GarlicNet: A pressure signal-based CNN-Transformer hybrid network for detecting the breakage and separation degree of garlic cloves

Xingyu Xia[1], Jie Tian[2], Mingxi Shao[1*], Yanan Zhang[3*]

(1. *College of Mechanical Engineering, Qinghai University, Xining 810016, China*;
2. *Qinghai Key Laboratory of Vegetable Genetics and Physiology, Academy of Agriculture and Forestry Sciences of Qinghai University, Xining 810016, China*;
3. *School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China*)

Abstract: Current research on garlic clove breaking primarily focuses on equipment development and optimization, with limited attention to elucidating the relationship between the breaking process and resultant damage or clove separation. To address this gap, this paper presents GarlicNet, a deep learning model that utilizes pressure signals generated during clove breaking to accurately detect breakage severity and separation extent. The model employs an enhanced adaptive channel attention mechanism to capture critical multi-channel information and a dual-branch attention structure to amplify salient features within pressure signals. Experimental results demonstrate high performance: GarlicNet achieved 95.6% accuracy, 95.9% recall, 95.9% precision, and 95.8% F1-score for breakage detection; corresponding values for separation detection were 96.5%, 96.5%, 96.5%, and 96.4%. Ablation studies confirm the efficacy of the Dynamic Multi-Channel Convolution Fusion (DMCF) module and Dual-Branch Attention Fusion (DAF) structure. Compared to benchmark models, GarlicNet exhibits superior detection performance and robustness. These findings validate the feasibility of predicting breakage and separation via pressure signal during clove breaking and underscore the model's practical utility. This approach shows significant potential for mechanized garlic processing by reducing losses, improving efficiency, and advancing industrial automation.

Keywords: garlic clove breaking, GarlicNet, pressure signal, damage detection, clove separation, deep learning
DOI: 10.25165/j.ijabe.20261903.10121

Citation: Xia X Y, Tian J, Shao M X, Zhang Y N. GarlicNet: A pressure signal-based CNN-Transformer hybrid network for detecting the breakage and separation degree of garlic cloves. Int J Agric & Biol Eng, 2026; 19(3): 212–224.

1 Introduction

Garlic (*Allium sativum*), a widely used food and dietary supplement, is well known for its rich phytochemical composition[1,2]. According to FAOSTAT, the global garlic planting area reached approximately 1.6 million hm^2 in 2023[3]. China, as one of the world's major garlic-producing regions, accounts for more than 70% of global production and over 80% of global exports[4]. Therefore, the development of mechanization in garlic production is particularly important. As a key step in garlic cultivation, mechanized clove separation can significantly improve operational efficiency; however, it can also easily cause mechanical damage to garlic seed cloves and incomplete separation. Existing studies have shown that mechanical damage can have significant negative effects on seed vigor and subsequent growth. For soybean, rice, and maize, damage incurred during threshing or transportation generally reduces germination rate and increases susceptibility to pests and diseases[5-7]. Similarly, damaged garlic seed cloves are more prone to pathogen infection at the injured sites, which may lead to seed rot, uneven emergence, and ultimately yield loss[4]. In addition, incompletely separated garlic cloves are likely to cause uneven sowing and poor development. Therefore, conducting online and quantitative detection of damage and separation status during the garlic clove separation process is of great significance for improving the operational performance of clove separation equipment and ensuring sowing quality.

Current research on garlic production mechanization mainly focuses on planting, harvesting, and post-harvest processing, whereas the development of garlic clove separation machinery is still at the stage of technological exploration. According to the operating principles of clove separation, existing equipment mainly adopts two technical approaches: the rubbing-tearing method and the rolling-rubbing method, corresponding to rubber-roller clove separation devices and cone-disc clove separation devices, respectively[8]. Typical machines include the SFB-400 double rubber-roller clove separator produced by Dayang in Zhucheng, Shandong; the JH-A bionic cone-disc clove separator produced by Jiahe in Shandong; and the JJ.BROCH garlic clove separator from Spain[9,10]. Although the performance of such equipment has continued to improve, systematic research on the clove condition during and after the separation process remains lacking[11]. At present, the status inspection of garlic seed cloves after clove separation mainly relies on manual identification or offline recognition[12,13]. Manual identification is highly subjective, as well as time-consuming and labor-intensive. Moreover, since damage and incomplete separation are caused during the clove separation process, offline recognition is inherently delayed and therefore unable to provide timely feedback for adjusting machine parameters to reduce damage during

Received date: 2025-08-21 **Accepted date:** 2026-04-20
Biographies: Xingyu Xia, ME, research interest: intelligent agricultural machinery equipment, Email: 13033004925@163.com; **Jie Tian**, PhD, Associate Professor, research interest: physiology of vegetable cultivation, Email: 345409210@qq.com.
***Corresponding author: Mingxi Shao**, PhD, Associate Professor, research interest: intelligent agricultural machinery equipment. College of Mechanical Engineering, Qinghai University, 251 Ningda Road, Xining 810016, China. Tel: +86-18110072802, Email: nifengfeiyang@163.com; **Yanan Zhang**, PhD, Research Assistant, research interest: electric vehicle chassis control. School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China. Tel: +86-15063511839, Email: 15063511839@163.com.

operation. In recent years, deep learning has offered a feasible and efficient solution for a wide range of agricultural classification[14] and detection[15] tasks. Applications of deep learning in garlic-related research remain relatively limited and are mainly focused on sowing and the detection of garlic bulb volume and mass. Zhang et al.[16] developed a real-time garlic clove bud detection system for a clove orientation metering device. Son et al.[17] combined a depth camera with machine learning to develop a non-contact system for predicting the volume and mass of garlic bulbs. In other agricultural products, deep learning has been more widely applied to object detection tasks. For example, Li et al.[18] improved the YOLOv8 network and integrated multi-scale features to achieve accurate identification of tea shoots. Chen et al.[19] designed a novel attention module for crop disease image recognition. Chen et al.[20] also proposed a new lightweight detection model for identifying tomato diseases. At present, there is still a lack of research on and development of real-time detection algorithms for the garlic clove separation process. Therefore, it is of clear research significance to carry out online detection specifically targeting the separation process itself and to establish process-signal characterization methods capable of reflecting the mechanisms underlying damage formation and separation behavior.

Mechanical damage caused by garlic clove separation exhibits diverse manifestations. Some cloves show only slight surface cracks, while others may suffer internal damage despite remaining wrapped in intact outer skin; in addition, mutual occlusion frequently occurs. As a result, traditional machine vision is insufficient to fully capture subtle surface damage as well as damage and separation conditions concealed by occlusion[13]. Compared with visual information, pressure signals are derived directly from the contact, compression, and impact behaviors occurring during the clove separation process, and can therefore more directly reflect the mechanical behavior of the operation itself as well as the formation mechanisms of clove damage and separation states. Meanwhile, the acquisition of pressure signals is not affected by factors such as illumination, occlusion, or imaging angle[21], making them more suitable for continuous online collection. Therefore, online detection of clove separation quality based on pressure signals offers stronger real-time capability and closer process relevance. The signals generated during garlic clove separation typically exhibit temporal characteristics in which local abrupt changes coexist with global multi-sensor coordination. On the one hand, damage and separation events lead to pronounced local variations, such as peaks, troughs, and transient impacts. On the other hand, different clove separation states are reflected over the full operating cycle in the overall differences among response patterns from multiple sensors and in their cross-temporal dependency relationships. Based on this, this study constructs a CNN–Transformer hybrid architecture, in which CNN is used to extract local features from individual sensor arrays, while Transformer is employed to model the global representations and dependency relationships among multiple sensors[22], thereby improving the recognition of garlic damage and separation states.

The major contributions of this work are summarized as follows: (1) Collects dynamic pressure signals directly from conical-disc separation equipment; (2) Transforms signals into spatiotemporal pressure maps for feature extraction; (3) Develops GarlicNet—a hybrid CNN-Transformer architecture—to simultaneously predict breakage severity and separation completeness.

2 Materials and methods

2.1 Sample preparation and measurement of physical properties

The experimental samples comprised two garlic cultivars sourced from the local market (Figure 1): dry-skin garlic from Kaifeng (Henan Province) and fresh-harvest garlic from Kunming (Yunnan Province). The Kaifeng specimens were naturally air-dried with desiccated outer skins, whereas Kunming samples retained field-moist pericarps.

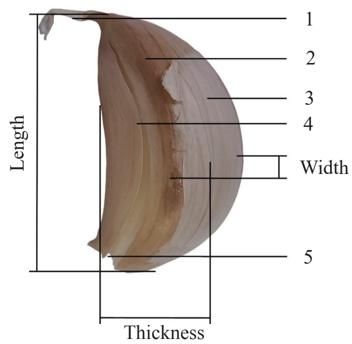
a. Purple-skinned garlic from Kaifeng, Henan

b. Purple-skinned garlic from Kunming, Yunnan

Figure 1 Garlic samples

For each cultivar, 200 intact cloves were manually separated. Randomly selected subsets underwent characterization:

(1) Physical dimensions (n=100/clove): Length (base of scale leaf to root plate), width (maximum dorsal radial span at peak curvature), thickness (dorsal-ridge distance at same curvature point) were measured using digital calipers (Figure 2)[23]. Individual mass was recorded via analytical balance (0.001 g resolution).



1. Scale leaf; 2. Longitudinal ridge; 3. Dorsal side; 4. Lateral side; 5. Root plate

Figure 2 Definition of garlic shape and size

(2) Moisture content (n=100/clove): Groups of 20 cloves were oven-dried at 90°C (DHG-9013A, Shanghai, China) for 360 min until constant mass, with outer skins retained. Moisture content (MC) was calculated per Equation (1).

$$MC = \frac{m_2 - m_3}{m_2 - m_1} \tag{1}$$

where, MC is the moisture content of garlic, m_1 is the weight of the beaker, m_2 is the total weight of garlic and beaker before drying, and m_3 is the total mass of the dried garlic and the beaker after drying.

2.2 Pressure signal acquisition and image transformation

The process of collecting the pressure experienced during garlic clove breaking is shown in Figure 3. Garlic clove-breaking experiments were conducted using a conical-disc clove-breaking device; the forces acting on the assembly are shown in Figure 4. Pressure data were collected using RX-M0808MS distributed thin-film flexible pressure sensors (Rouxi Electronic Technology Co., Ltd., Changzhou, China). Each sensor consisted of an 8×8 array with 64 sensing units. Four sensors were mounted at fixed intervals in the north, south, east, and west positions of the device to monitor the pressure distribution during the clove-breaking process. Signals were recorded at 200 ms intervals. Before data collection, each sensing unit was calibrated using 10%, 50%, and 90% full-scale standard loads to ensure measurement consistency. A layer of high-density foam slightly larger than the sensor surface was attached to each sensor to stabilize the installation and reduce impact during the breaking process.

Each garlic clove-breaking trial follows the same data acquisition procedure. The data acquisition board is first connected to the computer via a USB cable, and the data acquisition and visualization software is launched to start recording. The clove-splitting device is then loaded with groups of eight whole garlic cloves, while four pressure sensors simultaneously and continuously record pressure signals during the garlic clove-breaking process. After all cloves in each group are split, recording is stopped, and the time-series pressure signals from each sensor are saved. For each sensor, the pressure signals of all sensing units are averaged along the time dimension, compressing the original N×64 data into a 1×64 feature vector that characterizes the overall response of the sensor. Thus, each trial yields four 1×64 dimensional response vectors, which are subsequently normalized and spatially mapped according to the physical layout of the sensing units to generate heatmaps composed of 64 rectangular elements. These heatmaps provide an intuitive representation of the pressure distribution during the garlic clove-breaking process and establish the basis for subsequent model development.

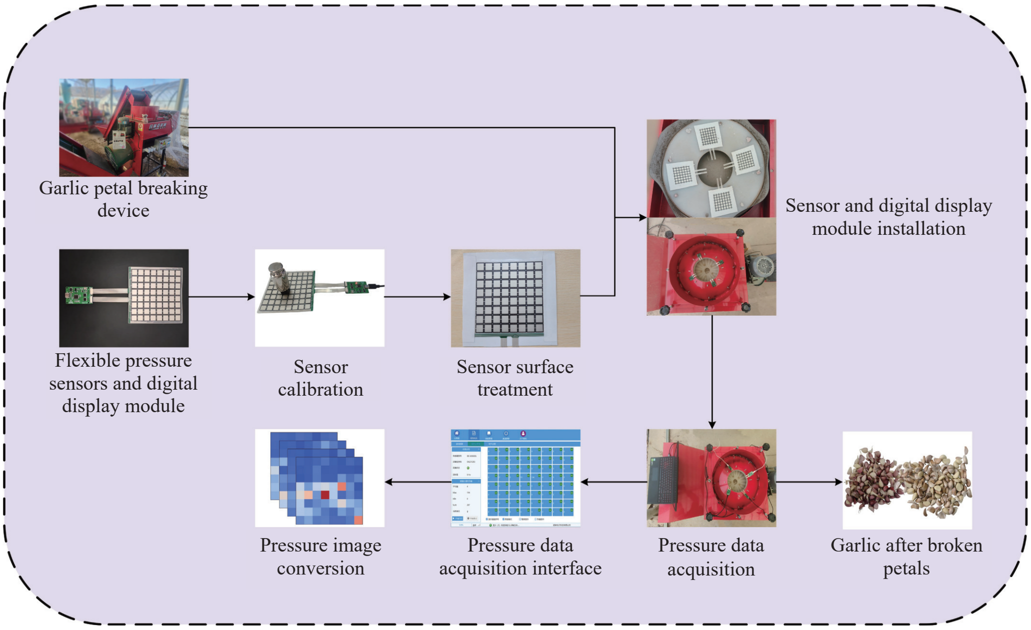
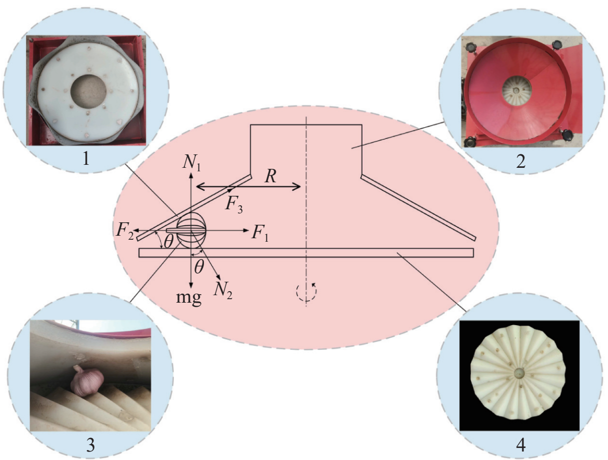


Figure 3 Pressure data acquisition process during garlic clove breaking



1. Conical disc 2. Feeding inlet 3. Garlic condition 4. Rotating disc

Figure 4 Force diagram during the garlic clove breaking process

2.3 Breakage rate and separation rate calculation

After each clove-breaking trial, the resulting garlic cloves were classified into three categories: intact, broken, and unseparated, as illustrated in Figure 5. For each group, the total weight of the separated garlic cloves was recorded as M_t. Garlic cloves exhibiting breakage, including visibly broken or fractured cloves, as shown in Figure 5b, were manually identified, and their total weight was recorded as M_b. Clove samples with two or more cloves remaining linked were considered unseparated, as shown in Figure 5c, with their total weight denoted as M_u. The calculation formulas for breakage rate (Br) and separation rate (Sr) are given in Equation (2) and Equation (3), respectively. Based on existing studies[24], the severity of breakage and degree of separation are defined as follows: a breakage rate between 0–3% is classified as slight breakage, 3%-6% as moderate breakage, and above 6% as severe breakage. For separation, a rate exceeding 95% is considered high separation, 90%-95% as moderate separation, and below 90% as low separation.

$$\mathrm{Br} = \frac{M_b}{M_t} \tag{2}$$

$$\mathrm{Sr} = \frac{M_t - M_u}{M_t} \tag{3}$$

2.4 Dataset structure

The pressure data collected during the garlic clove-breaking process, originally stored in CSV format, were converted into images, as shown in Figure 6. The resulting images were saved in

PNG format with a resolution of 1200×1200 pixels. Each image is divided into 64 equal-sized square regions arranged in an 8×8 grid, corresponding to the layout of the sensor's individual sensing units. For each group of experiments, four pressure images were generated. A blue-to-red gradient was employed for visualization: following data normalization, deep blue indicates values approaching zero, while deep red corresponds to values near one. As shown in Figure 6, the force distribution during the clove-breaking process is concentrated in the lower portion of the sensor, where the sensing units exhibit more intense responses. This phenomenon can be attributed to the installation configuration of the sensor on the conical disc. The upper portion has a relatively large gap between the conical disc and the rotating plate, resulting in lower pressure or even no contact at that region; in contrast, the lower portion has a smaller gap, causing the garlic to make firm contact and undergo compression under centrifugal force. These pressure images not only provide an intuitive representation of the force distribution during clove breaking but also capture inter-sensor relationships and subtle variations, offering a reliable foundation for subsequent feature extraction.

In this study, various breakage and separation rates of seed garlic were obtained by adjusting the gap between the conical disc and the rotating plate. Among the 455 collected samples, breakage rates ranged from 1.2% to 10.9%, and separation rates ranged from 76.7% to 100%. A Pearson correlation analysis was performed between the breakage rate and separation rate. The analysis revealed a significant positive correlation between the two variables (r= 0.632, p<0.001, n=455), indicating that under the current operating conditions, an increase in separation rate may be accompanied by an increase in breakage rate.

a. Intact garlic cloves b. Broken garlic cloves

c. Unseparated garlic cloves

Figure 5 Classification of garlic after clove-breaking experiments

Group 1	Sensor unit 1	Sensor unit 2	Sensor unit 3	···	Sensor unit 64
Sensor 1	0	423	125	···	21
Sensor 2	126	21	283	···	114
Sensor 3	271	0	876	···	0
Sensor 4	9	1062	0	···	7
Group n	···	···	···	···	···

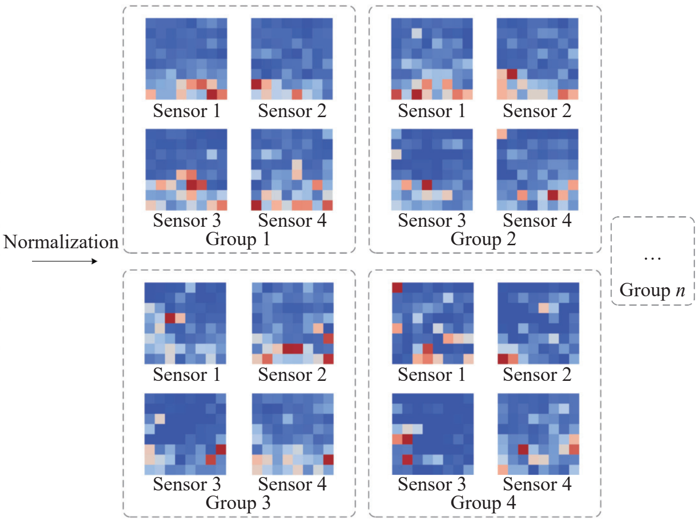


Figure 6 Pressure image transformation during the garlic clove-breaking process

According to the classification method described in Section 2.3, breakage rate and separation rate were labeled independently. For breakage rate, slight, moderate, and severe breakage were assigned labels 0, 1, and 2, respectively. For separation rate, high, moderate, and low separation were assigned labels 0, 1, and 2, respectively. As a result, the number of samples for breakage labels was: 135 for label 0, 165 for label 1, and 155 for label 2. For separation labels, 140 samples were assigned to label 0, 165 to label 1, and 150 to label 2. In both breakage and separation categories, the distribution of the three labels was relatively balanced, approximately in a 1:1:1 ratio, which is favorable for model training and overall performance. In the subsequent experiments, the breakage degree and separation degree were treated as independent classification tasks.

2.5 GarlicNet model

Given the characteristics of the pressure image classification task during garlic clove separation, this paper proposes GarlicNet, a CNN–Transformer network that integrates a dynamic gating mechanism to jointly assess the degree of clove damage and separation, as shown in Figure 7. Unlike natural images, discriminative information in pressure images is typically concentrated in locally high-response regions, subtle changes at contact boundaries, and non-local dependencies across multiple pressure images. Therefore, this task requires the model not only to capture fine-grained local force patterns but also to integrate global contextual relationships among the image sets. Based on these task characteristics, GarlicNet adopts a cascaded CNN–Transformer architecture. As shown in Figure 7a, the front-end CNN first extracts local structural patterns from each pressure image, including pressure peaks, the morphology of compressed regions, and intensity changes, providing stable local inductive biases for subsequent modeling. Subsequently, the Transformer encoder further models long-range dependencies and global contextual relationships based on these features, which possess stronger semantic expressiveness, thereby achieving deep integration of single-image information with group-level information. Compared to a pure CNN, this design enhances the modeling capability of non-local interactions; compared to a pure Transformer, the front-end CNN reduces spatial redundancy, introduces local prior knowledge, and effectively lowers the computational overhead of self-attention. Furthermore, compared to parallel fusion architectures, the serial design features a clearer information flow, enabling gradual feature modeling by first extracting local stress patterns and then integrating global relationships, which better aligns with the intrinsic patterns of the garlic stress image classification task.

Since each sample contains four pressure images, resulting in a total of 12 input channels, this paper introduces an enhanced Dynamic Multi-Channel Convolution Fusion (DMCF) module at the front end of the network (Figure 7b); its specific structure is described in Section 2.4.1. Following the DMCF, this paper employs an ultra-lightweight two-layer CNN backbone network (Figure 7c): first, a 3×3 max-pooling layer with a stride of 2 is used to halve the spatial resolution, and then two layers of 3×3 convolutions are used to expand the number of feature channels from 64 to 512. This backbone introduces only approximately 0.3 million additional parameters, yet it reduces the feature map area to one-quarter of the input size and cuts the GPU memory consumption of the subsequent Transformer encoder by about 40%. More importantly, it makes the feature flow more compact while preserving key local activation patterns, thereby providing more efficient input representations for subsequent attention mechanisms. To further balance representational capacity and computational efficiency, this paper improves the original multi-head self-attention mechanism in the Transformer encoder to a residual dual-branch structure that combines dynamic Top-*k* sparse attention branches; its specific design is detailed in Section 2.4.2.

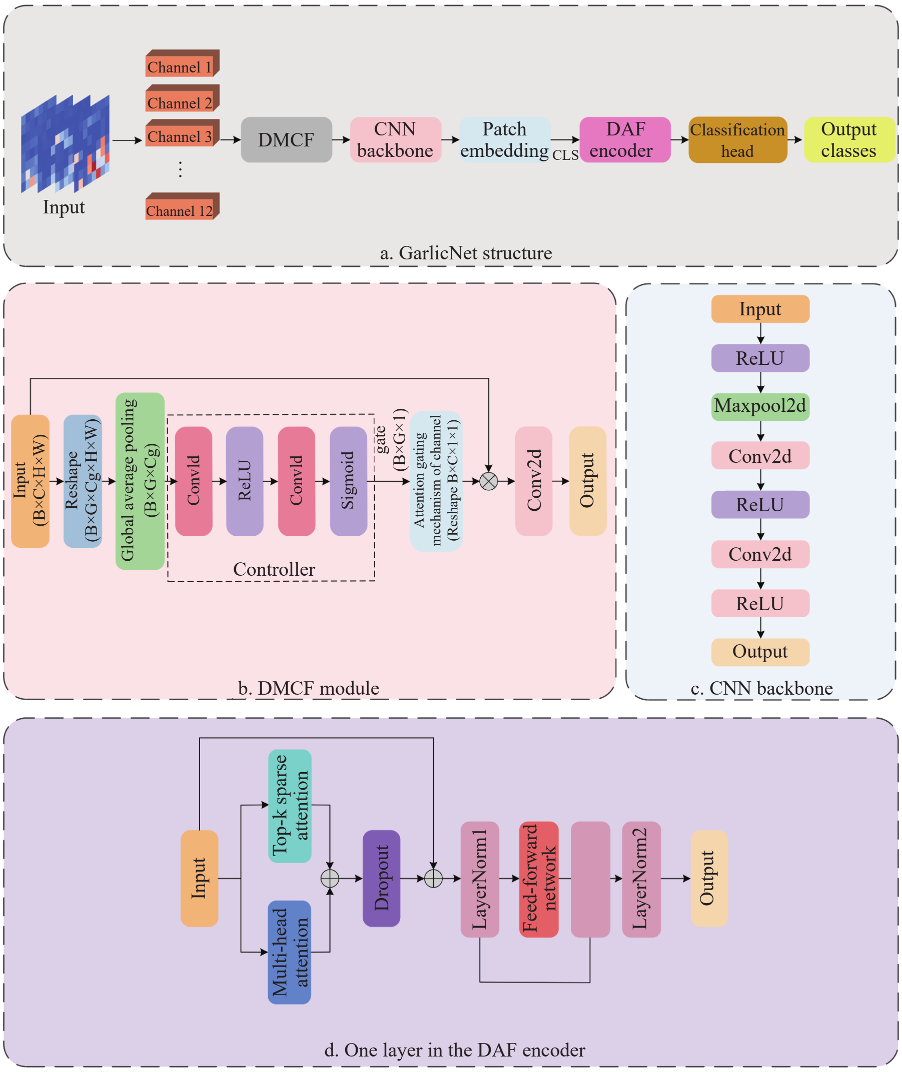


Figure 7 GarlicNet architecture and module details

2.5.1 DMCF module

The DMCF module is a multi-channel convolutional fusion architecture specifically designed for the characteristics of garlic clove pressure images, aiming to enhance representational capacity and computational efficiency during the integration of multi-channel features. Since each sample consists of four pressure images, the input comprises a total of 12 channels. In this scenario, the contribution of different channels to the classification task is not constant: some channels are better at highlighting local stress peaks, while others are more effective at reflecting changes in the contact area or separation status. However, traditional CNNs employ fixed convolution operations across all channels and lack the ability to dynamically adjust channel interactions based on input content, making it difficult to effectively model complex cross-channel dependencies[25].

To address this issue, DMCF combines a dynamic gating mechanism with a channel-grouped convolution strategy. Specifically, the input feature map is first divided into multiple channel-grouped sub-feature sets, and the convolution outputs within each sub-group are then adaptively modulated by dynamic gating. In this way, the network can assign input-relevant importance weights to different channel subgroups for the current sample, selectively enhancing effective responses related to pressure discrimination while suppressing redundant or irrelevant activations. This adaptive, selective fusion mechanism enables more effective cross-channel interaction, thereby yielding more discriminative group pressure feature representations and enhancing the model's adaptability and robustness in the garlic clove pressure image classification task.

Assuming the input feature map is denoted as $X \in R^{B\times C\times H\times W}$, where B is the batch size, C is the number of channels, and H and W represent the height and width, respectively, the computation process proceeds as follows:

First, the input feature map X is partitioned into groups G according to a predefined number, resulting in feature representations $X_g \in R^{B\times G\times C_g\times H\times W}$ for each group, where $C_g = \dfrac{C}{G}$ (Equation (4)).

$$X_g = reshape\left(X, B, G, C_g, H, W\right) \tag{4}$$

Then, global pooling is applied to each group of features to extract the corresponding statistical features $s_g \in R^{B\times G\times C_g}$. Specifically, average pooling is performed over the spatial dimensions H and W of each group feature map X_g (Equation (5)).

$$s_g = mean\left(X_g, \dim[3,4]\right) \tag{5}$$

Subsequently, a two-layer convolutional network with 1×1 kernels is applied to the statistical features of each group to generate the corresponding gating signal $Gate \in R^{B\times G\times 1}$. The output gating signal is then normalized using a sigmoid activation function, as in Equation (6). The generated gating signal $Gate$ is used to modulate the input feature map X through channel-wise multiplication, thereby enabling dynamic weighting of the feature representations, as in Equation (7).

$$Gate = \sigma\left(Conv_{1\times 1}(s_g)\right) \tag{6}$$

$$X_{adjusted} = X \cdot Gate \tag{7}$$

Finally, the adjusted feature map is processed through a convolution layer (3×3 convolution kernel) to generate the final output.

2.5.2 Dual-Branch Attention Fusion structure

In modern deep learning models, attention mechanisms have been widely used to model long-range dependencies in input data. However, when applied to high-resolution or long-sequence inputs, traditional multi-head self-attention (MHSA) mechanisms typically incur high computational and memory costs[26]. More importantly, for the task of classifying images of garlic clove crushing force, the truly discriminative information is often not uniformly distributed across all spatial locations, but rather concentrated in a small number of high-response force regions and their key dependencies.

Based on this characteristic, this paper proposes the Dual-Branch Attention Fusion (DAF) architecture, which combines MHSA with Top-k sparse attention to better balance global dependency modeling capabilities and computational efficiency. In this architecture, the pressure feature representations are fed into two complementary branches for processing simultaneously. Specifically, the MHSA branch is responsible for capturing comprehensive global interaction relationships and retaining sufficient contextual information to model the overall force distribution pattern; meanwhile, the Top-k sparse branch introduces sparsity into the attention map by retaining only the most significant attention connections, thereby highlighting the most informative regions with strong responses and their critical dependencies, while suppressing a large number of weakly correlated or redundant interactions. The outputs of the two branches are ultimately fused using weighted residuals. In this way, the MHSA branch ensures the comprehensiveness of feature representations, while the Top-k branch further enhances feature selectivity, noise resistance, and computational efficiency. By combining dense global modeling with sparse discriminative attention, the proposed DAF module enables GarlicNet to extract pressure information at different granularity levels, making it more suitable for the dual classification task of garlic clove splitting and separation. The specific computational steps are detailed as follows:

1) Multi-head self-attention (MHSA)

The conventional Multi-Head Self-Attention (MHSA) mechanism generates attention distributions by computing dot-product relationships among the Query (Q), Key (K), and Value (V) representations derived from the input features, followed by normalization using the softmax function. The computation of MHSA is given by Equation (8):

$$Attention(Q,K,V) = softmax\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{8}$$

where, $Q,K,V \in R^{B\times N\times D}$, B denotes the batch size; N is the sequence length; and D represents the feature dimension. MHSA concatenates the outputs of multiple attention heads and passes them through a linear transformation to produce the final output.

2) Top-k sparse attention mechanism

In the Top-k sparse attention branch, to reduce computational complexity, only the top k most significant connections are retained for each position. First, attention scores are computed, and a Top-k operation is applied to select the k largest values. These selected scores are then normalized using the softmax function to obtain sparse attention weights. The computation process of Top-k attention is as Equation (9):

$$\text{Top-}k\ Attention(Q,K,V) = soft\max\left(mask\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)\right)V \tag{9}$$

where, the mask is used to set the non-Top-k attention scores to negative infinity, ensuring that only the Top-k highest values are retained.

3) Residual fusion

The outputs of the Multi-Head Self-Attention (MHSA) branch y_{MHSA} and the Top-k Sparse Attention branch $y_{\text{Top-}k}$ are fused using a weighted sum with learnable coefficients. The fusion is computed as follows in Equation (10):

$$y_{fused} = y_{MHSA} + \lambda \cdot y_{\text{Top-}k} \tag{10}$$

where, the fused output is subsequently passed through dropout and normalization layer to facilitate further processing.

3 Experimental setup

3.1 Test environment and data segmentation

The deep learning models in this study were trained on a workstation running the Windows 10 operating system. The hardware configuration included an Intel Core i5-14600KF CPU, an NVIDIA GeForce RTX 4060 Ti GPU with 16 GB of memory, and 32 GB of RAM operating at 6400 MHz. In terms of the software environment, Python 3.12 was used as the programming language, with PyTorch 2.6.0 serving as the primary deep learning framework. GPU acceleration was enabled via CUDA 12.6. The versions used are: NumPy 2.1.2, Pandas 2.2.3, scikit-learn 1.6.1, Pillow 11.0.0, and Matplotlib 3.10.1. To ensure reproducibility and stability of the results, a fixed random seed was set for all experiments. A total of 455 experimental samples were collected for this study. The dataset was split into training and testing sets with a ratio of 3:1. The number of training epochs is set to 50, the batch size to 8, and the image size to 224. The Adam optimizer is used with a learning rate of 1e-4. To reduce the impact of randomness, each dataset is run five times.

3.2 Evaluation metrics

Accuracy (Acc), Recall (R), Precision (P), and F1-score (F1) were employed to evaluate the model's performance in detecting the two classification tasks. The calculation formulas for each metric are as follows in Equations (11)-(14):

$$\mathrm{Acc} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{11}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{12}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{13}$$

$$\mathrm{F1} = \frac{2\times P\times R}{P+R} \tag{14}$$

where, True Positive (TP) refers to the number of instances correctly predicted as positive; True Negative (TN) denotes the number of instances correctly predicted as negative; False Positive (FP) represents the number of false positive cases, where the model incorrectly predicts a negative instance as positive; and False Negative (FN) indicates the number of false negative cases, where the model incorrectly predicts a positive instance as negative.

The macro-averaging method assigns equal importance to each class and reflects the overall predictive performance of the model. In multi-class classification, macro-averaging calculates Precision (Mac_P), Recall (Mac_R), and F1-score (Mac_F1) by first evaluating each class independently and then computing the unweighted average of the results, as Equations (15)-(17), where n denotes the number of classes, which is 3 in this study.

$$\mathrm{Mac\_P} = \frac{\sum_{n}^{i} P_i}{n} \tag{15}$$

$$\mathrm{Mac\_R} = \frac{\sum_{n}^{i} R_i}{n} \tag{16}$$

$$\mathrm{Mac\_}F1 = \frac{\sum_{n}^{i} F1_i}{n} \tag{17}$$

3.3 Baseline model selection

In the model comparison experiments, this study selected three models that rely solely on pure convolutional computations, namely ResNet18[27], DenseNet121[28], and ConvNeXt-Tiny[29]. We selected three models based on the pure Transformer architecture: Vision Transformer (ViT)[30], Swin Transformer[31], and Data-efficient image Transformers (DeiT)[32], as well as three representative CNN–Transformer hybrid architectures. ResNet18 is a convolutional neural network composed of 18 stacked layers connected through residual connections. DenseNet121 is a convolutional neural network that strengthens feature propagation and reuse through dense connections between layers. ConvNeXt-Tiny is a modernized architecture that adopts larger convolution kernels and optimized design strategies to improve feature representation ability. ViT is a pure Transformer architecture designed specifically for image recognition tasks. Swin Transformer is a hierarchical Vision Transformer that models visual features through self-attention within shifted local windows. DeiT is a Vision Transformer designed to improve image classification performance under limited training data. CoAtNet[33] and ConTNet[34] are architectures that connect CNNs and transformers in series, while Conformer[35] is an architecture that connects them in parallel. To ensure the fairness of the comparative experiments and the reliability of the results, all models were trained using the same random seed, and the training parameters were kept as consistent as possible. Model performance was evaluated using four metrics, namely Acc, Mac_P, Mac_R, and Mac_F1. Each model was independently run five times, and the final results were reported as the average values.

3.4 Ablation strategy

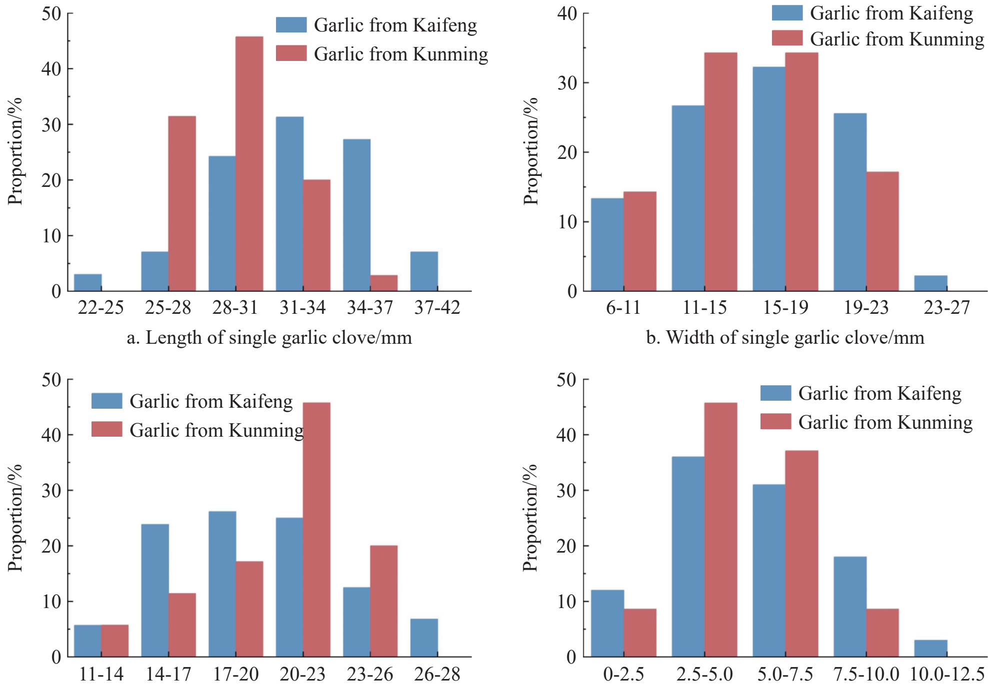
To evaluate the contribution of each module to the overall performance of the model, a series of ablation studies were conducted. In each experiment, a key component was either removed or replaced, and the resulting changes in model performance were observed. Specifically, the DMCF module was removed and replaced with a standard Conv2d layer to assess its impact on channel-wise feature selectivity. For the DAF structure, we first substituted the attention mechanism with a conventional feedforward neural network to evaluate the overall effectiveness of attention. Then, we replaced the dual-branch structure with a single-branch-only MHSA to examine the role of the Top-k residual branch in compensating for information loss. Each model configuration in the ablation study was executed five times, and the average results were reported to ensure robustness and minimize the impact of randomness.

4 Results and discussion

4.1 Analysis of the physical properties of individual garlic cloves

Figure 8 shows the geometric dimensions and mass distributions of 100 individual garlic cloves from each of the two varieties. Overall, clove length, width, thickness, and mass exhibit approximately normal distributions. Some differences are observed between the two varieties. Compared with the Kunming variety, the Kaifeng variety shows greater dispersion in length and thickness, indicating higher variability among individual cloves. In contrast, the Kunming variety exhibits more concentrated distributions and relatively greater uniformity in size and mass. In terms of width, both varieties show noticeable variability. These results reflect the natural variability of the samples used in subsequent garlic clove crushing experiments.

As shown in Table 1, the measurements of both garlic varieties exhibit small standard deviations and coefficients of variation, indicating good stability and consistency. The moisture content of Kaifeng garlic is lower than that of Kunming garlic, mainly because the former undergoes natural air-drying before clove breaking, resulting in noticeable moisture loss. In contrast, Kunming garlic is freshly harvested and retains a higher moisture content. As a result, Kaifeng garlic tends to be more brittle and easier to break during the garlic clove-breaking process, whereas Kunming garlic, with its tighter clove wrapping and higher moisture content, generally requires greater breaking force and is more prone to incomplete separation. Since these two varieties represent different physical states commonly encountered in practical production, combining their pressure signal data for model development helps cover a wider range of actual operating conditions. This improves the representativeness and rationality of the dataset and enhances the robustness and generalization ability of the proposed model.



a. Length of single garlic clove/mm b. Width of single garlic clove/mm

c. Thickness of single garlic clove/mm d. Mass of single garlic clove/g

Figure 8 Statistical analysis of geometric dimensions and weight of individual garlic cloves

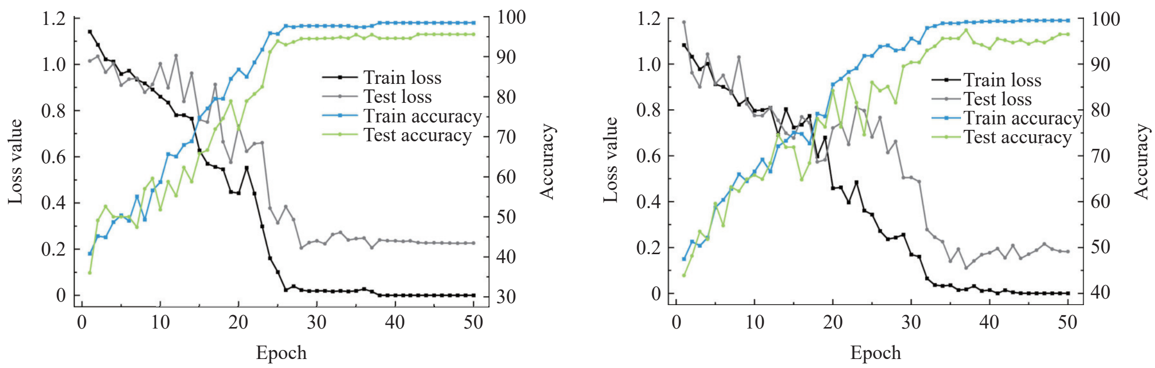
Table 1 Moisture content analysis of garlic cloves from different varieties

Sample	Mean value/%	Standard deviation	Coefficient of variation/%
Kaifeng garlic	60.9	0.033	5.42
Kunming garlic	62.5	0.027	4.32

4.2 Model performance analysis

Model performance was evaluated based on the changes in training and testing loss during training, along with standard metrics including accuracy (Acc), recall (R), precision (P), and F1-score (F1). The confusion matrix provides an intuitive representation of the prediction results for each class, where rows correspond to the actual class labels and columns represent the predicted class labels.

As shown in Figure 9, for the two classification tasks associated with the garlic clove-breaking process—namely, breakage and separation degree—the proposed model demonstrates clear convergence trends in both training and testing loss, as well as accuracy, indicating strong generalization capability. Specifically, for breakage classification, the training loss rapidly decreases from an initial value of 1.1 to below 0.05 within the first 25 epochs, and then stabilizes near zero. The testing loss similarly drops to approximately 0.22 within 30 epochs and remains stable thereafter. Meanwhile, training accuracy surpasses 90% after 22 epochs and eventually stabilizes around 98.5%. The testing accuracy also increases rapidly, plateauing at approximately 95.6%. The small gap between training and testing accuracy indicates the absence of significant overfitting. For separation degree classification, the model exhibits a similar convergence pattern. The training loss decreases from approximately 1.18 to below 0.1 within 32 epochs, while the testing loss converges around 0.18. Both training and testing accuracy steadily improve throughout training, ultimately reaching approximately 99.5% and 96.5%, respectively. Overall, the model achieves stable convergence for both classification tasks within approximately 25 epochs. The performance gap between training and testing sets remains around 3%, demonstrating that the proposed architecture maintains both high learning efficiency and strong generalization performance in classifying breakage and separation degree.



a. Using the degree of breakage as the target label b. Using the degree of separation as the target label

Figure 9 Changes in training and testing loss and accuracy

As shown in the Tables 2 and 3 for the three-class classification tasks, the model achieved an overall accuracy of 95.6% in predicting the degree of breakage. All samples labeled as slight breakage were correctly classified, with Mac_R, Mac_P, and Mac_F1 all reaching 100%. In the case of moderate breakage, the recall, precision, and F1-score were 90.2%, 97.4%, and 93.7%, respectively, with the primary source of error stemming from four samples being misclassified as severe breakage. For severe breakage, the recall, precision, and F1-score were 97.4%, 90.4%, and 93.8%, respectively, with only one sample misclassified as moderate breakage. The Mac_R, Mac_P, and Mac_F1 were 95.9%, 95.9%, and 95.8%, respectively, indicating that the model effectively distinguishes between breakage levels, though minor confusion still persists between adjacent categories. For separation degree classification, the model performed even better, achieving an accuracy rate of 96.5%. Misclassifications between high-separability and medium-separability samples were extremely rare, with only one sample misclassified in each category (for high-separability: recall=97.1%, precision=91.9%, F1 score=94.4%; for medium-separability: recall=97.6%, precision=97.6%, F1 score=97.6%). Among the low-resolution samples, only two were misclassified as high-resolution (recall=94.7%, precision=100.0%, F1 score=97.3%). There were no instances of misclassification between the medium- and low-resolution samples, indicating that the pressure images between these two resolution levels exhibit distinct features that facilitate differentiation.

The prediction results indicate that the proposed GarlicNet model exhibits high classification performance in classifying both breakage and separation levels, demonstrating its strong capability to effectively distinguish between different grade-specific features. This highlights the model's robustness and reliability in garlic clove breaking assessment tasks. Notably, the model shows excellent performance in recognizing extreme cases such as slight breakage and high separation, which further confirms its suitability for practical applications in breakage and separation detection. However, a small number of misclassifications were observed between adjacent categories (e.g., moderate and severe breakage), suggesting that distinguishing subtle boundaries between neighboring levels remains a challenging aspect.

Table 2 Prediction results for each class of degree of breakage

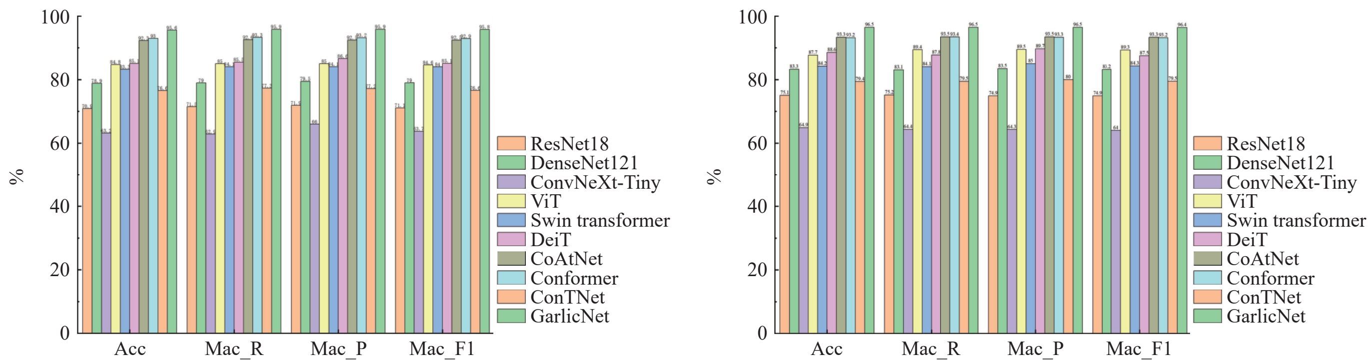
Breakage		Predicted label			R	p	F1
		0	1	2			
True label	0	34	0	0	100%	100%	100%
	1	0	37	4	90.2%	97.4%	93.7%
	2	0	1	38	97.4%	90.4%	93.8%
Macro average					95.9%	95.9%	95.8%

Table 3 Prediction results for each class of degree of separation

Separation		Predicted label			R	p	F1
		0	1	2			
True label	0	34	1	0	97.1%	91.9%	94.4%
	1	1	40	0	97.6%	97.6%	97.6%
	2	2	0	36	94.7%	100.0%	97.3%
Macro average					96.5%	96.5%	96.4%%

4.3 Model comparison experiments

Figure 10 compares GarlicNet with nine representative baseline models on two classification tasks, namely degree of breakage and degree of separation. For the degree-of-breakage task, compared with Conformer, which is the best baseline on this task, GarlicNet improves these four metrics by 2.6, 2.6, 2.7, and 2.9 percentage points, respectively. Compared with CoAtNet, the gains are 3.3, 3.3, 3.4, and 3.3 percentage points, respectively. For the degree-of-separation task, compared with the strongest baselines on this task, the gains over CoAtNet are 3.2, 3.0, 3.0, and 3.1 percentage points, while the gains over Conformer are 3.3, 3.1, 3.2, and 3.2 percentage points, respectively. These results indicate that GarlicNet not only improves overall classification accuracy, but also achieves a more balanced performance across classes.



a. Degree of breakage b. Degree of separation

Figure 10 Performance comparison of different models

Among the CNN-based models, DenseNet121 consistently outperforms ResNet18 and ConTNet on both tasks, indicating that denser feature reuse is beneficial for extracting informative local patterns from pressure images. In contrast, ConvNeXt-Tiny shows the weakest performance among all compared models, suggesting that under the relatively limited-data setting of this study, its pure convolutional design does not generalize well to this task. Among the pure Transformer models, DeiT performs slightly better than ViT and Swin Transformer on the breakage task and remains competitive on the separation task, implying that data-efficient training is helpful for pressure-image classification. However, all pure Transformer models still lag behind the hybrid architectures. For example, GarlicNet improves breakage-task accuracy by 10.8, 12.3, and 10.5 percentage points over ViT, Swin Transformer, and DeiT, respectively, and separation-task accuracy by 8.8, 12.3, and 7.9 percentage points. The gains over CNN baselines are even larger, reaching 16.7-32.4 percentage points for breakage and 13.2-31.6 percentage points for separation. This performance pattern shows that neither pure local modeling nor pure global modeling is sufficient for the current task. The above results suggest that pressure-image classification requires simultaneous modeling of localized pressure concentration patterns and global spatial dependency structures. CNN-based models are effective at capturing local textures and short-range spatial cues, but they are less capable of modeling long-range dependencies. In contrast, pure Transformer models can better describe global relationships, but their weaker convolutional inductive bias limits their ability to extract stable local structures, especially when the sample size is limited. Therefore, hybrid architectures such as CoAtNet and Conformer, which combine convolution and self-attention, achieve substantially better results than either pure CNNs or pure Transformers.

GarlicNet further improves upon these hybrid baselines because its architecture is more specifically tailored to the present task. The dual-branch attention mechanism enables the network to capture complementary local and global discriminative cues from pressure images, while the dynamic gating mechanism adaptively reweights different feature streams according to the characteristics of each input sample. This design is particularly beneficial for distinguishing subtle inter-class differences in breakage and separation severity, where local pressure concentration, shape continuity, and overall force-distribution patterns all contribute to the final label. As a result, GarlicNet achieves more accurate and more robust classification than the compared baseline models.

In addition, all models perform better on the degree-of-separation task than on the degree-of-breakage task. Depending on the model and metric, the performance improvement ranges from 0.1 to 4.4 percentage points. This indicates that the class boundaries for separation degree are more distinct, whereas the pressure patterns associated with breakage degree are more overlapping and therefore more difficult to classify correctly.

To further validate the proposed model's classification performance across different categories, this paper plots the confusion matrices for the proposed model and the best-performing models from each architecture, as shown in Figure 11. Unlike comparisons based solely on overall accuracy metrics, confusion matrices reveal the model's prediction distribution characteristics and misclassification patterns at the category level. Figure 11 shows that the sample distribution of the proposed model is more concentrated along the diagonal, indicating higher recognition accuracy for samples across all categories; simultaneously, the distribution of errors in the off-diagonal regions is significantly reduced, suggesting that the model effectively minimizes cross-class confusion.

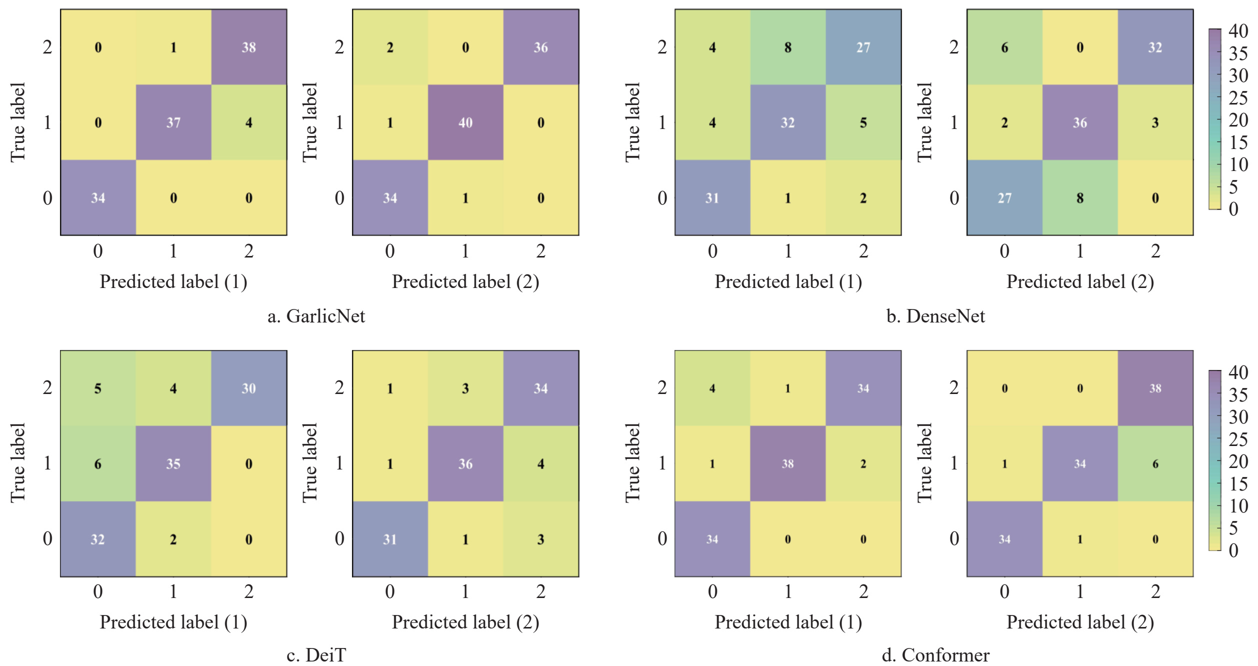


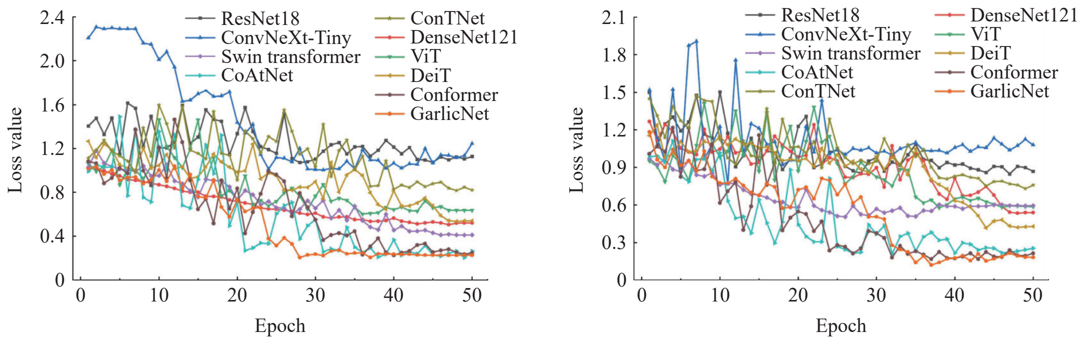
Figure 11 Confusion matrix for the degree of breakage (1) and separation (2) prediction

Regardless of whether predicting damage severity or separation degree, the baseline models DenseNet and DeiT exhibit significant misclassification across all three categories. In the hybrid architecture Conformer, confusion among the three categories is effectively mitigated. Regarding damage severity prediction, Conformer misclassifies Category 1 and Category 2 as the other two categories, whereas GarlicNet exhibits only minor misclassifications for Category 1 and Category 2. Regarding separation, while Conformer did not exhibit classification confusion for Class 2, it produced significantly more misclassifications for Class 1 compared to GarlicNet. These results indicate that the proposed model can more effectively extract discriminative features between different classes, thereby enhancing the ability to distinguish between similar classes. This is because the proposed method enhances the completeness and discriminative power of feature representations through dynamic cross-channel fusion and global complementary discrimination, thereby achieving more balanced and stable classification results while improving overall performance.

The testing-loss curves in Figure 12 provide further support for the above findings. GarlicNet exhibits the fastest convergence and the lowest and most stable loss values on both tasks, indicating superior feature extraction capability and optimization stability. CoAtNet and Conformer also show strong convergence behavior, but with slightly larger fluctuations. Other models exhibit relatively poor performance, with slower convergence and consistently higher loss values.

4.4 Ablation study

The results of the dissolution test are shown in Table 4. The full GarlicNet integrates both the DMCF module and the DAF architecture, achieving state-of-the-art performance on both classification tasks—garlic clove damage and separation—while requiring only 18.56 million parameters and 6.26 GFLOPs, demonstrating that the model strikes a good balance between accuracy and computational complexity. When standard convolutions were used in place of DMCF while retaining the dual-branch attention structure, the model complexity remained virtually unchanged, but the accuracy for the two tasks dropped to 90.4% and 91.2%, respectively. This indicates that the performance improvement does not stem from higher model complexity, but is primarily attributable to DMCF's ability to dynamically calibrate channel features. This module enhances key responses and suppresses redundant information, thereby improving the model's ability to distinguish subtle category differences. When retaining DMCF but replacing the DAF structure with a traditional feedforward network (FFN), the model size was reduced to 12.23 million parameters and 6.08 GFLOPs; however, performance degraded significantly, with accuracy dropping to 33.3% and 35.9%, and F1-scores falling below 18% in both cases. This indicates that simple linear mappings struggle to effectively model complex feature dependencies and cannot extract representations that support reliable classification; thus, the DAF structure is a critical component for ensuring stable model learning. Furthermore, when the Top-*k* sparse attention branches were removed and only the MHSA main branch was retained, the accuracy rates for the two tasks were 92.1% and 93.0%, respectively—still approximately 3 percentage points lower than the full model. This indicates that, in addition to global modeling, the Top-*k* branches can further supplement local fine-grained information, aiding in the identification of subtle differences between categories.



a. Degree of breakage b. Degree of separation

Figure 12 Loss values on the test set for different models

Table 4 Ablation study of GarlicNet

DMCF	DAF	Parameters	Flops	Acc	Mac_R	Mac_P	Mac_F1
√	√	18.56M	6.26G	95.6%	95.9%	95.9%	95.8%
				96.5%	96.5%	96.5%	96.4%
	√	18.53M	6.21G	90.4%	90.5%	91.2%	90.4%
				91.2%	90.9%	92.1%	91.0%
√	FFN replace	12.23M	6.08G	33.3%	33.3%	11.1%	16.7%
				35.9%	33.3%	12.0%	17.6%
√	MHSA replace	13.85M	6.41G	92.1%	92.2%	92.7%	92.4%
				93.0%	92.9%	93.6%	93.1%

Note: In each set of experiments,the first row uses breakage degree as the label; the second row uses separation degree as the label.

Overall, DMCF and DAF are the core modules responsible for enhancing GarlicNet's performance, while the Top-*k* sparse attention branch further strengthens its feature refinement capabilities. The outstanding performance of the complete model stems from the synergistic interaction of dynamic channel calibration, adaptive feature modeling, and the global-local complementary attention mechanism, rather than simply relying on an increase in the number of parameters or computational complexity. This enables the model to achieve the best detection performance for both breakage and separation tasks, all while maintaining a moderate computational footprint.

4.5 Discussion

This study is the first to systematically explore the use of pressure data collected during the garlic clove-breaking process for predicting both the degree of breakage and the degree of separation, and to develop a dedicated model GarlicNet tailored to this application. The results show that GarlicNet achieves prediction accuracies of 95.6% and 96.55% for breakage and separation, respectively, successfully enabling classification across different severity levels. These findings validate the central hypothesis of this work: that pressure experienced during the clove-breaking process is directly correlated with both breakage and separation outcomes. Furthermore, the study broadens the applicability boundary of image classification algorithms. From a theoretical standpoint, the results demonstrate that GarlicNet effectively captures discriminative features across different classes in pressure images and accurately distinguishes subtle differences among severity levels. This confirms the feasibility of classifying breakage and separation conditions based solely on pressure information obtained during the clove-breaking process. From a practical perspective, the study provides valuable guidance for garlic clove breaking and cultivation. Real-time monitoring of breakage and separation conditions enables timely adjustment of the spacing between the conical disc and the rotating plate, helping to reduce breakage rates, enhance separation efficiency, minimize garlic damage and waste, improve planting success rates, and ultimately boost yield.

This study still presents several limitations that should be addressed in future work. First, the dataset size is relatively limited, with a total of 455 samples collected, averaging approximately 150 per class. Such a sample size may constrain the model's ability to generalize under more complex operating conditions. Second, the experiments did not incorporate size or weight grading of garlic cloves, and all samples were treated uniformly in the modeling process. However, in practical seeding scenarios, the selection of high-quality cloves significantly impacts both yield and crop quality. Third, although Section 3.2 demonstrates a strong correlation between breakage rate and separation rate, the study did not explore strategies for achieving an optimal balance between the two by adjusting the spacing between the conical disc and the rotating plate.

Future work should expand the dataset and include multi-grade garlic cloves to verify the model's applicability across varying quality levels. Additionally, optimization algorithms should be employed to systematically investigate the parameter combinations that minimize breakage while maximizing separation. Such efforts would support the development of intelligent, precision-oriented clove-breaking equipment.

In conclusion, this study not only offers new empirical evidence for predicting garlic breakage and separation degrees using pressure signals during the clove-breaking process, but also lays a solid foundation for future theoretical advancements and practical implementations, with meaningful implications for real-world applications.

5 Conclusions

This study pioneers the transformation of pressure signals from garlic clove separation processes into spatiotemporal images, proposing GarlicNet—a hybrid CNN–Transformer architecture integrating Dynamic Multi-Channel Convolution Fusion (DMCF) and Dual-Branch Attention Fusion (DAF). The framework achieves breakage detection (95.6% accuracy) and separation assessment (96.5% accuracy) by simultaneously: (1) Capturing force-distribution signatures in pressure images; (2) Modeling local channel-specific features via DMCF; (3) Learning global dependencies through DAF. GarlicNet outperformed CNN architecture (ResNet18/DenseNet121/ConvNeXt-Tiny Δacc = 16.7%-32.4%/13.2%-31.6%), Transformer architecture (ViT/Swin Transformer/DeiT Δacc = 10.5%-12.3%/7.9%-12.3%), and hybrid architecture benchmarks (CoAtNet/Conformer/ConTNet Δacc = 2.6%-19%/3.2%-17.1%) while demonstrating superior training stability, validating pressure imaging as an efficient feature representation for mechanical damage monitoring.

Building on this foundation, future research will extend validation to diverse garlic quality grades. In addition, we will explore real-time adaptive control for spacing during clove separation, aiming to promote the reliable application of this method in complex agricultural environments.

Acknowledgements

This research was funded by the central government for guiding local science and technology development (Grant Number 2024ZY044). The authors would like to thank the editor and anonymous reviewers for their helpful suggestions on improving the quality of the paper.

[References]

[1] Wakamatsu J, Stark T D, Hofmann T. Antioxidative maillard reaction products generated in processed aged garlic extract. Journal of Agricultural and Food Chemistry, 2019; 67(8): 2190–200.

[2] Shang A, Cao S Y, Xu X Y, Gan R Y, Tang G Y, Corke H, et al. Bioactive compounds and biological functions of garlic (*Allium sativum* L.). Foods, 2019; 8(7): 246. DOI: 10.3390/foods8070246.

[3] FAO. Faostat: FAO Statistical Databases. Food & Agriculture Organization of the United Nations (FAO), 2023. Available: https://www.fao.org/faostat/es/#data/QCL.

[4] Shagun S, Bains A, Sridhar K, Dhull S B, Patil S, Gupta V K, et al. A comprehensive review on impact of post-harvest management and treatment practices on the quality of garlic (*Allium sativum* L) during storage. Scientia Horticulturae, 2024; 337: 113586.

[5] Parde S R, Kausal R T, Jayas D S, White N D G. Mechanical damage to soybean seed during processing. Journal of Stored Products Research, 2002; 38(4): 385–394.

[6] Wang J W, Chen H X, Li Q H, Huo S W, Wang Q, Tang H, et al. Study on rice sprout damage in UAV direct seeding with auger mechanisms. Computers and Electronics in Agriculture, 2025; 229: 109809.

[7] Gu R L, Huang R, Jia G Y, Yuan Z P, Ren L S, Li L, et al. Effect of mechanical threshing on damage and vigor of maize seed threshed at different moisture contents. Journal of Integrative Agriculture, 2019; 18(7): 1571–1578.

[8] Xu P F, Li N, Xu H C, Song J L, Xia L M, Jin C Q. Comparisons and analysis of rubber roller and conical disc garlic breaking technology. Journal of Chinese Agricultural Mechanization, 2020; 41(1): 95–99. (in Chinese)

[9] Shao M X, Zhang X M, Fang L Z, Ma G C, Tian J, Fu S H. Research progress on mechanized garlic seeding technology in China. Int J Agric & Biol Eng, 2025; 18(2): 9–20.

[10] Li N, Sun Q, Xu P F, Song J L, Xia L M, Jin C Q. Experimental study on conical disc garlic breaking device. Journal of Agricultural Mechanization Research, 2020; 42(2): 168–172. (in Chinese)

[11] Song Y Q, Jiao J K, Wu L Y, Chi B W, Chen T Y, Zhang B F, et al. Effect of mechanical compression of garlic clove on emergence and growth of *Allium sativum* L. Transactions of the CSAE, 2019; 35(5): 26–32. (in Chinese)

[12] Zhang M, Liang Z, Tong S. An image identification system for agricultural items based on machine vision. Journal of Combinatorial Mathematics Combinatorial Computing, 2024; 120: 169–75.

[13] Gao Y, Ma X D, Xia Z N, Qi T, Wang X, He Z, et al. Garlic-YOLO-DD: a lightweight object detection algorithm for garlic damage detection. Frontiers in Plant Science, 2026; 16: 1702045.

[14] Diaz R, Faus G, Blasco M, Blasco J, Moltó E. The application of a fast algorithm for the classification of olives by machine vision. Food Research International, 2000; 33(3): 305–309.

[15] Kuan Y N, Goh K M, Lim L L. Systematic review on machine learning and computer vision in precision agriculture: Applications, trends, and emerging techniques. Engineering Applications of Artificial Intelligence, 2025; 148: 110401.

[16] Zhang Y Z, Song J L, Zhou L. Real-time detection system of garlic clove bud for garlic clove orientation metering device based on machine vision. Computers and Electronics in Agriculture, 2025; 234: 110300.

[17] Son J H, Park H G, Han Y J, Kang S H, Woo S M, Ha Y S. Prediction of garlic clove volume and mass using a depth camera and machine learning models. Postharvest Biology and Technology, 2025; 226: 113526.

[18] Li Y T, Tan L H, Zhong Z H, He L Y, Chen J N, Wu C Y, et al. Improved YOLOv8 network using multi-scale feature fusion for detecting small tea shoots in complex environments. Int J Agric & Biol Eng, 2025; 18(5): 223–233.

[19] Chen L, Yuan Y, He H Y. Image recognition for crop diseases using a novel multi-attention module. Int J Agric & Biol Eng, 2024; 18(1): 238–244.

[20] Chen D X, Liu L, Zhao L, Shi Y, Wang F, Meng D, et al. Method for lightweight tomato leaf disease recognition based on improved YOLOv11s. Int J Agric & Biol Eng, 2025; 18(5): 298–305.

[21] Kallingal A V. The integration of pressure sensors in the field of robotics. International Journal of Science and Research Archive, 2025; 14(3): 30–35.

[22] Borugadda P, Kalluri H K. A comprehensive analysis of artificial intelligence, machine learning, deep learning and computer vision in food science. Journal of Future Foods, 2026; 6(6): 971–987.

[23] Geng A J, Li X Y, Hou J L, Zhang Z L, Zhang J, Chong J. Design and experiment of automatic directing garlic planter. Transactions of the CSAE, 2018; 34(11): 17–25. (in Chinese)

[24] Tullo Wako A, Abera Haptegebriel S. Performance evaluation of a power-operated garlic (*Allium sativum* L.) bulb breaker for clove separation. Heliyon, 2024; 10(8): e29835.

[25] Chen Y P, Dai X Y, Liu M C, Chen D D, Yuan L, Liu Z C. Dynamic convolution: attention over convolution kernels. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020; pp.13–19. DOI: 10.1109/CVPR42600.2020.01104.

[26] Wang S, Li B Z, Khabsa M, Fang H, Ma H. Linformer: Self-attention with linear complexity. arXiv: 2006; 04768v3. DOI: 10.48550/arXiv.2006.04768.

[27] He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016; DOI: 10.1109/CVPR.2016.90.

[28] Huang G, Liu Z, Maaten L V D, Weinberger K Q. Densely connected convolutional networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. DOI:10.1109/CVPR.2017.243.

[29] Liu Z, Mao H, Wu C Y, Feichtenhofer C, Darrell T, Xie S. A ConvNet for

the 2020s. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. DOI: 10.1109/CVPR52688.2022.01167.

[30] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. International Conference on Learning Representations, 2021. DOI: 10.48550/arXiv.2010.11929.

[31] Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. IEEE/CVF International Conference on Computer Vision, 2021. DOI: 10.48550/arXiv.2103.14030.

[32] Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, J'egou H E. Training data-efficient image transformers & distillation through attention. International Conference on Machine Learning, 2020. DOI: 10.48550/arXiv.2012.12877.

[33] Dai Z, Liu H, Le Q V, Tan M. CoAtNet: marrying convolution and attention for all data sizes. Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021. DOI: 10.48550/arXiv.2106.04803.

[34] Yan H T, Li Z, Li W J, Wang C H, Zhang C. ConTNet: Why not use convolution and transformer at the same time? arXiv: 2104; 13497v3. DOI: 10.48550/arXiv.2104.13497.

[35] Peng Z L, Huang W, Gu S Z, Xie L X, Wang Y W, Jiao J B, et al. Conformer: Local features coupling global representations for visual recognition. arXiv: 2105; 03889v1. DOI: 10.48550/arXiv.2105.03889.