial numerical gains across a more challenging grading space. Prior research addressed anywhere from 3 to 41 grades, yet often relied on small datasets (e.g., only 1588 and 1498 images for 41-grade evaluations) or large datasets covering only a few classes, such as 66,966 images for just six grades. Against this backdrop, the proposed model delivers higher accuracy while handling 35 expert-defined grades, demonstrating strong scalability and generalization.

Along with strong predictive performance, this study contributes a clear pathway for improving tobacco grading practices. The combination of a controlled imaging setup, a large expert-annotated dataset, and an MTL-based architecture provides a foundation that can be expanded and refined over time. With further development, such systems could help reduce disputes in buying stations, support fairer pricing for farmers, and enhance quality control for manufacturers.

In practical deployment scenarios, such as conveyor-belt grading systems, the operational design typically enforces single-leaf presentation per frame to ensure reliable inspection. Under this assumption, the current preprocessing pipeline based on the Largest Connected Component (LCC) segmentation remains effective and robust. However, in rare cases where partial overlap or occlusion may still occur due to handling or mechanical irregularities, adjacent leaves could be merged into a single region, potentially affecting segmentation quality and downstream classification. To enhance robustness in such situations, the proposed framework can be complemented with confidence-based rejection mechanisms, whereby samples producing low prediction confidence in one or more task branches (stalk group, quality, or color) are automatically flagged for manual review. This human-in-the-loop strategy allows ambiguous or non-ideal samples to be handled reliably without interrupting automated throughput, while preserving grading accuracy in high-volume industrial environments.

While the results are promising, several limitations should be noted. The model was trained on only 35 of the 58 official Burley grades, as some rare classes lacked sufficient samples, which limits full-grade generalization. In addition, all images were collected under controlled indoor conditions, which may not fully capture the variability encountered in real grading environments, such as changes in lighting, background, camera type, or acquisition setup. The system also relies solely on RGB images and does not incorporate other potentially informative cues, such as moisture content, chemical attributes, or multispectral information that human graders may implicitly consider during manual assessment.

Future work should address these limitations by expanding the dataset to include the complete 58-grade classification system, capturing images under diverse and uncontrolled field conditions, and exploring the integration of multimodal sensing, such as hyperspectral imaging or IoT-based leaf condition sensors. More advanced architectures, including Vision Transformers, hybrid CNN-ViT models, and self-supervised feature learning approaches, may further improve robustness to domain shifts and enhance interpretability.

Additionally, on-field testing and deployment within real-time IoT-enabled grading systems, such as conveyor-belt inspection units, would help transition the framework from controlled laboratory settings to practical use in tobacco-buying stations and supply chains. Future implementations should also consider security aspects, including secure data storage, access control, and protected deployment within closed or on-premise networks, to ensure data integrity and trustworthy operation in real-world grading environments, supporting scalable, sensor-driven automated grading. Reliability-aware deployment strategies, such as confidence-based decision thresholds and human-in-the-loop verification for borderline cases, will also be explored to further enhance dependable grading performance, supporting scalable, sensor-driven automated grading.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

CM led the entire research process, including conceptualizing the study, designing the methodology, developing the dataset, conducting all primary experiments, analyzing the results, and writing the full manuscript. CR and OA contributed by assisting with additional experiments, validating the findings, and providing critical feedback that strengthened the final analysis. All authors reviewed and approved the final version of the manuscript.

## Acknowledgment

The authors wish to thank Universal Leaf Philippines Inc., De La Salle University, Don Mariano Marcos Memorial State University, the grading experts, and the dedicated tobacco farmers for their invaluable support and contributions to this study.

## References

[1] L. Han, "Recognition of the part of growth of flue-cured tobacco leaves based on support vector machine," in *Proc. 7th World Congress on Intelligent Control and Automation*, 2008, pp. 3624–3627.

[2] D. S. Guru, P. B. Mallikarjuna, and S. Manjunath *et al.*, "Machine vision-based classification of tobacco leaves for automatic harvesting," *Intelligent Automation & Soft Computing*, vol. 18, no. 5, pp. 581–590, 2012.

[3] F. Zhang and X. Zhang, "Classification and quality evaluation of tobacco leaves based on image processing and fuzzy comprehensive evaluation," *Sensors*, vol. 11, no. 3, pp. 2369–2384, 2011.

[4] J. Liu, J. Shen, and Z. Shen *et al.*, "Grading tobacco leaves based on image processing and generalized regression neural network," in *Proc. Intelligent Conference on Intelligent Control, Automatic Detection and High-End Equipment*, 2012, pp. 89–93.

[5] S. K. Dasari, K. R. Chintada, and M. Patruni, "Flue-cured tobacco leaves classification: A generalized approach using deep convolutional neural networks," in *Cognitive Science and Artificial Intelligence*, Singapore: Springer Singapore, 2017, pp. 13–21.

[6] C. S. Marzan and C. R. Ruiz, "Automated tobacco grading using image processing techniques and a convolutional neural network," *International Journal of Machine Learning and Computing*, vol. 9, no. 6, pp. 807–813, 2019.

[7] L. Pei, J. Shen, and R. Liu, "Deep feature of image screened by improved clustering algorithm cascaded with genetic algorithm," in *Proc. 29th Chinese Control and Decision Conference,* 2017, pp. 452–455.

[8] G. Li, H. Zhen, and F. Jiao *et al.*, "Research on tobacco leaf grading algorithm based on transfer learning," in *Proc. International Conference on Artificial Intelligence and Computer Applications*, 2021, pp. 32–35.

[9] M. Lu, C. Wang, W. Wu *et al.*, "Intelligent grading of tobacco leaves using an improved bilinear convolutional neural network," *IEEE Access*, vol. 11, pp. 68153–68170, 2023.

[10] X. Xin, H. Gong, R. Hu *et al.*, "Intelligent large-scale flue-cured tobacco grading based on deep densely convolutional network," *Scientific Reports*, vol. 13, no. 1, 2023.

[11] Z. He, G. Chen, Y. Zhang *et al.*, "Pyramid feature fusion through shifted window self-attention for tobacco leaf classification," *Expert Systems with Applications*, vol. 230, 120601, 2023.

[12] H. Luo and C. Zhang, "Features representation for flue-cured tobacco grading based on transfer learning to hard sample," in *Proc. 14th IEEE International Conference on Signal Processing*, 2018, pp. 591–595.

[13] M. Lu, S. Jiang, C. Wang *et al.*, "Tobacco leaf grading based on deep convolutional neural networks and machine vision," *Journal of the ASABE*, vol. 65, no. 1, pp. 11–22, 2022.

[14] L. Mengyao, Z. Qiang, J. Shuwen *et al.*, "Flue-cured tobacco leaf grading method based on deep learning and multi-scale feature fusion," *Journal of Chinese Agricultural Mechanization*, vol. 43, no. 1, pp. 158–166, 2022.

[15] Q. Li, H. Lin, J. Hu *et al.*, "Automated tobacco leaf grading system based on deep learning," in *Proc. IEEE 3rd International Conference on Software Engineering and Artificial Intelligence*, 2023, pp. 40–46.

[16] W. Gu, C. Duan, X. Gai *et al.*, "Automatic grouping method of flue-cured tobacco based on MobileViT," *Chinese Tobacco Science*, vol. 45, no. 1, pp. 104–111, 2024. (in Chinese)

[17] C. Feng, S. Zhu, H. Huang *et al.*, "Research on high efficiency flue-cured tobacco grade identification model based on deep learning," *Journal of Southwest University Natural Science Edition*, vol. 47, no. 1, pp. 213–225, 2025.

[18] Y. Chen, C. Lian, X. Chen *et al.*, "Fine-grained visual classification of flue-cured tobacco leaf grades based on vein information," in *Proc. 10th International Conference on Systems and Informatics*, 2024, pp. 1–6.

[19] D. Wang, Y Li, C. Wang *et al.*, "Tobacco leaf grading algorithm based on expert ViT model," in *Proc. 2025 Asia-Europe Conference on Cybersecurity, Internet of Things and Soft Computing*, 2025, pp. 234–238.

[20] L. Huabo, L. He, and L. Tao, "The barrel theory-based decision-making algorithm and its application," in *Proc. 2009 International Conference on Computational Intelligence and Natural Computing,* 2009, pp. 11–14.

[21] X. Ma, J. Shen, R. Liu *et al.*, "Choice of tobacco leaf features based on selected probability of particle swarm algorithm," in *Proc. 2016 Chinese Control and Decision Conference*, 2016, pp. 3041–3044.

[22] S. Yang, C. Dong, F. Wang *et al.*, "Fisher's tobacco leaf grading method based on image multi-features," in *Proc. 2022 International Conference on Artificial Intelligence and Computer Information Technology*, 2022, pp. 1–6.

[23] D. Wang, G. Li, H. Zhen *et al.*, "Intelligent classification of tobacco leaves based on residual network," in *Proc. International Conference on Artificial Intelligence and Computer Applications*, 2022, pp. 156–159.

[24] C. Zhang, L. Yun, and Z. Chen, "Flue-cured tobacco grading method based on a convolutional neural network," in *Proc. CCF National Conference on Computer Applications*, 2023, pp. 54–66.

[25] R. Surendran, C. P. Sree, and K. Sumathy, "Texture- and symptom-based coffee leaf disease classification using CNN and ResNet algorithms," in *Proc. 4th International Conference on Advancement in Electronics & Communication Engineering*, 2024, pp. 1287–1292.

[26] P. Sree, R. Surendran, and R. Selvanarayanan, "Leveraging CNN and AlexNet algoritms for improved coffee leaf disease identification and detection," in *Proc. 2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering,* 2024, pp. 1–6.

[27] T. Tamilvizhi, R. Surendran, K. Anbazhagan *et al.,* "Quantum-behaved particle swarm optimization-based deep transfer learning model for sugarcane leaf disease detection and classification," *Mathematical Problems in Engineering,* vol. 1, 3452413, 2022.

[28] D. Chen, Y. Zhang, Z. He *et al.*, "Feature-reinforced dual-encoder aggregation network for flue-cured tobacco grading," *Computers and Electronics in Agriculture*, vol. 210, 107887, 2023.

[29] Q. He, X. Zhang, J. Hu *et al.*, "CMENet: A cros-modal enhancement network for tobacco leaf grading," *IEEE Access*, vol. 11, pp. 109201–109212, Oct. 2023.

[30] M. Xu, J. Gao, Z. Zhang *et al.*, "Multi-channel and multi-scale separable dilated convolutional neural network with attention mechanism for flue-cured tobacco classification," *Neural Computing and Applications*, vol. 35, no. 21, pp. 15511–15529, 2023.

[31] T. Hu, K. Hou, Y. Su *et al.*, "Research on synchronous recognition method of multiple tobacco leaf grades based on MDCNet," in *Proc.*

*5th International Conference on Frontiers Technology of Information and Computer*, 2023, pp. 483–490.

[32] S. Wang, R. Zhuang, W. Lu *et al.*, "Intelligent classification of flue-cured tobacco based on improved network model of six-channel ResNet," *Journal of Biotech Research*, vol. 20, pp. 98–110, 2025.

[33] H. Benrong, F. Zhaofeng, W. Fei *et al.*, "Intelligent grading of flue-cured tobacco based on VGG16-DenseNet integrated model," *Chinese Tobacco Science*, vol. 45, no. 3, pp. 102–112, 2024.

[34] A. Bensaoud and J. Kalita, "Deep multi-task learning for malware image classification," *Journal of Information Security and Applications*, vol. 64, 103057, 2022.

[35] Y. J. Wu, C. M. Tsai, and F. Shih, "Improving leaf classification rate via background removal and ROI extraction," *Journal of Image and Graphics,* vol. 4, no. 2, pp. 93–98, 2016.

[36] R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed., Switzerland: Springer, 2022.