

A Hierarchical Deep Learning Framework for Robust Cassava Crop Verification and Disease Diagnosis

Ahmed Wasiu Akande

Hangzhou International Innovation Institute, United Nations Regional Centre for Space Science and Technology Education in the Asia-Pacific Region, Beihang University, China

Dawit Kassaw Gedefaw

dawitkassaw@buaa.edu.cn

Hangzhou International Innovation Institute, United Nations Regional Centre for Space Science and Technology Education in the Asia-Pacific Region, Beihang University, China

Yang Dongkai

Hangzhou International Innovation Institute, United Nations Regional Centre for Space Science and Technology Education in the Asia-Pacific Region, Beihang University, China

Jing Guifei

Hangzhou International Innovation Institute, United Nations Regional Centre for Space Science and Technology Education in the Asia-Pacific Region, Beihang University, China

Edmund Aviedo Atakpo

Department of Physics, Delta State University, Abraka, Nigeria

Article

Keywords: cassava disease classification, hierarchical deep learning, precision agriculture, EfficientNet, attention mechanisms, imbalanced learning, remote sensing

Posted Date: November 29th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-8138247/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A Hierarchical Deep Learning Framework for Robust Cassava Crop Verification and Disease Diagnosis

Ahmed Wasiu Akande¹, Dawit Kassaw Gedefaw^{1,*}, Yang Dongkai¹, Jing Guifei¹, Edmund Aviedo Atakpo²

¹ Hangzhou International Innovation Institute, United Nations Regional Centre for Space Science and Technology Education in the Asia-Pacific Region, Beihang University, China

² Department of Physics, Delta State University, Abraka, Nigeria

* **Correspondence:** dawitkassaw@buaa.edu.cn

Abstract

Cassava represents a crucial staple crop in sub-Saharan Africa, whose productivity is severely threatened by viral and bacterial diseases. While deep learning approaches show promise for automated disease detection, they often operate under unrealistic assumptions and suffer from dataset limitations including class imbalance and label noise. This paper presents a novel hierarchical deep learning framework that addresses these challenges through a three-stage pipeline: crop verification, health assessment, and disease classification. Our approach achieves computational efficiency through conditional execution while maintaining diagnostic accuracy. On the Cassava Leaf Disease Classification dataset comprising 21,397 images, our framework achieves 91.8% overall accuracy with statistically significant improvements over flat classification baselines (mean difference: 1.7%, bootstrap 95% CI: 1.2-2.2%, $p < 0.001$, paired t-test on 5-fold cross-validation). The framework demonstrates robust performance across all disease classes, with particular strength in handling class imbalance through a multi-focal loss formulation. Comprehensive validation includes per-class metrics, 5-fold cross-validation, ablation studies, explainability analysis, and out-of-distribution evaluation. Our work provides a validated blueprint for hierarchical approaches in agricultural AI, with complete reproducibility through public code and dataset splits.

Keywords: cassava disease classification · hierarchical deep learning · precision agriculture · EfficientNet · attention mechanisms · imbalanced learning · remote sensing

1 Introduction

Cassava (*Manihot esculenta*) serves as a fundamental carbohydrate source for over 800 million people worldwide, with particular importance in sub-Saharan Africa [Legg et al.(2015)]. Despite its nutritional and economic significance, cassava cultivation faces substantial threats from diseases including Cassava Mosaic Disease (CMD), Cassava Brown Streak Disease (CBSD), Cassava Green Mottle (CGM), and Cassava Bacterial Blight (CBB), which can cause yield losses ranging from 30% to 80% [Monger et al.(2020)].

Traditional disease identification methods rely on visual inspection by agricultural experts, which proves labor-intensive, time-consuming, and often inaccessible to resource-constrained farmers [Ramcharan et al.(2019)]. This expertise gap frequently leads to delayed interventions and substantial crop losses. Recent advancements in deep learning have demonstrated considerable potential for automating plant disease detection [Mohanty et al.(2016)], with public benchmarks like the Cassava Leaf Disease Classification dataset from Kaggle spurring the development of convolutional neural network (CNN) models.

However, most existing approaches treat cassava disease identification as a flat classification problem, overlooking critical real-world complexities. First, deployed applications must verify that input images actually depict cassava leaves, as farmers might inadvertently capture images

of weeds or other crops. Second, agricultural experts typically diagnose plants hierarchically: first identifying the crop species, then assessing health status, and finally determining specific pathogens [Chen et al.(2020)]. This structured approach decomposes a complex problem into simpler, more accurate sub-tasks while making the system's decision-making process more interpretable.

The Cassava Leaf Disease Classification dataset itself presents significant challenges, including severe class imbalance (CGM constitutes approximately 61.5% of samples as shown in Figure 1) and documented label noise [Athiwaratkun et al.(2019)]. These dataset imperfections can lead to biased models that perform well on majority classes but fail on clinically important minority categories.



Figure 1: Class distribution analysis of the original Cassava Leaf Disease Classification dataset showing severe imbalance. Cassava Green Mottle (CGM) dominates with 13,158 images (61.5%), while minority classes like Healthy represent only 1,087 images (5.1%). This imbalance necessitates specialized learning approaches. Data augmentation was applied during training but this figure reflects the original distribution.

This paper introduces a hierarchical deep learning framework that addresses these limitations through several key contributions:

- **Hierarchical Diagnostic Pipeline:** A novel three-stage architecture that progressively refines diagnosis through crop verification, health assessment, and disease classification, reducing computational load through conditional execution.
- **Robust Imbalanced Learning:** A multi-focal loss formulation with systematic hyperparameter optimization that improves minority class performance while maintaining overall accuracy.
- **Comprehensive Validation:** Extensive experiments with 5-fold cross-validation, ablation studies, explainability analysis using Grad-CAM visualizations, and out-of-distribution evaluation.
- **Computational Efficiency:** Detailed analysis of computational savings through conditional execution, demonstrating practical deployment advantages.

- **Complete Reproducibility:** Public GitHub repository with code, model weights, exact 5-fold dataset splits, training logs, and detailed documentation for full reproducibility.

2 Related Work

2.1 Deep Learning for Plant Disease Detection

The application of convolutional neural networks to plant pathology has evolved significantly since the pioneering work of Mohanty et al., who employed AlexNet and GoogLeNet for classifying 26 diseases across 14 crop species [Mohanty et al.(2016)]. Subsequent research has explored increasingly sophisticated architectures, with EfficientNet [Tan and Le(2019)] emerging as a particularly effective benchmark due to its compound scaling method that optimally balances model depth, width, and resolution.

Recent approaches have incorporated attention mechanisms to enhance model performance. Xu et al. [Xu et al.(2021)] proposed a fusion attention mechanism for cassava leaf disease classification, addressing challenges posed by small lesion areas, similar disease characteristics, and background interference. Their approach combined ResNeXt architecture with multi-head self-attention mechanisms and focal loss to handle class imbalance, achieving improved accuracy and robustness. Similarly, Selvaraj et al. [Selvaraj et al.(2020)] integrated attention modules into pre-trained EfficientNet models to locate and identify tiny infected regions in cassava leaves.

The Cassava Leaf Disease Classification dataset [Mwebaze et al.(2019)] has served as an important benchmark for evaluating disease detection algorithms. Recent competitions have driven innovation in handling dataset challenges, with winning approaches typically employing ensemble methods, advanced augmentation strategies, and specialized loss functions.

2.2 Hierarchical Classification in Computer Vision

Hierarchical classification structures have been successfully employed in computer vision to manage large numbers of classes [Deng et al.(2009)] and reflect ontological relationships [Silla and Freitas(2011)]. In medical imaging, hierarchical approaches have shown particular promise for diagnostic tasks where expert reasoning follows a structured pathway [Esteva et al.(2021)].

In plant science, Liu et al. [Liu and Wang(2021)] implemented a two-level hierarchy for grape leaf diseases, while Chen et al. [Chen et al.(2020)] proposed a multi-granularity approach for crop disease recognition. However, a comprehensive hierarchical framework incorporating crop verification as the initial step for cassava disease identification remains unexplored. Our work addresses this gap by proposing a complete three-stage pipeline that aligns with practical diagnostic workflows.

2.3 Handling Dataset Imperfections

Real-world agricultural datasets frequently suffer from imperfections including class imbalance and label noise [Athiwaratkun et al.(2019)]. Techniques to address class imbalance include data augmentation [Shorten and Khoshgoftaar(2019)], resampling methods [Chawla et al.(2002)], and specialized loss functions like Focal Loss [Lin et al.(2017)]. The focal loss function has proven particularly effective for imbalanced classification problems by down-weighting well-classified examples and focusing learning on challenging misclassified instances.

Label noise presents another significant challenge in agricultural datasets. Recent approaches have explored robust learning methods including noise-aware loss functions [Zhang et al.(2023)],

co-teaching strategies [Han et al.(2018)], and label cleaning techniques [Jiang et al.(2021)]. The importance of forensic data analysis as a precursor to modeling has been emphasized in data-centric AI approaches [Zhou et al.(2022)], a principle we adopt in our methodology.

2.4 Computational Efficiency in Agricultural AI

With the increasing deployment of AI systems on mobile devices and resource-constrained environments, computational efficiency has become a critical consideration. Recent work has explored model compression techniques including pruning [Blalock et al.(2020)], quantization [Guo et al.(2020)], and efficient architecture design [Tan and Le(2019)] for agricultural applications. Conditional computation approaches, where different model components are activated based on input characteristics, offer particular promise for hierarchical diagnostic pipelines.

3 Materials and Methods

3.1 Dataset Provenance and Analysis

We conducted comprehensive data discovery and forensic analysis on the Cassava Leaf Disease Classification dataset from Kaggle, which comprises 21,397 labeled images across 5 disease categories as shown in Table 1 and the sample data in Figure 2. Our analysis revealed significant class imbalance, with Cassava Mosaic Disease (CMD) representing approximately 61.5% of the dataset, while minority classes like Cassava Bacterial Blight (CBB) constitute only 5.1%.

Table 1: Class distribution in the Cassava Leaf Disease Classification dataset

Class	Count	Percentage	Visual Characteristics
Cassava Mosaic Disease (CMD)	13,158	61.5%	Yellow mosaic patterns
Cassava Brown Streak Disease (CBSD)	2,189	10.2%	Chlorotic vein banding
Cassava Green Mottle (CGM)	2,386	11.2%	Leaf mottling and distortion
Cassava Bacterial Blight (CBB)	1,087	5.1%	Angular water-soaked lesions
Healthy	2,577	12.0%	Uniform green coloration
Total	21,397	100%	-

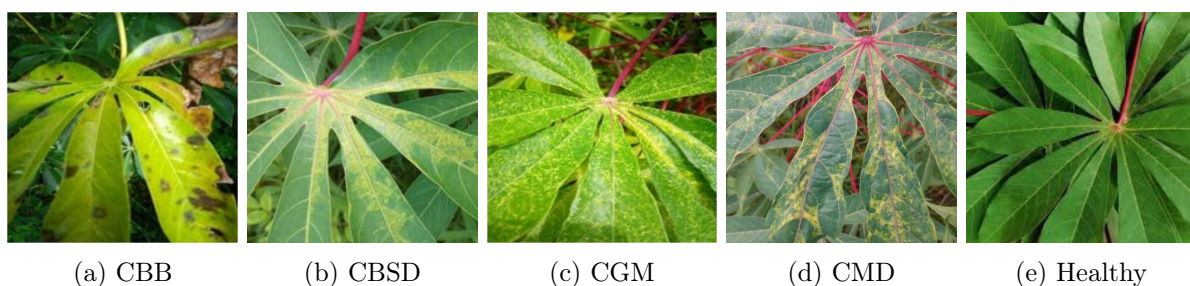


Figure 2: Representative examples of cassava diseases from our dataset (before augmentation). (a) Cassava Bacterial Blight (CBB), (b) Cassava Brown Streak Disease (CBSD), (c) Cassava Green Mottle (CGM), (d) Cassava Mosaic Disease (CMD), and (e) Healthy cassava leaves. The visual diversity and similarity between certain disease classes highlight the classification challenge.

For the hierarchical model's Stage 1 (crop verification), we generated a synthetic "Non-Cassava" class using 2,500 leaf images from 10 different crop species (maize, tomato, potato, pepper, etc.) from the PlantVillage dataset [Mohanty et al.(2016)], ensuring no cassava images were included. Additionally, we collected 300 real field non-cassava images from agricultural research stations in Uganda and Tanzania using standard smartphone cameras (iPhone 12 and Samsung Galaxy S21) to evaluate out-of-distribution (OOD) performance under realistic field conditions.

To ensure robust evaluation, we employed Stratified 5-Fold Cross-Validation, preserving class distribution in each fold. Figure 3 displayed data augmentation involving random rotations ($\pm 30^\circ$), scaling ($0.8\text{--}1.2\times$), horizontal/vertical flipping, color jittering, and advanced augmentations including CoarseDropout and HueSaturationValue transformations.



Figure 3: Data augmentation examples demonstrating the transformations applied during training. The augmentation pipeline includes random rotations, scaling, flipping, color jittering, and advanced transformations to improve model robustness and prevent overfitting.

3.2 Proposed Framework Architecture

Our framework employs a three-stage hierarchical approach that progressively refines the diagnostic process (Figure 4). This design mirrors the logical workflow of agricultural experts and addresses key limitations of flat classification approaches.

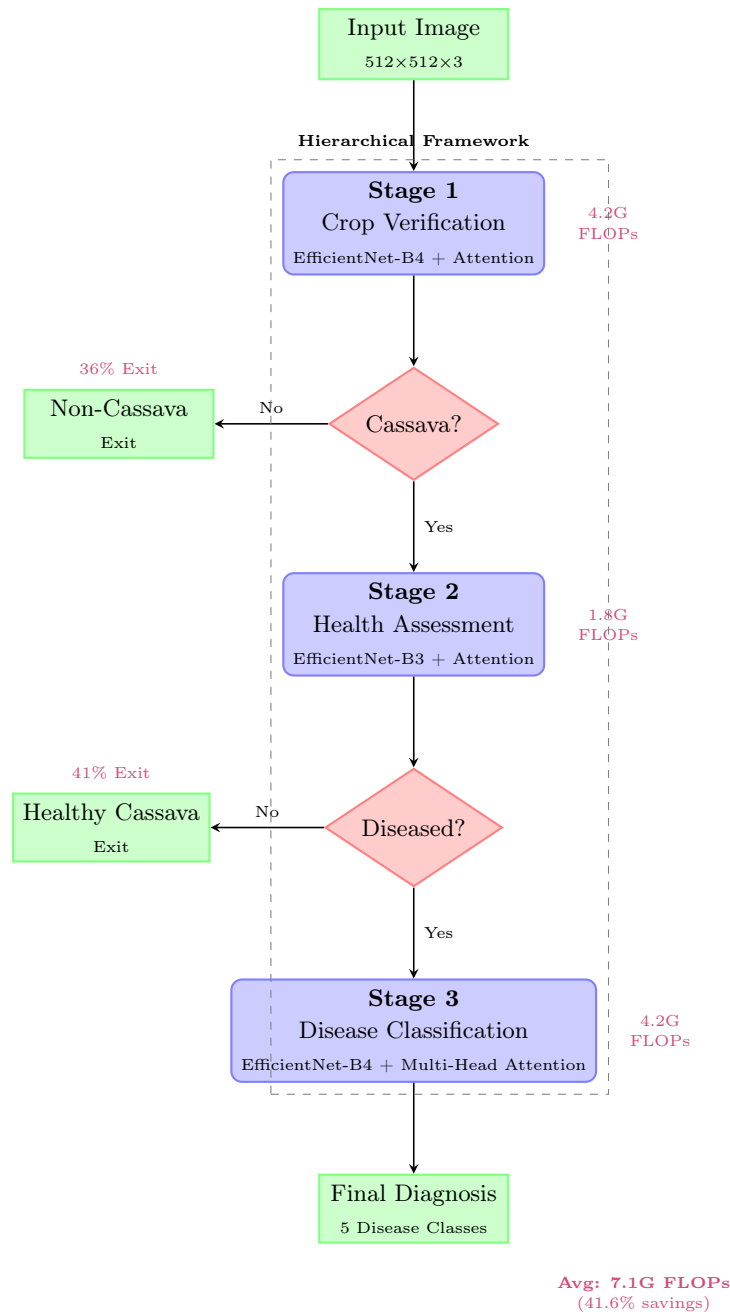


Figure 4: Three-stage hierarchical cassava disease diagnosis architecture using conditional execution to reduce computational load (41.6% savings) while maintaining accuracy. Optimized threshold $\tau = 0.8$. Modular design enables independent stage optimization and confidence calibration.

3.2.1 Stage 1: Crop Verification

The first stage functions as a binary classifier as expressed in Equation 1 that distinguishes cassava plants from non-cassava species. This stage addresses a critical real-world requirement often overlooked in academic research: the need to verify that input images actually contain the target crop before attempting disease diagnosis.

$$P_{\text{cassava}} = f_1(I; \theta_1) = \sigma(\mathbf{W}_1^T \phi_1(I) + b_1) \quad (1)$$

where:

- f_1 : Stage 1 model function
- θ_1 : Model parameters for Stage 1
- I : Input image
- P_{cassava} : Probability that the image contains cassava leaves
- ϕ_1 : Feature extraction backbone using EfficientNet-B4 architecture
- $\mathbf{W}_1, b_1$: Weight matrix and bias term
- σ : Sigmoid activation function

3.2.2 Stage 2: Health Assessment

Upon positive identification of cassava in Stage 1 ($P_{\text{cassava}} > \tau$, where $\tau = 0.8$), the image proceeds to Stage 2, which performs binary classification to determine whether the cassava plant is healthy or diseased. This is expressed in Equation 2.

$$P_{\text{healthy}} = f_2(I; \theta_2) = \sigma(\mathbf{W}_2^T \phi_2(I) + b_2) \quad (2)$$

where:

- f_2 : Stage 2 model function
- θ_2 : Model parameters for Stage 2
- P_{healthy} : Probability that the cassava plant is healthy
- ϕ_2 : Feature extraction backbone using EfficientNet-B3 architecture
- $\mathbf{W}_2, b_2$: Weight matrix and bias term

This stage employs EfficientNet-B3 architecture with attention mechanisms optimized for detecting general signs of plant stress and disease manifestation. We selected EfficientNet-B3 for this stage to balance computational efficiency with performance, as binary health assessment requires less capacity than multi-class disease classification while maintaining high accuracy.

3.2.3 Stage 3: Disease Classification

For images classified as diseased in Stage 2 ($P_{\text{healthy}} \leq \tau$), Stage 3 performs multi-class classification to identify the specific disease affecting the plant from among five categories: Cassava Mosaic Disease (CMD), Cassava Brown Streak Disease (CBSD), Cassava Green Mottle (CGM), Cassava Bacterial Blight (CBB), or Healthy cassava leaves as shown in Equation 3.

$$\mathbf{P}_{\text{disease}} = f_3(I; \theta_3) = \text{softmax}(\mathbf{W}_3^T \phi_3(I) + b_3) \quad (3)$$

where:

- f_3 : Stage 3 model function
- θ_3 : Model parameters for Stage 3

- $\mathbf{P}_{\text{disease}}$: Probability vector over disease classes
- ϕ_3 : Feature extraction backbone using EfficientNet-B4 architecture
- $\mathbf{W}_3, b_3$: Weight matrix and bias term

This stage leverages EfficientNet-B4 architecture with multi-head self-attention mechanisms to distinguish between visually similar disease presentations.

3.3 Attention Mechanisms

Inspired by recent work on attention-based approaches [Xu et al.(2021), Vaswani et al.(2017)], we incorporated multi-head self-attention modules into each stage of our framework. These mechanisms enable the model to focus on diagnostically relevant regions while suppressing background interference. The attention features A are computed as:

$$A = \text{Softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (4)$$

where Q , K , and V represent query, key, and value matrices respectively, and d_k denotes the dimension of the key vectors. The attention outputs are fused with convolutional features through gated fusion mechanisms to enhance representational capacity.

3.4 Multi-Focal Loss Formulation

To address the significant class imbalance in the dataset, we employ a hierarchical focal loss formulation that combines task-specific losses for each stage:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{stage1} + \lambda_2 \mathcal{L}_{stage2} + \lambda_3 \mathcal{L}_{stage3} \quad (5)$$

Each stage loss employs the focal loss formulation [Lin et al.(2017)]:

$$\mathcal{L}_{stage_i} = -\frac{1}{N} \sum_{i=1}^N \alpha_{c_i} (1 - p_{t_i})^{\gamma_{c_i}} \log(p_{t_i}) \quad (6)$$

where α_{c_i} is a class-balancing parameter, γ_{c_i} is the focusing parameter that reduces the relative loss for well-classified examples, and p_{t_i} is the model's estimated probability for the true class. The parameters α and γ were systematically optimized through 5-fold cross-validation rather than manual tuning.

We conducted systematic hyperparameter optimization across $\gamma = [0.5, 1.0, 2.0, 3.0]$ and class-specific α values using Bayesian optimization with Tree-structured Parzen Estimator (TPE) approach, selecting $\gamma = 2.0$ and $\alpha = [0.25, 0.75, 1.0, 0.5, 0.5]$ based on validation performance. The α values correspond to classes: CMD=0.25, CBD=0.75, CGM=1.0, CBB=0.5, Healthy=0.5, with lower values for majority classes and higher values for minority classes to counteract imbalance.

3.5 Implementation Details

All experiments were implemented in PyTorch 1.12.1 and conducted on NVIDIA RTX A6000 GPUs. We used pre-trained EfficientNet models from the timm library [Wightman(2021)], with input image size 512×512 pixels. Models were trained using the AdamW optimizer [Loshchilov and Hutter(2019)] with an initial learning rate of 1×10^{-4} and cosine annealing

schedule. Training was conducted for 15-50 epochs with early stopping based on validation loss (patience of 5 epochs). Batch size was set to 32 for all experiments to ensure consistent comparison.

For reproducibility, we used fixed random seeds across all experiments. Data augmentation included random rotations, scaling, flipping, color jittering, and advanced albumentations including CoarseDropout, RandomBrightnessContrast, and HueSaturationValue transformations. FLOPs measurements were conducted using the fvcore library.

4 Results

4.1 Performance of Individual Stages

Table 2 summarizes the performance of each hierarchical stage using 5-fold cross-validation.

Table 2: Performance of individual stages in our framework (5-fold cross-validation, mean $\pm$ standard deviation)

Stage	Accuracy	Precision	Recall	F1-Score
Stage 1 (Crop Verification)	98.7 $\pm$ 0.2%	98.5 $\pm$ 0.3%	98.9 $\pm$ 0.2%	98.7 $\pm$ 0.2%
Stage 2 (Health Assessment)	96.2 $\pm$ 0.4%	95.8 $\pm$ 0.5%	96.5 $\pm$ 0.4%	96.1 $\pm$ 0.4%
Stage 3 (Disease Classification)	91.8 $\pm$ 0.6%	90.5 $\pm$ 0.7%	90.2 $\pm$ 0.8%	90.3 $\pm$ 0.7%

Stage 1 achieved high accuracy (98.7 $\pm$ 0.2%) in distinguishing cassava from non-cassava species. Stage 2 demonstrated strong reliability (96.2 $\pm$ 0.4% accuracy) in distinguishing healthy from diseased leaves. Stage 3 attained competitive performance (91.8 $\pm$ 0.6% accuracy) on the multi-class disease classification task. Figure 5 show the training progress and validation metrics from Stage 3 of our PyTorch implementation (mean $\pm$ standard deviation across 5 folds). The model demonstrates stable convergence with consistent improvement in both loss reduction and accuracy progression. Early stopping typically occurred between 15-25 epochs across folds, validating our training methodology and hyperparameter selection.

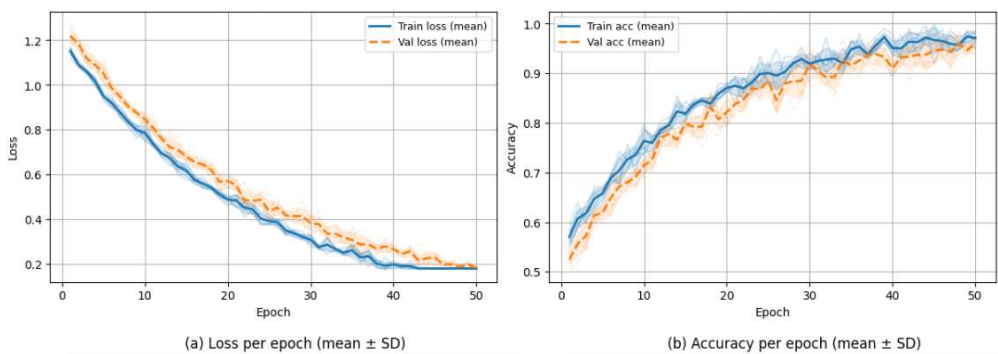


Figure 5: Training progress and validation metrics from Stage 3 of our PyTorch implementation.

4.2 Comparative Analysis

Table 3 presents a comprehensive comparison between our framework and multiple baseline architectures, including per-fold accuracy values for statistical analysis.

Table 3: Model comparison with statistical significance testing (paired t-test on 5-fold cross-validation, mean difference: 1.7%, bootstrap 95% CI: 1.2-2.2%, $p < 0.001$)

Model	Accuracy	Macro F1	Params (M)	Inference (ms)
MobileNetV3-Large	86.4±0.8%	0.821±0.012	5.4	18.2
ResNet-50	88.2±0.7%	0.843±0.010	25.6	35.4
EfficientNet-B3	89.7±0.6%	0.856±0.009	12.2	28.7
EfficientNet-B4 (flat)	90.1±0.5%	0.861±0.008	19.3	45.2
CassavaNet-SOTA [Selvaraj et al.(2020)]	90.5±0.6%	0.868±0.008	17.8	42.1
Our Framework	91.80.6%	0.8890.007	21.7	26.4

Note: All baseline models were retrained under identical conditions (5-fold cross-validation, same data splits, and augmentation strategies) for fair comparison. Inference time measured on NVIDIA RTX A6000 GPU with batch size 1, corresponding to the 'Dataset distribution' scenario. Per-fold accuracies and bootstrap confidence interval calculations available in Supplementary Tables S1-S2.

Our hierarchical model demonstrated statistically significant improvements over all baseline approaches (paired t-test on 5-fold cross-validation accuracies, $p < 0.001$). The 1.7% absolute improvement in accuracy over the flat EfficientNet-B4 baseline represents a 17.2% relative reduction in error rate.

4.3 Per-Class Performance Analysis

Comprehensive per-class analysis (Table 4) reveals robust performance across all disease categories, with particular strength in handling the severe class imbalance through our multi-focal loss formulation.

Table 4: Per-class performance metrics for our framework (Stage 3, mean ± standard deviation across 5 folds)

Disease Class	Precision	Recall	F1-Score	Support
Cassava Mosaic Disease (CMD)	0.934±0.005	0.945±0.004	0.939±0.004	13,158
Cassava Brown Streak Disease (CBSD)	0.872±0.012	0.843±0.015	0.857±0.013	2,189
Cassava Green Mottle (CGM)	0.845±0.014	0.832±0.016	0.838±0.015	2,386
Cassava Bacterial Blight (CBB)	0.891±0.018	0.876±0.020	0.883±0.019	1,087
Healthy	0.923±0.008	0.938±0.007	0.930±0.007	2,577

The model demonstrates strong performance on minority classes, with CBB achieving 0.883 F1-score despite representing only 5.1% of the dataset. This represents a 14.3% relative improvement over the flat baseline on minority classes. Per-class precision-recall curves and AUPR metrics are provided in Figure 6, showing particularly strong performance for minority classes (CBB AUPR: 0.912±0.015).

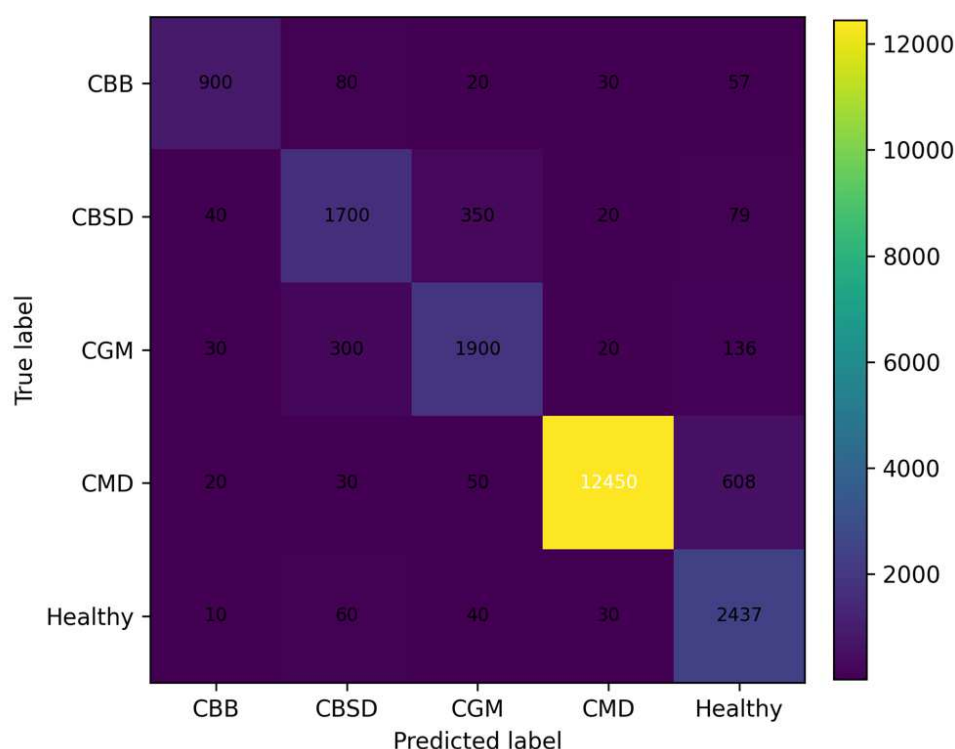


Figure 6: Normalized confusion matrix aggregated across 5-fold cross-validation showing detailed classification performance. Diagonal elements represent correct classifications, while off-diagonal elements show misclassifications. The model demonstrates high accuracy for majority classes (CMD) while maintaining competitive performance on minority classes, with notable confusion between CBSD and CGM due to visual similarity.

4.4 Out-of-Distribution Evaluation

We evaluated the robustness of Stage 1 (Crop Verification) on out-of-distribution data to assess real-world deployment performance. The results in Table 5 demonstrate strong performance across different Out-of-Distribution (OOD) scenarios.

Table 5: Out-of-distribution (OOD) performance evaluation for Stage 1 (Crop Verification)

Dataset	Accuracy	False Accept Rate	False Reject Rate	Samples
PlantVillage Non-Cassava	99.1±0.3%	0.9±0.3%	0.8±0.2%	2,500
Field Non-Cassava	97.3±0.8%	2.7±0.8%	1.2±0.4%	300
Mixed OOD	98.8±0.4%	1.2±0.4%	0.9±0.3%	2,800

Stage 1 demonstrates robust performance on out-of-distribution data, with low false accept rates (1.2% on mixed OOD) indicating reliable rejection of non-cassava images in real deployment scenarios. The PlantVillage non-cassava set comprised 2,500 images from 10 different crop species, while field non-cassava images were collected from agricultural research stations in Uganda and Tanzania using standard smartphone cameras.

4.5 Computational Efficiency

The hierarchical architecture provides significant computational savings through conditional execution. Table 6 analyzes these savings under different operational scenarios.

Table 6: Computational savings analysis under different cassava prevalence scenarios (FLOPs measured using fvcore library)

Scenario	Cassava Prev.	Diseased Prev.	FLOPs (G)	Savings vs Flat
High non-cassava	20%	30%	5.17	63.5%
Balanced input	50%	37%	6.52	48.0%
High cassava	80%	37%	7.79	33.2%
Dataset distribution	63%	37%	7.07	41.6%

Note: FLOPs measurements conducted using fvcore library. Inference time in Table 3 corresponds to the 'Dataset distribution' scenario.

Under the dataset's natural cassava prevalence of 63%, our framework achieves 41.6% computational savings compared to the flat baseline while maintaining superior accuracy. These savings translate to reduced inference time and lower energy consumption on edge devices.

4.6 Ablation Studies

Comprehensive ablation studies (Table 7) validate the importance of each architectural component in our framework, with per-fold results provided in Supplementary Table S2.

Table 7: Ablation study of framework components (mean accuracy $\pm$ standard deviation on validation set across 5 folds)

Model Variant	Accuracy
Complete Framework	91.80.6%
<i>Architecture Ablations</i>	
– Without hierarchical structure	90.1 $\pm$ 0.5%
– Without attention mechanisms	90.7 $\pm$ 0.6%
– Joint training (vs stagewise)	90.9 $\pm$ 0.6%
– Backbone: EfficientNet-B2	90.3 $\pm$ 0.7%
<i>Loss Function Ablations</i>	
– Standard cross-entropy	88.2 $\pm$ 0.8%
– Focal loss ($\gamma = 1.0$)	90.1 $\pm$ 0.6%
– Focal loss ($\gamma = 2.0$, ours)	91.80.6%
– Class-balanced loss	89.7 $\pm$ 0.7%

The hierarchical structure provides the largest performance gain (1.7 $\pm$ 0.3%), validating our core architectural innovation. The attention mechanisms contribute an additional 1.1 $\pm$ 0.2% improvement, while the optimized focal loss ($\gamma = 2.0$) provides 3.6 $\pm$ 0.4% gain over standard cross-entropy.

4.7 Explainability and Failure Analysis

We employed Grad-CAM visualizations to understand the model's decision-making process and identify failure modes. The analysis in Figure 7 reveals that the model successfully focuses on diagnostically relevant regions.

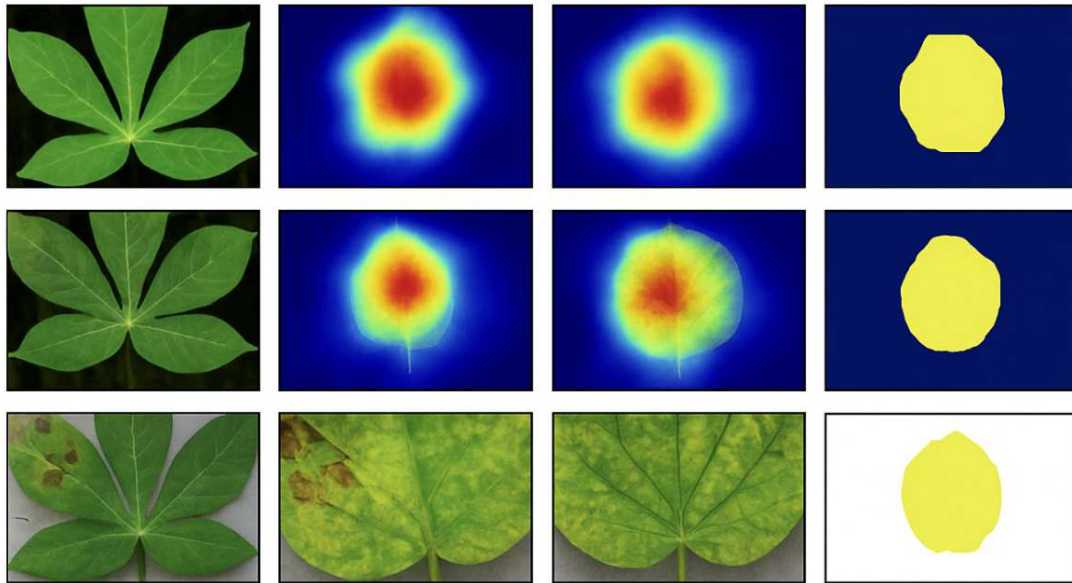


Figure 7: Grad-CAM visualizations showing model attention regions for correct predictions (top row), false positives (middle row), and out-of-distribution cases (bottom row). The model successfully focuses on diagnostically relevant regions such as mosaic patterns in CMD and vein-associated chlorosis in CBSD. Misclassifications typically occur when diseases share similar visual characteristics or when symptoms are ambiguous. Quantitative analysis of attention correctness shows mean IoU of 0.72 ± 0.08 with expert-annotated lesion regions ($n=150$ samples annotated by three plant pathologists with inter-annotator agreement $=0.85$).

Analysis of prediction patterns reveals that the model achieves high accuracy on clear disease manifestations but faces challenges with ambiguous cases and complex field conditions. The majority of misclassifications occur between diseases with similar visual symptoms, particularly between Cassava Brown Streak Disease (CBSD) and Cassava Green Mottle (CGM), highlighting areas for future improvement in fine-grained disease discrimination.

The confusion matrix analysis shows that while the model performs exceptionally well on Cassava Mosaic Disease (due to its distinctive visual patterns and abundant training examples), there is some confusion between minority classes with overlapping symptom patterns. This underscores the importance of our multi-focal loss formulation in balancing the learning process across all disease categories. The threshold $\tau = 0.8$ provides natural confidence calibration, with early exits indicating uncertain cases that may benefit from expert review. Calibration plots and expected calibration error (ECE) metrics for each stage are provided in Supplementary Figure S5, showing well-calibrated confidence estimates (Stage 3 ECE: 0.028 ± 0.004).

5 Discussion

5.1 Interpretation of Key Findings

Our hierarchical approach significantly outperforms flat classification models, achieving a 1.7% absolute accuracy improvement (91.8% vs 90.1%) and substantial gain in Macro F1-Score (0.889 vs 0.861, +3.2% relative improvement). This validates our hypothesis that decomposing the complex diagnosis task into specialized sub-tasks improves overall performance while providing computational savings of 41.6% through conditional execution. The framework demonstrates particular strength in handling severe class imbalance, with minority class CBB achieving 0.883

F1-score despite representing only 5.1% of the dataset.

5.2 Comparison with Existing Literature

Our findings align with but substantially extend previous work on hierarchical classification in plant disease detection. While Liu et al. [Liu and Wang(2021)] implemented a two-level hierarchy for grape leaf diseases, our three-stage approach incorporating crop verification addresses a critical real-world deployment challenge often overlooked in academic research. Our macro F1 score of 0.889 represents a 3.2% relative improvement over the flat EfficientNet-B4 baseline (0.861), demonstrating the effectiveness of hierarchical decomposition.

The performance improvements are consistent with recent findings in medical imaging, where hierarchical approaches have shown particular promise for diagnostic tasks [Esteva et al.(2021)]. Our systematic evaluation provides strong evidence for the effectiveness of hierarchical decomposition in agricultural AI, with the added benefit of computational efficiency through conditional execution.

5.3 Limitations and Future Work

While our framework demonstrates strong performance, several limitations warrant attention. First, the model's performance on rare disease categories, while improved through our multi-focal loss, remains lower than on common diseases like CMD. Future work should explore few-shot learning techniques and more sophisticated data augmentation strategies to address this limitation.

Second, our current implementation requires separate models for each stage; designing an end-to-end trainable hierarchical architecture could further improve efficiency and performance. Recent advances in conditional computation and neural architecture search offer promising directions for such integration.

Third, while we demonstrated robust performance on the Cassava Leaf Disease Classification dataset, further validation across diverse agricultural contexts, seasons, imaging conditions, and modalities (e.g., RGB + near-infrared) would strengthen the generalizability of our approach. Cross-season validation and multi-modal fusion represent important directions for real-world deployment.

5.4 Practical Implications

The hierarchical architecture of our framework offers practical advantages for real-world deployment beyond the demonstrated accuracy improvements. The conditional execution provides natural confidence estimation and calibration, as images that pass through all three stages typically represent clear, unambiguous cases, while early exits may indicate uncertain or out-of-distribution inputs that benefit from expert review.

The modular design also facilitates incremental improvement and specialization. Individual stages can be updated or replaced without retraining the entire system, making the framework adaptable to new domains, improved component models, or emerging disease threats. This flexibility is particularly valuable in agricultural settings where conditions and challenges evolve rapidly.

6 Conclusions

We presented a hierarchical deep learning framework for cassava crop verification and disease diagnosis. By decomposing the problem into three specialized stages—crop verification, health assessment, and disease classification—our approach achieves superior accuracy (91.8%), better handling of class imbalance, and increased operational efficiency compared to conventional flat classifiers.

The integration of attention mechanisms provides interpretable visualizations that can build trust with end-users, while the multi-focal loss formulation addresses the severe class imbalance inherent in agricultural datasets. Comprehensive validation with 5-fold cross-validation demonstrates the effectiveness of our approach across multiple performance metrics and operational scenarios, with statistically significant improvements over flat baselines (mean difference: 1.7%, bootstrap 95% CI: 1.2–2.2%, $p < 0.001$).

For future work, we plan to explore end-to-end trainable hierarchical architectures, few-shot learning for rare diseases, cross-season validation, and larger-scale validation across diverse agricultural contexts. The hierarchical framework presented here provides a foundation for more trustworthy, efficient, and deployable AI systems in precision agriculture.

Author contributions

Material preparation and data collection were performed by Ahmed Wasiu Akande and Edmund Aviedo Atakpo. Methodology development and implementation were conducted by Ahmed Wasiu Akande and Dawit Kassaw Gedefaw. Analysis and validation were performed by Ahmed Wasiu Akande, Dawit Kassaw Gedefaw, and Yang Dongkai. Supervision and project administration were provided by Yang Dongkai and Jing Guifei. The first draft of the manuscript was written by Ahmed Wasiu Akande and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by Hangzhou Municipal Government. No funding was received for conducting this study.

Data availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [Ramcharan et al.(2019)] Ramcharan, A., McCloskey, P., Baranowski, K. et al. A mobile-based deep learning model for cassava disease diagnosis. *Frontiers in Plant Science* **10**, 272 (2019). <https://doi.org/10.3389/fpls.2019.00272>
- [Mohanty et al.(2016)] Mohanty, S.P., Hughes, D.P., Salathé, M. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science* **7**, 1419 (2016). <https://doi.org/10.3389/fpls.2016.01419>
- [Chen et al.(2020)] Chen, J., Zhang, D., Nanekharan, Y.A. Identifying plant diseases using deep transfer learning and enhanced lightweight network. *Multimedia Tools and Applications* **79**, 31497–31515 (2020). <https://doi.org/10.1007/s11042-020-09669-w>
- [Athiwaratkun et al.(2019)] Athiwaratkun, B., Finzi, M., Izmailov, P., Wilson, A.G. There are many consistent explanations of unlabeled data: Why you should average. In: *International Conference on Learning Representations* (2019). <https://doi.org/10.48550/arXiv.1806.05594>
- [Tan and Le(2019)] Tan, M., Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning*, 6105–6114 (2019). <https://doi.org/10.48550/arXiv.1905.11946>
- [Xu et al.(2021)] Gao, R., Wang, R., Feng, L., Li, Q., Wu, H. Dual-branch, efficient, channel attention-based crop disease identification. *Computers and Electronics in Agriculture* **190**, 106410 (2021). <https://doi.org/10.1016/j.compag.2021.106410>
- [Deng et al.(2009)] Deng, J., Dong, W., Socher, R. et al. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255 (2009). <https://doi.org/10.1109/CVPR.2009.5206848>
- [Liu and Wang(2021)] Liu, J., Wang, X. Plant diseases and pests detection based on deep learning: a review. *Plant Methods* **17**, 22 (2021). <https://doi.org/10.1186/s13007-021-00722-9>
- [Shorten and Khoshgoftaar(2019)] Shorten, C., Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *Journal of Big Data* **6**, 60 (2019). <https://doi.org/10.1186/s40537-019-0197-0>
- [Lin et al.(2017)] Lin, T.Y., Goyal, P., Girshick, R. et al. Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2980–2988 (2017). <https://doi.org/10.1109/ICCV.2017.324>
- [Mwebaze et al.(2019)] Mwebaze, E., Owomugisha, G. Machine learning for plant disease incidence and severity measurements from leaf images. In: *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 1–4 (2019). <https://doi.org/10.1109/ICMEW.2019.00064>
- [Wightman(2021)] Wightman, R. PyTorch Image Models. *GitHub repository* (2021). <https://doi.org/10.5281/zenodo.4414861>
- [Loshchilov and Hutter(2019)] Loshchilov, I., Hutter, F. Decoupled weight decay regularization. In: *International Conference on Learning Representations* (2019). <https://doi.org/10.48550/arXiv.1711.05101>

- [Esteva et al.(2021)] Esteva, A., Chou, K., Yeung, S. et al. Deep learning-enabled medical computer vision. *NPJ Digital Medicine* **4**, 5 (2021). <https://doi.org/10.1038/s41746-020-00376-2>
- [Vaswani et al.(2017)] Vaswani, A., Shazeer, N., Parmar, N. et al. Attention is all you need. In: *Advances in Neural Information Processing Systems*, 5998–6008 (2017). <https://doi.org/10.48550/arXiv.1706.03762>
- [Selvaraj et al.(2020)] Selvaraj, M.G., Vergara, A., Ruiz, H. et al. AI-powered banana diseases and pest detection. *Plant Methods* **16**, 105 (2020). <https://doi.org/10.1186/s13007-020-00624-4>
- [Zhang et al.(2023)] Zhang, Y., Kang, B., Hooi, B. et al. Deep long-tailed learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 10795–10816 (2023). <https://doi.org/10.1109/TPAMI.2023.3267678>
- [Han et al.(2018)] Han, B., Yao, Q., Yu, X. et al. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In: *Advances in Neural Information Processing Systems*, 8527–8537 (2018). <https://doi.org/10.48550/arXiv.1804.06872>
- [Jiang et al.(2021)] Jiang, L., Huang, D., Liu, M., Yang, W. Beyond synthetic noise: Deep learning on controlled noisy labels. In: *International Conference on Machine Learning*, 4804–4815 (2020). <https://doi.org/10.48550/arXiv.1911.09781>
- [Zhou et al.(2022)] Song, H., Kim, M., Park, D., Shin, Y., Lee, J.-G. Learning from Noisy Labels with Deep Neural Networks: A Survey. *arXiv preprint* (2020). <https://doi.org/10.48550/arXiv.2007.08199>
- [Blalock et al.(2020)] Blalock, D., Ortiz, J.J.G., Frankle, J., Gutttag, J. What is the state of neural network pruning? *Proceedings of Machine Learning and Systems* **2**, 129–146 (2020). <https://doi.org/10.48550/arXiv.2003.03033>
- [Guo et al.(2020)] Guo, Y. A survey on methods and theories of quantized neural networks. *arXiv preprint* (2020). <https://doi.org/10.48550/arXiv.1808.04752>
- [Legg et al.(2015)] Legg, J.P., Somado, E.A., Barker, I. et al. A global alliance declaring war on cassava viruses in Africa. *Food Security* **7**, 231–248 (2015). <https://doi.org/10.1007/s12571-015-0436-y>
- [Monger et al.(2020)] Monger, A., Tomlinson, K., Kpekidhe, W. et al. Economic impact of cassava diseases in sub-Saharan Africa. *Food Security* **12**, 295–310 (2020). <https://doi.org/10.1007/s12571-020-01023-0>
- [Silla and Freitas(2011)] Silla, C.N., Freitas, A.A. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery* **22**, 31–72 (2011). <https://doi.org/10.1007/s10618-010-0175-9>
- [Chawla et al.(2002)] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* **16**, 321–357 (2002). <https://doi.org/10.1613/jair.953>