

Grad-CAM-Assisted Adaptive Scheme for Uplink Traffic Reduction in Plant Health Screening Services

Jiun-Yu Tu, Chi-Yuan Yeh

Dept. of Computer Science and Information Engineering
National Formosa University
Yunlin, Taiwan
jiunyu.tu@nfu.edu.tw, 41243158@nfu.edu.tw

Chih-Lin Hu

Dept. of Communication Engineering
National Central University
Taoyuan, Taiwan
clhu@ce.ncu.edu.tw

Abstract—Cloud-based plant health monitoring systems typically transmit full plant images from edge cameras to remote servers for inference and inspection. Although this design is straightforward, it may introduce unnecessary uplink traffic, transmission delay, and storage overhead in long-term monitoring scenarios. This paper presents a Gradient-weighted Class Activation Mapping (Grad-CAM)-assisted adaptive upload framework for reducing uplink traffic in edge plant health screening. The proposed framework employs a YOLO26x-cls model at the edge to perform plant health screening and conditionally generate Grad-CAM heatmaps for Region of Interest (ROI) proposal extraction. Based on the prediction confidence and Grad-CAM response, the edge device adaptively determines whether to upload only metadata, compact ROI evidence, or the full image. A proof-of-concept evaluation using the PlantVillage dataset is conducted to analyze binary screening performance, edge latency, payload reduction, and estimated uplink transmission delay. Experimental results show that the proposed adaptive upload strategy reduces the average upload payload from 13.64 KB to 2.82 KB, with a per-image average saving of 77.71% compared with full-image upload.

Index Terms—Edge AI, plant health monitoring, Grad-CAM, ROI, adaptive upload, YOLO26x-cls.

I. INTRODUCTION

Visual plant health monitoring is an important component in intelligent agriculture and edge-assisted plant disease management systems [1]–[3]. It is also relevant to smart retail and high-value ornamental plant monitoring, where periodic image acquisition can support early screening of symptoms such as discoloration, spots, or local degradation. As camera nodes and edge devices become more common, plant monitoring systems are expected to support long-term and multi-point image acquisition in practical environments.

A common design is to connect edge cameras or sensing nodes to cloud-assisted backends for model inference, visualization, storage, and user access [2], [4]. However, repeatedly uploading full images may introduce uplink traffic, delay, and storage overhead, especially in long-term monitoring scenarios [5]. Therefore, reducing uplink traffic while preserving

useful inspection evidence is a key issue for plant monitoring systems with constrained uplink bandwidth.

Edge Artificial Intelligence (Edge AI) can reduce unnecessary transmission by performing initial screening near the image source. Instead of uploading every image, an edge device can first estimate the plant health condition and then decide what information should be transmitted. However, transmitting only a label or confidence score may be insufficient for remote inspection. Compact visual evidence, such as Grad-CAM-guided ROI crops, can support remote review while avoiding full-image transmission for every sample.

In this paper, we propose a Grad-CAM-assisted adaptive upload framework for uplink traffic reduction in edge plant health screening services. The framework applies YOLO26x-cls at the edge for binary health screening and conditionally generates Grad-CAM heatmaps for ROI proposal extraction. Based on the screening confidence and Grad-CAM response, the proposed policy selects metadata-only reporting, ROI-crop upload, or full-image fallback. The objective is not to replace expert diagnosis, but to reduce unnecessary upload payload while preserving visual evidence for suspicious plant health conditions.

The main contributions of this paper are summarized as follows.

- We propose an edge-assisted plant health screening framework that integrates YOLO26x-cls inference, conditional Grad-CAM generation, and adaptive upload control.
- We design a confidence- and Grad-CAM-guided upload policy that selects metadata-only reporting, ROI-crop upload, or full-image fallback.
- We evaluate the proposed framework on PlantVillage images in terms of binary screening, edge latency, upload payload reduction, and estimated uplink delay.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Edge-Assisted Plant Monitoring Model

We consider an edge-assisted plant health monitoring system consisting of a camera node, an edge device, a wireless

Corresponding author: Chih-Lin Hu

TABLE I
PREDEFINED UPLOAD MODES.

Mode	Content	Purpose
M_{msg}	Metadata	Confident healthy case
M_{roi}	ROI crop	Localized visual evidence
M_{full}	Full image	Uncertain fallback

uplink, and a cloud server. The camera node periodically captures plant images. The edge device performs local screening and adaptive upload control, while the cloud server receives the uploaded information for visualization, storage, or human inspection.

Let x denote an input image captured by the camera node. The edge model produces a predicted class $\hat{y}$ and a confidence score δ . When needed, a Grad-CAM heatmap $\mathcal{H}$ is generated internally for ROI proposal extraction. The edge device then selects an upload mode $m(x) \in \mathcal{M}$, where $\mathcal{M} = \{M_{\text{msg}}, M_{\text{roi}}, M_{\text{full}}\}$, as summarized in Table I.

B. Communication Cost Model

Let S_m denote the payload size of upload mode $m \in \mathcal{M}$, and let S_{full} denote the payload size of the full-image upload baseline. For an uplink data rate R , the estimated transmission delay of mode m is

$$\tau(m) = 8 \times \frac{S_m}{R}, \quad (1)$$

where S_m is measured in bytes and R is measured in bits per second. The payload saving ratio of mode m is defined as:

$$\rho(m) = 1 - \frac{S_m}{S_{\text{full}}}. \quad (2)$$

For adaptive upload, let $P(m)$ denote the trigger probability of upload mode m . The average upload payload is defined as:

$$\bar{S} = \sum_{m \in \mathcal{M}} P(m) S_m. \quad (3)$$

C. Problem Formulation

The objective is to reduce the average upload payload while preserving sufficient visual evidence for remote inspection. In this work, the edge device selects an upload mode $m(x) \in \mathcal{M}$ for each input image x . The adaptive upload problem is formulated as

$$\min_{m(\cdot)} \bar{S}, \quad (4)$$

subject to preserving visual evidence for suspicious or uncertain plant health conditions.

In the proposed design, visual evidence is transmitted through ROI crops or full-image upload. Grad-CAM is generated only as an internal mechanism for visual explanation and ROI proposal extraction, and is not treated as an independent upload mode.

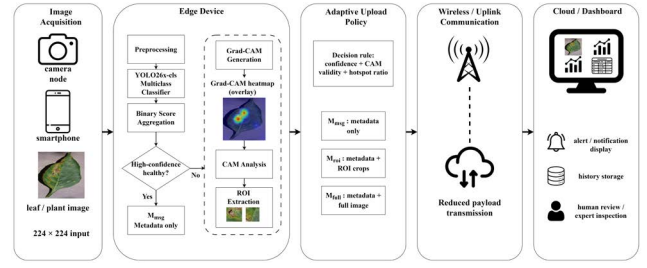


Fig. 1. Proposed Grad-CAM-assisted adaptive upload architecture for edge plant health screening. The edge device performs local inference, conditionally generates visual evidence, and adaptively selects the upload mode to reduce uplink traffic.

III. PROPOSED GRAD-CAM-ASSISTED ADAPTIVE UPLOAD FRAMEWORK

A. Overview

The proposed framework consists of three stages: edge-side plant health screening, conditional Grad-CAM-based ROI extraction, and adaptive upload decision. Fig. 1 illustrates the overall architecture. Given an input image, the YOLO26x-clas model first produces the plant health prediction and confidence score δ . Grad-CAM is generated only for selected diseased candidates to support ROI proposal extraction. The upload policy then determines whether metadata, ROI crops, or the full image should be transmitted. The framework is designed for screening and evidence reduction rather than final disease diagnosis. The edge model is used to identify suspicious cases and generate compact visual evidence. The final interpretation can still be performed by a cloud-side system or by a human inspector when necessary.

B. Edge-Side YOLO26x-clas Screening

Given an input image x , the edge-side classifier $f_\theta(\cdot)$ produces a predicted class $\hat{y}$ and a confidence score δ ,

$$\hat{y}, \delta = f_\theta(x). \quad (5)$$

In this study, YOLO26x-clas is used as the edge classifier. Its original multiclass outputs are aggregated into binary healthy and diseased scores for screening-oriented upload control. The resulting confidence score is used by the adaptive policy to determine whether the image can be represented by metadata, ROI crops, or the full image.

C. Grad-CAM Heatmap Generation

Grad-CAM [6] is generated only for selected diseased candidates that require compact visual evidence. Let $P(i)$ denote the softmax probability of class i , and let $\mathcal{C}_d$ denote the set of diseased class indices. For a diseased candidate that requires Grad-CAM, the target class is selected as:

$$i^* = \arg \max_{i \in \mathcal{C}_d} P(i). \quad (6)$$

The corresponding class-specific heatmap is then computed as:

$$H_{i^*}(x) = \text{GradCAM}(f_\theta, x, i^*), \quad (7)$$

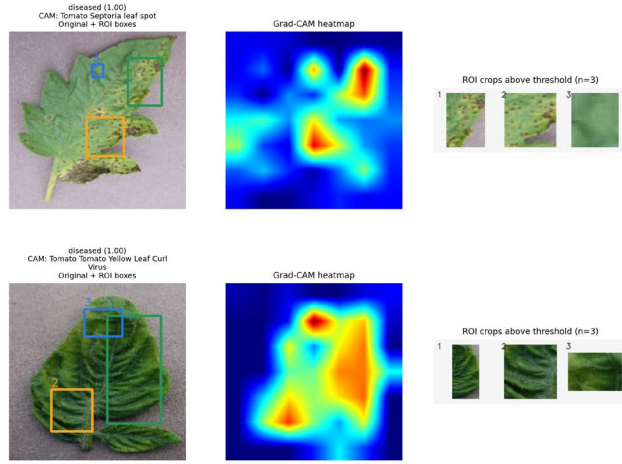


Fig. 2. Examples of Grad-CAM-assisted ROI proposal generation. The left column shows the original image with ROI boxes, the middle column shows the Grad-CAM heatmap, and the right column shows the extracted ROI crops above the activation threshold. The ROI crops are used as compact visual evidence rather than ground-truth lesion localization.

where $H_{i^*}(x)$ denotes the Grad-CAM heatmap generated for input image x with respect to the selected diseased target class i^* . The heatmap highlights the regions that contribute to the selected diseased class and is used internally for ROI proposal extraction. It is treated as attention-guided evidence rather than ground-truth lesion localization.

D. Heatmap-Based ROI Proposal Extraction

For samples that require compact visual evidence, the Grad-CAM heatmap is converted into ROI proposals. Let Ω_H denote the spatial domain of the heatmap and let λ_H denote the activation threshold. The set of activated heatmap locations is defined as:

$$\mathcal{A}_H = \{(u, v) \in \Omega_H \mid H(u, v) > \lambda_H\}. \quad (8)$$

Connected components are extracted from $\mathcal{A}_H$, and each component is mapped back to the original image to form an ROI crop. Small components are removed using a minimum-area constraint. When multiple components are detected, the top K ROI crops are retained according to component area or average activation intensity. These ROI crops are used as compact visual evidence for upload, as illustrated in Fig. 2.

E. Conservative Adaptive Upload Policy

The proposed policy uses the screening confidence and the compactness of the Grad-CAM response to select the upload mode. Based on the activated region defined in Eq. (8), the heatmap activation ratio is computed as

$$\gamma_H = \frac{|\mathcal{A}_H|}{|\Omega_H|}. \quad (9)$$

A smaller γ_H indicates a more localized response and is therefore more suitable for ROI-crop upload.

Let δ_h and δ_d denote the binary confidence scores of the healthy and diseased categories, respectively. Let τ_h , τ_d , and τ_γ denote the corresponding decision thresholds. We define $C_{\text{roi}}(x)$ as a Boolean condition that is true when at least one ROI crop remains after CAM thresholding and component filtering, and the heatmap response is sufficiently localized, i.e., $\gamma_H \leq \tau_\gamma$. The upload mode is selected as:

$$m(x) = \begin{cases} M_{\text{msg}}, & \delta_h \geq \tau_h, \\ M_{\text{roi}}, & (\delta_d \geq \tau_d) \wedge C_{\text{roi}}(x), \\ M_{\text{full}}, & \text{otherwise.} \end{cases} \quad (10)$$

Thus, confident healthy samples are uploaded as metadata, confident diseased samples with valid and localized CAM responses are uploaded as ROI crops, and uncertain or diffuse cases fall back to full-image upload.

F. Evidence-Preserving Evaluation Metrics

In addition to classification and latency metrics, the proposed framework is evaluated using communication- and evidence-oriented metrics. The average payload, payload saving ratio, and estimated uplink delay follow the communication cost model in Section II.

Let N_d denote the number of diseased images, N_{dv} denote the number of diseased images uploaded with visual evidence, and N_{dm} denote the number of diseased images uploaded as metadata only. In this work, visual evidence includes ROI-crop upload and full-image fallback. The evidence recall and unsafe metadata-only rate are defined as:

$$R_{\text{ev}} = \frac{N_{dv}}{N_d}, \quad R_{\text{um}} = \frac{N_{dm}}{N_d}. \quad (11)$$

A higher R_{ev} and a lower R_{um} indicate that the upload policy preserves evidence for suspicious cases while reducing unnecessary payload.

IV. EXPERIMENTAL SETUP

A. Dataset and Binary Screening Task

The proof-of-concept evaluation is conducted on a PlantVillage split reconstructed from the raw/color images. A class-wise random 80/20 split with seed 42 produces 43,429 training images and 10,876 validation images across 38 crop-condition classes. According to the class-name aggregation rule, the validation set contains 3022 healthy images and 7854 diseased images.

The original PlantVillage labels distinguish both plant species and health conditions. Since the objective of this work is to reduce the uplink traffic in health screening rather than species recognition, the original multiclass labels are aggregated into binary healthy and diseased categories. Let $P(i)$ denote the predicted softmax probability of class i . The healthy class set $\mathcal{C}_h$ contains class indices whose names include “healthy”, and the diseased class set is defined as $\mathcal{C}_d = \mathcal{C} \setminus \mathcal{C}_h$.

TABLE II
EXPERIMENTAL PLATFORM AND SOFTWARE ENVIRONMENT.

Item	Specification
CPU and memory	Intel Core i9-14900K, 128 GB RAM
GPU	NVIDIA GeForce RTX 5070 Ti, 16 GiB memory
Framework	Python 3.9.25, PyTorch 2.8.0, TorchVision 0.23.0
Libraries	Ultralytics 8.3.225, OpenCV 4.12.0, NumPy 2.0.2

Let P_h and P_d denote the aggregated healthy and diseased scores, respectively. They are computed as:

$$\begin{aligned} P_h &= \sum_{i \in \mathcal{C}_h} P(i), \\ P_d &= \sum_{i \in \mathcal{C}_d} P(i). \end{aligned} \quad (12)$$

The image is classified as diseased when $P_d \geq 0.5$, and as healthy otherwise. For diseased candidates that require Grad-CAM, the target class is selected using the diseased class set defined above, as described in Eq. (6).

B. Implementation and Evaluation Metrics

The experimental platform and software environment are summarized in Table II. Latency is measured on a per-image basis with batch size 1 and averaged over all 10,876 validation images. The first three images are used for warm-up and are excluded from the reported latency averages. CUDA timing uses explicit device synchronization, and both CPU and CUDA inference are conducted using FP32 precision.

The adaptive pipeline is timed according to the selected upload mode. Samples assigned to M_{msg} require only pre-processing, classifier inference, binary score aggregation, and metadata serialization. Samples assigned to M_{full} bypass Grad-CAM and ROI extraction, while samples assigned to M_{roi} execute Grad-CAM generation, CAM validation, and ROI-crop extraction. Therefore, the reported selected-mode latency reflects the actual computation required by the adaptive upload decision.

The classifier input follows the Ultralytics classification transform attached to the checkpoint, including resize, center crop, tensor conversion, and normalization. For payload comparison, uploaded images are resized according to the 224-pixel classifier input setting and encoded in JPEG format with a quality factor of 85. Grad-CAM heatmaps are used only for internal analysis and ROI proposal generation, and are not counted as uploaded payloads.

The evaluation considers the full-image upload baseline and the three upload modes in Table I. ROI crops are extracted from connected components above the CAM threshold $\lambda_H = 0.60$. The confidence thresholds are set to $\tau_h = \tau_d = 0.90$, and the hotspot-ratio threshold is set to $\tau_\gamma = 0.25$. We report binary screening metrics, selected-mode latency, average payload size, payload saving ratio, evidence preservation metrics, and estimated uplink delay under 0.5, 1, 5, and 10 Mbps uplink rates.

For M_{msg} , the transmitted payload is represented as a compact JSON object containing only the sample identifier,

TABLE III
BINARY CLASSIFICATION METRICS ON THE VALIDATION SPLIT.

Metric	Value	Note
Samples	10876	3022 healthy / 7854 diseased
Accuracy	99.51%	Overall binary correctness
Precision	99.85%	Diseased as positive class
Recall	99.48%	Diseased as positive class
F1-score	99.66%	Diseased as positive class
AUROC	0.9999	Binary aggregation score
TN/FP/FN/TP	3010/12/41/7813	Confusion-matrix counts

TABLE IV
EDGE-SIDE LATENCY BREAKDOWN WITH BATCH SIZE 1.

Device	Pre. (ms)	Inf. (ms)	Meta. (ms)	CAM (ms)	ROI (ms)	Full (ms)	Total (ms)
CPU	4.06	72.08	0.09	160.75	0.23	0.01	237.22
CUDA	2.72	7.80	0.07	33.61	0.22	0.01	44.42

selected upload mode, binary screening result, binary confidence value, and decision threshold. Ground-truth labels, image content, and diagnostic analysis fields are not included. An example M_{msg} payload is shown below:

```
{
  "id": "PV_val_000123",
  "m": "M_msg",
  "y": "healthy",
  "conf": 0.9821,
  "thr": 0.90
}
```

V. RESULTS AND DISCUSSION

A. Binary Screening and Edge Latency

Table III summarizes the binary screening results. Under the controlled PlantVillage validation split and the probability aggregation rule, the evaluated checkpoint achieves high binary screening performance between healthy and diseased samples. This result should be interpreted as proof-of-concept verification on a controlled public dataset rather than as evidence of field-level diagnostic robustness. Among the 7854 diseased validation images, 7851 are assigned to visual-evidence upload modes, resulting in an evidence recall of 99.96% and an unsafe metadata-only rate of 0.04%.

Table IV reports the edge-side latency breakdown. The CUDA implementation reduces the selected-mode per-image processing time from 237.22 ms on CPU to 44.42 ms on GPU. Grad-CAM generation accounts for the largest portion of the processing time among samples that require visual evidence, while metadata-only M_{msg} samples bypass Grad-CAM and image compression.

The CAM, ROI, and full-image compression columns are averaged over all validation samples, with zero time assigned to samples that bypass the corresponding operation.

B. Payload and Uplink Delay Reduction

Table V summarizes the selection counts of the upload modes, average payloads, saving ratios, and average uploaded

TABLE V
PAYLOAD SIZE AND SAVING RATIO UNDER THE ADAPTIVE UPLOAD
POLICY.

Upload option	Count	Payload (KB)	Saving (%)	ROIs
Full baseline	All	13.64	0.00	0.00
M_{msg}	2947	0.11	99.20	0.00
M_{roi}	7551	3.35	73.22	2.09
M_{full}	378	13.18	0.00	0.00
Adaptive	All	2.82	77.71	1.45

TABLE VI
ESTIMATED UPLINK DELAY UNDER DIFFERENT WIRELESS RATES.

Upload option	0.5 Mbps (ms)	1 Mbps (ms)	5 Mbps (ms)	10 Mbps (ms)
Full baseline	223.46	111.73	22.35	11.17
M_{msg}	1.82	0.91	0.18	0.09
M_{roi}	54.91	27.46	5.49	2.75
M_{full}	215.92	107.96	21.59	10.80
Adaptive	46.12	23.06	4.61	2.31

ROI crop counts. The full-image baseline includes the compact metadata payload plus the full resized image and has an average payload of 13.64 KB. The adaptive policy reduces the average payload to 2.82 KB, corresponding to a per-image average saving of 77.71%. The metadata-only mode M_{msg} provides the largest saving for confident healthy cases, M_{roi} uploads localized ROI crops for high-confidence diseased cases with concentrated CAM responses, and M_{full} is retained as the conservative full-image fallback for uncertain, invalid-CAM, or diffuse-CAM cases. The saving ratio is computed for each image relative to the payload of uploading the same image in full-image form. Therefore, M_{full} has zero saving by definition, even though the average payload of the M_{full} subset may differ from the global full-image average.

Table VI converts the average payload size into estimated uplink transmission delay under different wireless rates. Since the transmission delay is proportional to the payload size, the adaptive policy reduces the estimated delay across all of the evaluated uplink settings. Under a constrained 0.5 Mbps uplink, the estimated average transmission delay is reduced from 223.46 ms with full-image upload to 46.12 ms with the adaptive policy. Fig. 3 further illustrates the payload and delay comparison between the full-image baseline and the adaptive upload policy.

The results show that the communication benefit comes from adapting the uploaded content to the screening confidence and Grad-CAM response of each image. The current evaluation is based on a controlled public dataset and should be regarded as a proof-of-concept study. In real-world plant monitoring scenarios, background clutter, illumination variation, occlusion, and domain shift may affect classification and Grad-CAM reliability. Future work will validate the framework using real-scene ornamental plant or bonsai images and investigate adaptive thresholds for different network and inspection requirements.

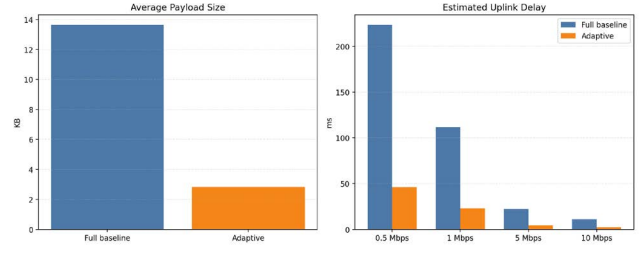


Fig. 3. Average payload size and estimated uplink delay comparison between the full-image baseline and the adaptive upload policy.

VI. CONCLUSION

This paper presented a Grad-CAM-assisted adaptive upload framework for reducing uplink traffic in plant health screening. The proposed framework uses edge-side classification confidence and Grad-CAM response to select metadata-only reporting, ROI-crop upload, or full-image fallback. A proof-of-concept evaluation on PlantVillage images showed that the adaptive policy reduced the average upload payload from 13.64 KB to 2.82 KB, corresponding to a per-image average saving of 77.71%. These results confirm that explainable edge inference can serve as an effective upload-control mechanism for reducing visual data transmission while preserving compact evidence for remote inspection.

REFERENCES

- [1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Front. Plant Sci.*, vol. 7, 2016, art. no. 1419.
- [2] V. Udutalapally, S. P. Mohanty, V. Pallagani, and V. Khandelwal, "sCrop: A novel device for sustainable automatic disease prediction, crop selection, and irrigation in Internet-of-Agro-Things for smart agriculture," *IEEE Sensors J.*, vol. 21, no. 16, pp. 17 525–17 538, 2021.
- [3] P. E. Correa da Silva and J. Almeida, "An edge computing-based solution for real-time leaf disease classification using thermal imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 22, pp. 1–5, 2025.
- [4] A. Khan, U. Nawaz, A. Ulhaq, and R. W. Robinson, "Real-time plant health assessment via implementing cloud-based scalable transfer learning on AWS DeepLens," *PLOS ONE*, vol. 15, no. 12, 2020, art. no. e0243243.
- [5] H. Zhu, X. Huang, H. Gao, M. Jiang, H. Que, and L. Mu, "A wireless collaborated inference acceleration framework for plant disease recognition," in *Advanced Intelligent Computing Technology and Applications*, ser. Lecture Notes in Computer Science, D.-S. Huang, C. Zhang, Q. Zhang, and Y. Pan, Eds. Singapore: Springer, 2025, vol. 15862, pp. 331–341.
- [6] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.