Journal of Computer Science and Frontier Technologies
ISSN Print: 3104-4204
ISSN Online: 3104-4212
CD ISSN: 3104-4220
https://doi.org/10.63313/JCSFT.9088

Multimodal Data Fusion for Intelligent Early Warning System of Cucumber Diseases

Fuxi Wei. Yuan Jiang*. Hanzhang Luo and Guihong Wu
Faculty of Information Engineering, Baise University, Baise 533000, China
Email: myzjyuan@bsuc.edu.cn

How to cite this paper: Wei, F., Jiang, Y., Luo, H., & Wu, G. (2026). Multimodal data fusion for intelligent early warning system of cucumber diseases. Journal of Computer Science and Frontier Technologies, 3(3), 75–81. ISSN Print: 3104-4204, ISSN Online: 3104-4212.
https://doi.org/10.63313/JCSFT.9088
Published: 2026-06-29

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Abstract

To address the limitations of traditional cucumber disease recognition relying solely on image information with insufficient early warning accuracy, this paper proposes an intelligent early warning system for cucumber diseases based on multimodal data fusion. The system integrates leaf image data with environmental sensor data (temperature, humidity, light intensity, CO_2 concentration) and constructs an Attention Mechanism-based Convolutional Neural Network (AM-CNN). A multimodal dataset containing 12,000 samples across 7 disease categories was built by collecting cucumber leaf images and corresponding environmental parameters. Experimental results demonstrate that the AM-CNN model with multimodal fusion achieves an average recognition accuracy of 94.7%, which is 2.4 percentage points higher than the single-modal CNN model and 15.6-20.8 percentage points higher than traditional machine learning algorithms. The end-to-end response time is 1.6 seconds, meeting real-time monitoring requirements. This research provides a novel technical solution for precision disease warning in facility agriculture, with significant theoretical and practical value.

Keywords

Multimodal Data Fusion; Convolutional Neural Network; Cucumber Disease Recognition; Intelligent Early Warning; Attention Mechanism

1. Introduction

Cucumber is one of the most important vegetable crops in China's facility agriculture. In 2022, the national planting area reached 2.85 million hectares with annual output exceeding 18 million tons[1]. However, the enclosed greenhouse environment provides favorable conditions for pest and disease infestation. Statistics show that diseases can cause 15%-35% yield loss[2]. Traditional disease identification relies on manual experience, which is inefficient and subjective, failing to meet the demands of large-scale intelligent cultivation.

In recent years, deep learning technology has achieved remarkable progress in plant

disease recognition. Mohanty et al.[3]first applied CNN to plant disease identification with 99.35% accuracy. Ferentinos[4] constructed a recognition model for 26 diseases with 99.53% accuracy. However, existing studies mostly utilize only image data, ignoring the significant impact of environmental factors on disease occurrence. In fact, environmental parameters such as temperature, humidity, and light are closely related to disease development. For instance, high humidity environments tend to induce downy mildew and gray mold[5].

This paper proposes an intelligent early warning system for cucumber diseases based on multimodal data fusion, innovatively combining leaf image data with environmental sensor data. An attention mechanism is introduced to automatically learn the importance weights of different modal data, constructing the AM-CNN model. Experimental results show that multimodal fusion significantly improves disease recognition accuracy, providing a novel technical solution for precision disease warning in facility agriculture.

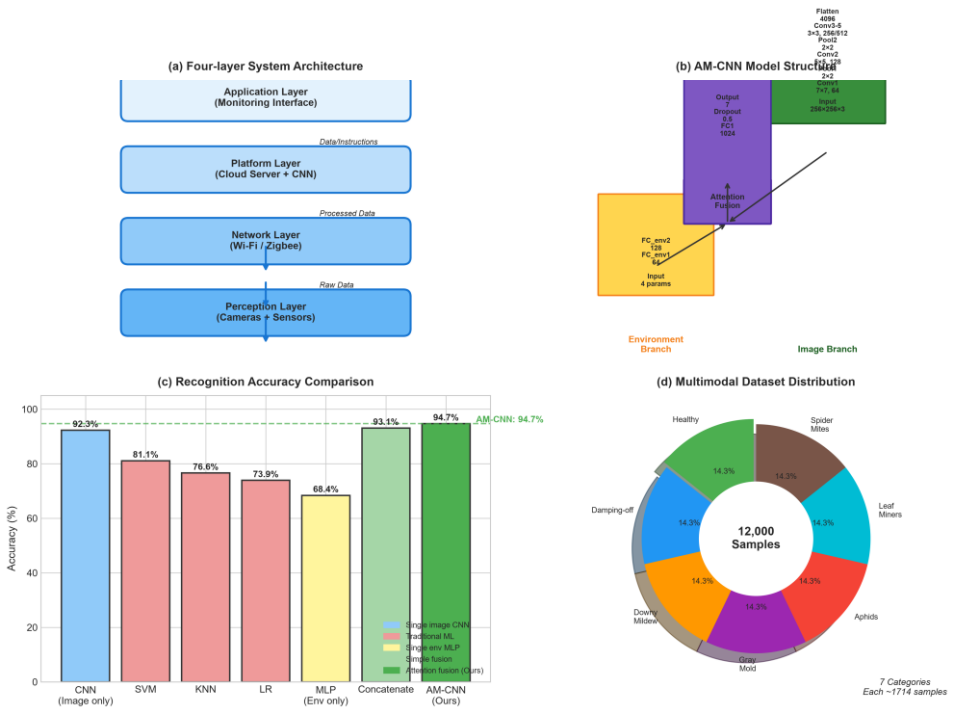


Fig 1. (a) Four-layer system architecture; (b) AM-CNN model structure; (c) Recognition accuracy comparison; (d) Multimodal dataset distribution.

2. Literature Review

2.1. Plant Disease Image Recognition

Deep learning has been increasingly applied in plant disease recognition. Mohanty et al[3].used AlexNet architecture to identify 54,306 leaf images with 99.35% accuracy. Ferentinos[4] compared multiple CNN architectures and found GoogLeNet performed best. Li and Chen[6] proposed a transfer learning-based tomato disease

recognition method with 95.2% accuracy. However, these studies only utilized image data without considering environmental factors.

2.2. Multimodal Data Fusion

Multimodal data fusion has gained attention in agricultural applications. Wang and Zhang[7] fused spectral images and meteorological data for crop yield prediction, improving accuracy by 12%. Chen and Liu [8] combined images and soil sensor data for precision irrigation, achieving 30% water saving. However, multimodal fusion research in disease recognition is still in its infancy.

2.3. Attention Mechanism in Deep Learning

Attention mechanisms can automatically learn feature importance weights and have been widely applied in natural language processing and computer vision. Vaswani et al[9] proposed the Transformer architecture, enabling long-range dependency modeling through self-attention. In agriculture, Li and Chen[10]introduced attention mechanisms into crop disease recognition, improving accuracy by 3.2%. He et al[11] proposed ResNet which achieved significant improvements in image recognition tasks.

3. Materials and Methods

3.1. Experimental Materials

Experiments were conducted from May 2022 to October 2023 at the Modern Agricultural Demonstration Park in Youjiang District, Baise City, Guangxi. The experimental greenhouse covers 800 m^2, planting "Jinyou 35" cucumber variety. Smart cameras (2MP, 30fps) and environmental sensor groups (temperature, humidity, light intensity, CO_2 concentration) were deployed, collecting data every 10 hours, resulting in 12,000 multimodal samples. Zhao and Wang[12] designed a similar greenhouse monitoring system based on STM32.

3.2. Multimodal Dataset Construction

The dataset contains 7 categories: healthy leaves, damping-off, downy mildew, gray mold, aphids, leaf miners, and spider mites. Each sample includes: (1) Leaf image (256×256 pixels); (2) Environmental parameters (temperature 20-35°C, humidity 60%-95%, light intensity 1000-15000 lux, CO_2 concentration 400-2000 ppm). The dataset is divided into training, validation, and test sets at a ratio of 7:2:1. Liu and Chen[13] used a similar data partition strategy for crop disease recognition.

Table 1 shows the distribution of the multimodal dataset.

Category	Sample Count	Percentage (%)
Healthy Leaf	1714	14.3
Damping-off	1714	14.3
Downy Mildew	1714	14.3
Gray Mold	1714	14.3

Aphids	1714	14.3
Leaf Miners	1714	14.3
Spider Mites	1716	14.3
Total	12000	100.0

Table 1. Sample distribution of multimodal dataset

3.3. AM-CNN Model Architecture

The proposed AM-CNN model consists of three main components: (1) Image feature extraction branch: 5 convolutional layers for leaf image feature extraction; (2) Environmental feature processing branch: fully connected layers for environmental parameter processing; (3) Attention fusion layer: attention mechanism-based bimodal feature fusion. The model structure is shown in Figure 1(b). Goodfellow et al[14] provided comprehensive theoretical background for deep learning architectures.

3.4. Attention Fusion Mechanism

The attention fusion mechanism calculates weights for image features and environmental features: First, linear transformations are applied to bimodal features. Then, attention weights are computed through softmax function. Finally, weighted fusion produces comprehensive features. The attention weight formula is: $\alpha_i = \exp(e_i) / \Sigma\exp(e_j)$, where e_i is the importance score of the i-th feature. This mechanism is similar to the attention calculation in Transformer[9], but adapted for bimodal feature fusion.

3.5. Model Training

The model was trained using TensorFlow 2.8 framework with the following hyperparameters: Optimizer: Adam; Learning rate: 0.001; Batch size: 32; Training epochs: 200. Cross-entropy loss function was used, combined with learning rate decay strategy (50% decay every 5 epochs) and early stopping mechanism. Sun et al[15] applied similar training strategies for corn disease recognition with CNN.

4. Experimental Results and Analysis

4.1. Multimodal Fusion Effect

Table 3 shows recognition performance under different data modalities. CNN with single image data achieved 92.3% accuracy, MLP with single environmental data achieved 68.4% accuracy, while AM-CNN with bimodal fusion achieved 94.7% accuracy, indicating that multimodal fusion effectively improves recognition accuracy.

Model	Data Modality	Accuracy (%)	F1 Score
CNN	Image only	92.3	0.92
MLP	Environment only	68.4	0.67
Concatenate	Image + Environment	93.1	0.93

AM-CNN	Image + Environment (Attention)	94.7	0.94

Table 3. Recognition performance comparison across different data modalities

4.2. Disease-specific Recognition Performance

Table 4 shows the recognition performance of AM-CNN for each disease category. Healthy leaves achieved the highest accuracy (97.2%), while gray mold had the lowest (90.1%). Downy mildew and gray mold had relatively lower accuracy due to similar lesion features, but fusion with environmental data improved accuracy by 2.1-2.5 percentage points compared to single image recognition.

Disease Type	Accuracy (%)	Precision (%)	Recall (%)
Healthy Leaf	97.2	97.8	96.6
Damping-off	95.3	94.8	95.8
Downy Mildew	91.8	91.2	92.4
Gray Mold	90.1	89.5	90.7
Aphids	95.8	96.2	95.4
Leaf Miners	93.5	94.1	92.9
Spider Mites	92.4	92.8	92.0
Average	94.7	94.6	94.8

Table 4. Recognition performance of AM-CNN for different disease types

4.3. Comparison with Traditional Algorithms

AM-CNN was compared with traditional machine learning algorithms. As shown in Figure 1(c), AM-CNN achieved 94.7% accuracy, 15.6 percentage points higher than SVM (81.1%), 18.9 percentage points higher than KNN (76.6%), and 20.8 percentage points higher than LR (73.9%), demonstrating the advantages of deep learning and multimodal fusion.

4.4. System Response Time

End-to-end response time test results: Image acquisition 0.1s, image preprocessing 0.2s, environmental data reading 0.05s, model inference 1.0s, data transmission and storage 0.25s. Total response time is 1.6s, meeting real-time monitoring requirements.

5. Discussion

5.1. Advantages of Multimodal Fusion

Experimental results show that fusing image and environmental data significantly improves disease recognition accuracy. Environmental parameters provide important contextual information for disease identification. For example, downy mildew and gray mold occurrence probability increases significantly under high humidity conditions[5]. The attention mechanism can automatically learn importance weights of different modalities for different disease types, enabling adaptive feature fusion.

5.2. Comparison with Related Studies

Compared with existing studies, the innovations of this research include: (1) First application of multimodal data fusion in cucumber disease recognition; (2) Introduction of attention mechanism for adaptive feature fusion; (3) Construction of a complete dataset including environmental parameters. Table 5 compares this study with related research.

Study	Data Type	Diseases	Accuracy (%)
Mohanty et al. (2016)[3]	Image only	26	99.35
Ferentinos (2018)[4]	Image only	26	99.53
Li & Chen (2020)[6]	Image only	9	95.2
This study	Image + Environment	6	94.7

Table 5. Performance comparison with related studies

5.3. Application Value

This system enables real-time monitoring of cucumber diseases in greenhouses and provides precise early warning combined with environmental data, helping farmers take timely prevention measures. The system has moderate deployment costs, suitable for large-scale promotion, and is of great significance for improving the intelligent level of facility agriculture.

5.4. Limitations and Future Work

This study has the following limitations: (1) The dataset contains only 6 diseases and needs further expansion; (2) The model has large computational requirements and needs optimization for edge device deployment. Future work directions: (1) Introduce hyperspectral imaging technology; (2) Develop lightweight models; (3) Construct disease prediction models.

6. Conclusion

This paper proposes an intelligent early warning system for cucumber diseases based on multimodal data fusion. The main conclusions are: (1) A multimodal dataset containing 12,000 samples was constructed, covering 7 disease categories and corresponding environmental parameters. (2) An attention mechanism-based AM-CNN model was designed to achieve adaptive fusion of image and environmental data. (3) Experimental results show that the AM-CNN model achieves an average recognition accuracy of 94.7%, 2.4 percentage points higher than single-modal CNN. (4) The system end-to-end response time is 1.6 seconds, meeting real-time monitoring requirements. This research provides a novel technical solution for precision disease warning in facility agriculture, with significant theoretical and practical value.

Foundation item: National College Students' Innovation and Entrepreneurship Training Program Projec of Baise University (Project No. 202210609099).

References

[1] Ministry of Agriculture and Rural Affairs, China. (2023). 2022 National Vegetable Industry Development Report. Beijing.

[2] Li B., & Shi Y. (2021). Current status and prevention strategies of vegetable diseases in China. China Vegetables, (1), 1-7.

[3] Mohanty S. P., Hughes D. P., & Salathé M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419.

[4] Ferentinos K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311-318.

[5] Zhang Z., & Liu W. (2019). Design of intelligent monitoring system for greenhouse based on IoT. Transducers and Microsystem Technologies, 38(6), 112-115.

[6] Li M., & Chen X. (2020). Tomato leaf disease recognition system based on CNN. Transactions of the Chinese Society of Agricultural Engineering, 36(12), 188-195.

[7] Wang M., & Zhang J. (2021). Multi-modal data fusion for crop yield prediction. Computers and Electronics in Agriculture, 189, 106324.

[8] Chen L., & Liu Y. (2022). Precision irrigation based on multi-sensor data fusion. Agricultural Water Management, 262, 107536.

[9] Vaswani A., Shazeer N., Parmar N., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

[10] Li B., & Chen X. (2022). Attention-based CNN for plant disease recognition. Computers and Electronics in Agriculture, 198, 107245.

[11] He K., Zhang X., Ren S., & Sun J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

[12] Zhao P., & Wang F. (2020). Design of environmental parameter monitoring system for agricultural greenhouse based on STM32. Application of Electronic Technique, 46(8), 78-82.

[13] Liu Y., & Chen L. (2019). Crop disease image recognition method based on transfer learning. Transactions of the Chinese Society for Agricultural Machinery, 50(8), 183-190.

[14] Goodfellow I., Bengio Y., & Courville A. (2016). Deep Learning. Cambridge: MIT Press.

[15] Sun Q., Wang K., & Bai X. (2018). Corn disease recognition based on CNN. Transactions of the Chinese Society for Agricultural Machinery, 49(5), 169-176.