RESEARCH-ARTICLE

Data-driven Root Tuber Biomass Estimation via a Wireless Network

TAO WANG

YANG ZHAO

YANG PENG

JINGHUA WANG

JIE LIU

YANG ZHONG

Published: 03 March 2026
Accepted: 12 February 2026
Revised: 26 January 2026
Received: 21 November 2025

Data-driven Root Tuber Biomass Estimation via a Wireless Network

TAO WANG, Harbin Institute of Technology Shenzhen, Shenzhen, China and State Key Laboratory of Smart Farm Technologies and System, Harbin, China
YANG ZHAO*, Harbin Institute of Technology Shenzhen, Shenzhen, China and State Key Laboratory of Smart Farm Technologies and System, Harbin, China
YANG PENG, The Chinese University of Hong Kong - Shenzhen, Shenzhen, China
JINGHUA WANG, Harbin Institute of Technology Shenzhen, Shenzhen, China and State Key Laboratory of Smart Farm Technologies and System, Harbin, China
JIE LIU, Harbin Institute of Technology Shenzhen, Shenzhen, China and State Key Laboratory of Smart Farm Technologies and System, Harbin, China
YANG ZHONG, Chinese Academy of Agricultural Sciences Agricultural Genomics Institute at Shenzhen, Shenzhen, China

Below-ground biomass (BGB) of root tubers is an important phenotypic trait in crop monitoring and other agricultural applications. This paper proposes a novel tuber biomass sensing (TBS) framework that uses internet of things (IoT) devices to enable non-destructive estimation of below-ground root tuber biomass. Specifically, we perform extensive experiments to build a new BGB dataset with more than 700,000 received signal strength (RSS) measurements collected by our low-cost wireless network. Then, we propose a novel data-driven model that integrates convolution neural networks, residual connections, and attention mechanisms to facilitate discriminative feature extraction from RSS data and achieve state-of-the-art (SOTA) performance in biomass estimation. In addition, to mitigate performance degradation caused by imbalanced training data, we propose a contrastive learning method that aligns feature representations of samples with similar biomass values while increasing the separation between those with significantly different values. This method reduces estimation bias toward high-frequency biomass labels, thereby improving the performance and generalizability of the data-driven model. Experimental results demonstrate the efficacy of the proposed TBS framework. Our dataset and pre-trained models are publicly available on https://zenodo.org/records/15000852.

CCS Concepts: • **Computer systems organization → Sensor networks**.

Additional Key Words and Phrases: Below-ground biomass, Wireless network, Deep learning

*Yang Zhao is the corresponding author.

Authors' Contact Information: Tao Wang, Harbin Institute of Technology Shenzhen, Shenzhen, China and State Key Laboratory of Smart Farm Technologies and System, Harbin, China; e-mail: hitszwangtao@gmail.com; Yang Zhao, Harbin Institute of Technology Shenzhen, Shenzhen, China and State Key Laboratory of Smart Farm Technologies and System, Harbin, China; e-mail: yang.zhao@hit.edu.cn; Yang Peng, The Chinese University of Hong Kong - Shenzhen, Shenzhen, Guangdong, China; e-mail: yangpeng@link.cuhk.edu.cn; Jinghua Wang, Harbin Institute of Technology Shenzhen, Shenzhen, Guangdong, China and State Key Laboratory of Smart Farm Technologies and System, Harbin, China; e-mail: wangjh2012@foxmail.com; Jie Liu, Harbin Institute of Technology Shenzhen, Shenzhen, Guangdong, China and State Key Laboratory of Smart Farm Technologies and System, Harbin, China; e-mail: jieliu@hit.edu.cn; Yang Zhong, Chinese Academy of Agricultural Sciences Agricultural Genomics Institute at Shenzhen, Shenzhen, Guangdong, China; e-mail: zhongyang@caas.cn.

ACM 2577-6207/2026/3-ART
https://doi.org/10.1145/3800586

1 Introduction

Below-ground biomass (BGB) is an important phenotypic trait that contributes to the development of strong and abundant root systems [9], higher crop productivity and yields [43], and crop breeding [2]. With the advancement of IoT technology, numerous methods have been developed for biomass estimation [6, 42]. For example, [36] combines a multispectral camera sensor with an unmanned aerial vehicle (UAV) to estimate the above-ground biomass (AGB) of potatoes. Although various methods have been proposed to estimate AGB [22, 44], it remains an open problem to estimate BGB of root tubers due to their underground nature [21].

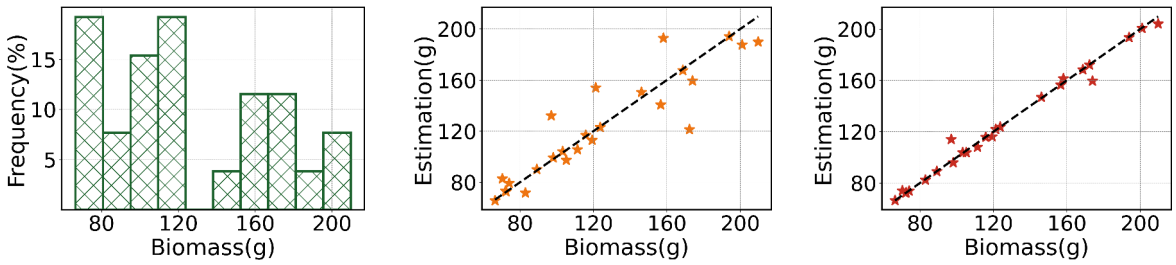
To enable non-destructive estimation of BGB, various studies [65, 66] have developed sensing systems based on ground penetrating radar (GPR), which uses radio frequency signals to penetrate soil. For example, [41] uses GPR to detect below-ground potato tubers and applies thresholding along with continuous wavelet transform analysis to estimate potato biomass. Although GPR has the potential for BGB estimation, its resolution is limited by the number of antennas [5], resulting in constrained sensing accuracy and range. In addition, the high cost of GPR devices hinders their widespread adoption in agriculture. In this paper, we propose an alternative approach: a novel TBS framework that uses low-cost IoT devices to estimate BGB of root tubers in a non-destructive way.

The TBS framework is inspired by the use of wireless network signals for non-line-of-sight (NLOS) sensing [1, 4, 31]. For example, [34] and [1] use wireless signals to assess fruit ripeness, while [31] uses RSS data to detect underground structures in coal mines. In addition, recent studies have shown that deep neural networks (DNNs) achieve SOTA performance in crop trait estimation [14, 26, 35, 58, 64]. For example, [14] and [26] use a multilayer perceptron (MLP) to estimate the biomass of soybeans and trees, respectively. [35] applies a deep convolution neural network (DCNN) to estimate above-ground rice biomass, while [64] employs a transformer network to estimate grass biomass. Although these systems and DNNs have achieved promising results in NLOS sensing and crop phenotypic trait estimation, several challenges remain to accurately estimate BGB of root tubers.

First, current wireless network-based sensing systems remain inadequate for estimating BGB of root tubers. Existing studies employ wireless networks composed of a limited number of sensor nodes or antennas for the detection of fruit ripeness [1, 34]. However, BGB estimation requires wireless signals to penetrate soil, which can cause greater attenuation and variability of wireless signals. In addition, although DNN models have demonstrated SOTA performance in estimating crop phenotypic traits [18, 22] and have shown great potential for below-ground sensing [50], they require a large amount of training data. To the best of our knowledge, no publicly available dataset exists for the networked wireless sensing of BGB of root tubers. In practice, collecting wireless data for below-ground tubers is labor-intensive and time-consuming, due to variations in tuber dimensions, positions, and weights.

Second, due to the diversity and continuity of biomass values, it is difficult for the training data to cover all possible biomass labels. Moreover, tubers with different shapes and sizes may exhibit identical or similar biomass, leading to an imbalanced data distribution. This means that certain biomass labels may be less observed than others or even missed during training, resulting in estimation bias in DNN models. For example, as shown in Fig. 1a, we construct an RSS dataset in a greenhouse for potato tuber BGB estimation, which exhibits an imbalanced distribution of training samples. Specifically, the number of samples with biomass labels in the ranges of 80-100, 120-140, and 170-190 is smaller than that in other ranges. Consequently, the parity plot in Fig. 1b demonstrates that the existing DNN model [47] yields larger estimation errors for these underrepresented ranges.

In addition to training data collection, further improvements are needed in DNN model design for BGB estimation. Our investigation and evaluation reveal that DNN models used in previous studies [47, 59] are insufficient to extract discriminative features from multi-sensor wireless data. Specifically, we use saliency maps [46] to quantify the contributions of RSS measurements from different wireless links in BGB estimation. These contributions are then used as link weights to visualize the wireless network, highlighting the links that strongly influence the estimation results. As shown in Fig. 2a, the MLP-based model [59] relies on many links

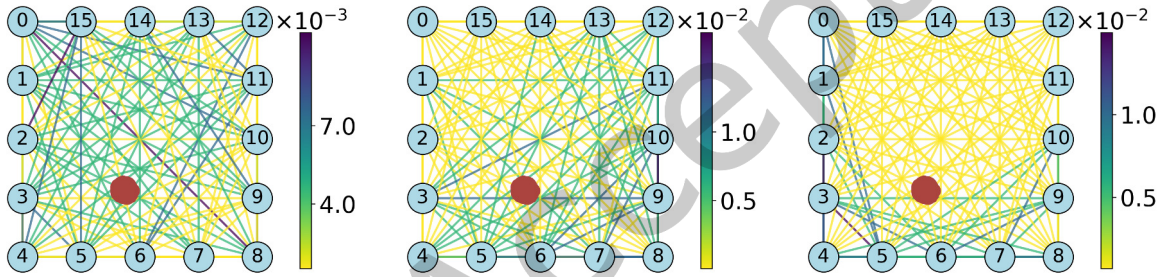


(a) Histogram of biomass label distribution.

(b) Parity plot of the CNN-based model [47].

(c) Parity plot of our model.

Fig. 1. Distribution of training biomass labels and estimation performance of different models. (a) Histogram of training biomass labels, showing data imbalance. (b) Parity plot of the CNN-based model, illustrating estimation errors for biomass labels underrepresented in the training set. (c) Parity plot of our model, showing improved estimation accuracy and robustness.



(a) MLP-based model [59]

(b) CNN-based model [47]

(c) Ours

Fig. 2. Visualization of link importance for estimating BGB of root tubers. Link color indicates its saliency in biomass estimation, and the red region denotes the tuber location and its maximum cross-section.

unaffected by tubers, resulting in noisy features with limited efficacy and robustness. Although the CNN-based model [47] reduces irrelevant links (Fig. 2b), some links remain unrelated to the tuber, degrading model estimation accuracy.

To address the issues mentioned above, we propose a novel tuber biomass sensing (TBS) framework that combines a scalable wireless network and a DNN model. The framework is suitable for greenhouse-based scenarios, where controlled environmental conditions enable relatively stable wireless sensing for non-invasive BGB estimation of root tubers. First, we build a data acquisition testbed using a low-cost wireless network with multiple sensor nodes [51] to collect RSS data to train BGB estimation models. This network not only provides soil penetration capability, but also achieves higher sensing resolution than GPR and previous wireless networks by increasing the number of frequency channels and sensor nodes. The frequency and spatial diversities of the networked sensing testbed provide richer information about below-ground root tubers for biomass estimation. Using this testbed, we perform extensive experiments in various scenarios, such as different tuber dimensions, weights, and placement positions, to construct a new root tuber biomass sensing dataset. In total, this dataset includes more than 700,000 RSS measurements for sensing model training and evaluation.

Second, we propose a novel DNN model with two modules specifically designed for BGB estimation. Convolution neural networks with residual connections have shown superior performance in various fields, such as remote sensing [16, 17] and wireless sensing [12, 37], by effectively addressing gradient-related issues [8]. Additionally, attention mechanism enhances DNN performance by adaptively emphasizing target-related features while suppressing irrelevant ones, and has been widely incorporated into various deep learning algorithms [18, 67, 68]. Thus, we introduce a DNN model that uses a convolution neural network and residual connections as the backbone for feature extraction, while incorporating two types of attention mechanisms [48, 56] to adjust the channel and spatial characteristics of features. As illustrated in Fig. 2c, our model uses measurements from links that cross or surround the tuber to generate discriminative features, demonstrating the efficacy of the attention mechanisms in capturing discriminative patterns from multi-frequency and multi-sensor RSS data. Moreover, to mitigate performance degradation caused by imbalanced training data, we propose to use contrastive learning to align feature representations of samples with similar biomass labels while increasing the separation between those with significantly different labels. As shown in Fig. 1c, our model effectively reduces estimation errors for samples with underrepresented biomass labels, thereby improving overall estimation accuracy and robustness.

Third, to verify the efficacy of the proposed model, we perform extensive evaluations and compare it with various baseline models. We introduce multiple evaluation metrics to assess performance from different aspects and perform evaluations under diverse scenarios, including various tuber planting positions, tuber weights, and wireless network configurations. The evaluation results show that the proposed model achieves an average mean absolute percentage error (MAPE) of 2.81%, outperforming SOTA baseline models and confirming its efficacy for non-destructive estimation of BGB of root tubers. In addition, we perform ablation studies to validate the contribution of different modules in our DNN model, each of which improves the accuracy of biomass estimation.

In summary, this paper makes the following contributions.

- We propose a novel TBS framework using a low-cost wireless network. This framework takes advantage of both frequency and spatial diversities of the wireless network, and a root tuber BGB sensing dataset is constructed based on it through extensive experiments.
- We propose a novel DNN model for root tuber BGB estimation. The model integrates a convolution neural network with residual connections and attention mechanisms to extract discriminative features from RSS data for accurate biomass estimation. It then uses a contrastive learning method to further enhance the robustness and generalizability of the DNN model.
- We perform extensive real-world experiments and evaluate our DNN model using over 700,000 RSS measurements. The evaluation results demonstrate the efficacy of our model, which achieves superior estimation accuracy compared to SOTA baseline models.

2 Problem Statement and Overview

2.1 Problem Statement

Given a two-dimensional sensing area surrounded by K wireless sensor nodes, a wireless network is formed with $M = K(K-1)$ wireless links, where the nodes operate on C frequency channels. The RSS value measured by the receiving node of link i at time t on frequency channel c can be described as [54]:

$$r_{i,c}[t] = P_c - Z_{i,c} - H_{i,c}[t] + F_{i,c}[t] - G_{i,c}[t],$$

where P_c is the transmit power, $Z_{i,c}$ is the larger scale path loss, $F_{i,c}$ is the fading gain, $G_{i,c}$ is the measurement noise, and $H_{i,c}$ is the shadowing loss caused by objects blocking the signal propagation path. Note that the transmit power P_c is constant for all links operating on the same frequency channel, and the larger scale path loss $Z_{i,c}$ remains unchanged over time. Therefore, we use a single subscript index for P_c and two subscript indices

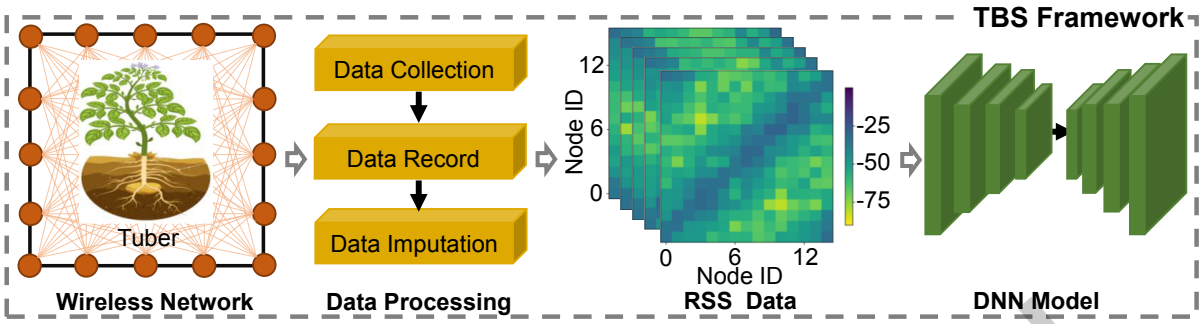


Fig. 3. Overview of the tuber biomass sensing framework. A wireless network monitors the below-ground root tuber, and the collected RSS data are processed and then input into a novel DNN model to estimate root tuber biomass.

for $Z_{i,c}$, respectively. By considering all links and channels, we construct an RSS matrix $\mathbf{R}$, where each column corresponds to a wireless link and each row represents a frequency channel.

To estimate biomass, we propose an end-to-end DNN model $\mathcal{F} : \mathbf{R} \rightarrow y$ that learns the mapping between RSS data $\mathbf{R}$ and root tuber biomass y. Formally, it is defined as:

$$\widetilde{y} = \mathcal{F}(\mathbf{R}; \mathbf{\Theta}),$$

where $\widetilde{y}$ is the estimate of y, and $\mathbf{\Theta}$ represents the set of parameters of the DNN model, which are iteratively optimized using various loss functions, as detailed in Section 3. After optimization, the well-trained model $\mathcal{F}$ is used to estimate BGB of root tubers using RSS data.

2.2 Framework Overview

Fig. 3 presents an overview of our framework, which consists of the following components. First, we build a data acquisition testbed [51] to collect RSS data. The testbed uses a wireless network to capture multi-frequency and multi-sensor RSS measurements from below-ground root tubers. Additionally, the testbed uses a novel sensing toolkit, including "plug-and-play" containers and a rotating platform, to facilitate data collection from below-ground root tubers with varying dimensions, weights, and locations. Based on this testbed, we perform extensive data collection experiments, thereby building a new biomass sensing dataset for training and evaluation. Further details of the testbed, experiments, and dataset are provided in Section 4.

Second, missing data due to wireless interference is a common challenge in wireless sensing. To address this issue, we propose a data processing component to perform data imputation using the most recent packets. Specifically, if a sensor node fails to receive packets from other nodes, we impute the missing RSS values using the latest available data for this node on the same frequency channel. If all packets in a given frequency channel are lost, we impute them with the latest values from the same channel.

Third, we propose a novel DNN model to estimate BGB of root tubers using multi-frequency and multi-sensor RSS data. The model consists of several key modules: a multi-layer convolution neural network with residual connections to extract high-dimensional features from RSS data, a channel attention mechanism to effectively capture channel dependencies, and a spatial attention mechanism to automatically emphasize relevant spatial information in the features. Additionally, a contrastive learning method is proposed to encourage features with similar biomass labels to have similar representations in feature space, while enhancing the distinction between features with significantly different labels. The network architecture of the proposed model is illustrated in Fig. 4, with additional details provided in Section 3.

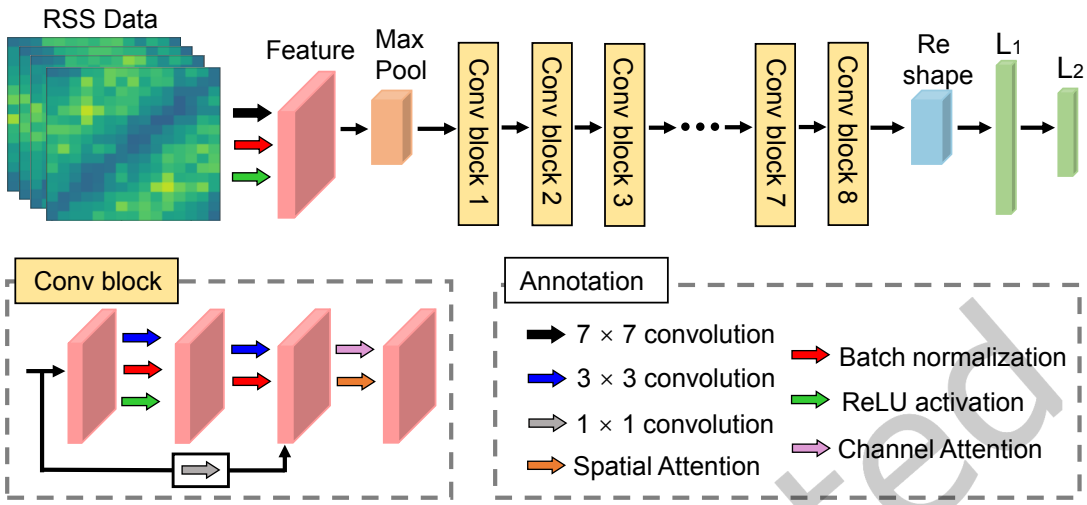


Fig. 4. Architecture of the proposed DNN model. "Conv block" represents the convolution block, while L_1 and L_2 denote two linear layers, respectively.

3 Model

As shown in Fig. 4, we propose a novel DNN model that uses a convolution neural network with residual connections and attention mechanisms for discriminative feature extraction and accurate biomass estimation. Moreover, it incorporates contrastive learning to enhance the generalizability and robustness of biomass estimation. Further details are discussed in the following.

3.1 Convolution Neural Network

We propose to use a convolution neural network with residual connections as the backbone to model relationships between RSS data and BGB of root tubers. The convolution network uses a series of convolution blocks to extract discriminative features from the input data. Specifically, as shown in Fig. 4, the neural network consists of eight convolution blocks, each applying convolution operations to the input to progressively generate high-dimensional features. Each block consists of two convolution layers for feature extraction, a BatchNorm layer to mitigate covariate shift and facilitate model convergence, and a ReLU activation layer to introduce nonlinearity. To mitigate gradient explosion or vanishing during network training and facilitate the learning of complex patterns, we incorporate residual connections [25] in each convolution block. The initial input of each block is passed through two convolution layers, and the resulting output is added to that of another convolution layer with 1×1 kernel size. Residual connections enhance training efficiency and allow for feature reuse, leading to more robust and comprehensive feature extraction. In addition, each convolution block includes a channel attention mechanism and a spatial attention mechanism, detailed in Section 3.2.

As shown in Fig. 4, before being fed into the convolution blocks, the input is first passed through a single convolution layer, followed by BatchNorm and ReLU layers for initial feature extraction. A max pooling operation is then used to reduce the feature map size, thereby decreasing the computational burden on subsequent convolution blocks. Finally, the feature map from the last convolution block is reshaped into a one-dimensional vector and is passed through two linear layers for root tuber BGB estimation.

3.2 Attention Mechanism

To enhance biomass estimation performance, we propose to use two attention mechanisms: the channel attention mechanism [48] and the spatial attention mechanism [56], both of which are integrated into the convolution block. Specifically, for a given feature map $\mathbf{v} \in \mathbb{R}^{D\times E\times Q}$, where D, E, and Q denote the number of channels, height, and width, respectively, the channel attention mechanism first aggregates spatial information across the width and height dimensions of the feature map using an average pooling operation. The pooled output is then fed into a one-dimensional convolution layer to generate a channel-wise attention vector $\mathbf{a}_{ch}$, defined as:

$$\mathbf{a}_{ch} = Soft(Conv_1(AvgPool(\mathbf{v}))),$$

where $Conv_1$ denotes the one-dimensional convolution layer and $AvgPool$ represents the average pooling operation. $Soft$ represents the softmax function, which normalizes the attention vector to produce attention weights. The channel attention weight vector is then used to multiply the input feature map $\mathbf{v}$ to generate a channel-weighted feature map: $\mathbf{u} = \mathbf{a}_{ch} \cdot \mathbf{v}$, allowing the network to focus on the most important channels.

Subsequently, the spatial attention mechanism takes $\mathbf{u}$ as input, and applies average pooling and max pooling operations to respectively aggregate the feature map along the channel dimension. The outputs from two pooling operations are concatenated along the channel axis to form a novel feature map, which is fed into a convolution layer followed by a softmax function to generate the spatial attention map $\mathbf{a}_{sp}$, defined as:

$$\mathbf{a}_{sp} = Soft(Conv([MaxPool(\mathbf{u}), AvgPool(\mathbf{u})])),$$

where $Conv$ and $MaxPool$ represent the convolution layer and the max pooling operation, respectively. The attention map is then used to weight $\mathbf{u}$ through multiplication to generate a spatial-weighted feature map: $\mathbf{f} = \mathbf{a}_{sp} \cdot \mathbf{u}$, enhancing the relevant spatial features while suppressing less informative ones. Finally, $\mathbf{f}$ is fed into the next convolution block for feature learning.

3.3 Contrastive Learning

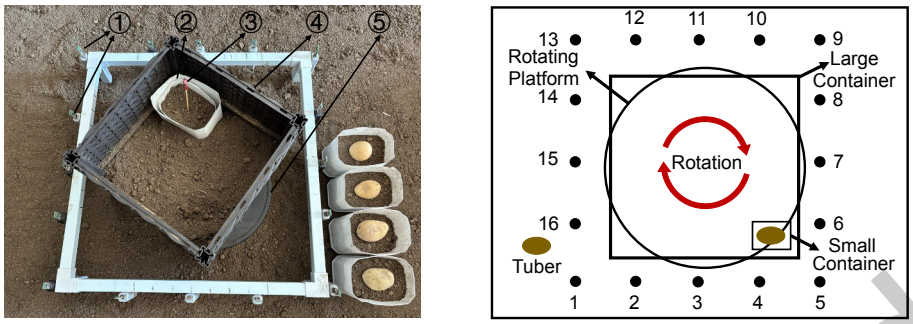
Contrastive learning has shown efficacy in mitigating model performance degradation caused by imbalanced training data [29]. In this study, we propose a contrastive learning method to reduce the estimation bias of root tuber BGB. This method is implemented through a loss function [62]. It encourages features with similar biomass labels to have similar representations in the feature space, while enhancing the distinction between features with significantly different labels. Specifically, the features extracted by the proposed convolution neural network are passed through two linear layers for biomass estimation, while the outputs of the first linear layer are used to compute the contrastive loss. Given an anchor feature $\mathbf{f}_n$ and another feature $\mathbf{f}_b$ within the same batch, we select the third feature $\mathbf{f}_o$ from the same batch and define the following set of features [62]:

$$\mathbf{S}_{n,b} = \{\mathbf{f}_o | d(y_n, y_o) > d(y_n, y_b)\},$$

where y_n, y_o and y_b denote the biomass labels of $\mathbf{f}_n$, $\mathbf{f}_o$ and $\mathbf{f}_b$, respectively, and $d(\cdot,\cdot)$ represents the distance measure between two biomass labels. $\mathbf{S}_{n,b}$ denotes a set of features whose label distances to y_n exceed the distance between y_n and y_b, indicating that $\mathbf{f}_n$ should be more similar to $\mathbf{f}_b$ while maintaining a greater distinction from each $\mathbf{f}_o \in \mathbf{S}_{n,b}$. To this end, we use a contrastive loss defined as [62]:

$$l_{n,b} = -log\frac{exp(sim(\mathbf{f}_n, \mathbf{f}_b)/\tau}{\sum_{\mathbf{f}_o\in\mathbf{S}_{n,b}} exp(sim(\mathbf{f}_n, \mathbf{f}_o))/\tau},$$

where $sim(\cdot,\cdot)$ denotes the similarity measure between two features, and τ represents the temperature parameter. Minimizing $l_{n,b}$ encourages increasing the similarity between $\mathbf{f}_n$ and $\mathbf{f}_b$ or decreasing the similarity between $\mathbf{f}_n$ and $\mathbf{f}_o$, both of which ensure that $\mathbf{f}_n$ is closer to $\mathbf{f}_b$ than to the features in $\mathbf{S}_{n,b}$.



(a) Testbed (b) Experimental layout

Fig. 5. Our testbed comprises a wireless network with 16 sensor nodes and a through-soil sensing toolkit featuring "plug-and-play" functionality and data augmentation capability. ① TI CC2531 nodes, ② Small container with the tuber, ③ Marker indicating the tuber position, ④ Larger container, ⑤ Rotating platform.

Furthermore, we compute the average contrastive loss for $\mathbf{f}_n$ using all features within the same batch, defined as $l_n = \frac{1}{N}\sum_{b=1}^{N} l_{n,b}$, where N represents the batch size. Intuitively, for an anchor feature $\mathbf{f}_n$, any other $\mathbf{f}_b$ in the same batch is contrasted with it, enforcing the similarity between $\mathbf{f}_n$ and $\mathbf{f}_b$ to be greater than that between $\mathbf{f}_n$ and $\mathbf{f}_o$, when the label distance between y_n and y_o is greater than that between y_n and y_b. Using each feature in the current batch as an anchor, we calculate the overall contrastive loss L_{con}, defined as:

$$L_{con} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{N}\sum_{b=1}^{N} l_{n,b}.$$

Additionally, these features are fed into the second linear layer of our model to estimate the below-ground tuber biomass, with the estimation loss L_{bio} defined as follows:

$$L_{bio} = \frac{1}{N}\sum_{n=1}^{N} ||y_n - \widetilde{y}_n||^2,$$

where $\widetilde{y}_n$ denotes the estimated biomass value for the n-th sample in the batch. Finally, the contrastive loss and the estimation loss are summed to obtain the total loss, $L = L_{con} + L_{bio}$, which is used to guide the iterative learning process of the proposed DNN model.

4 Experiments and Dataset

In this paper, we select potato tubers as a representative type of root tuber. We construct a testbed based on a low-cost wireless network to collect RSS data to estimate BGB of potted potato plants grown in a greenhouse environment. Using the collected data, we build a potato tuber biomass sensing dataset to train and evaluate DNN models. Further details are discussed in the following.

4.1 Data Acquisition Testbed

In this paper, we use the testbed proposed in [51], which consists of a ZigBee wireless network and a through-soil sensing toolkit, as shown in Fig. 5.

4.1.1 Wireless Network. Fig. 5 illustrates the wireless network used for data collection, which consists of 16 TI CC2531 sensor nodes deployed on a white rack. Each sensor node is programmed with a multi-channel

Table 1. Comparison of hardware cost, energy consumption, and sample size among the ZigBee wireless network, the GPR device [41], and the CT scanner [38].

	Cost (USD)	Power (W)	Battery	Mobility	Sample Size (KB)
GPR device [41]	39,950	13.3	Yes	Yes	477.27
CT scanner [38]	150, 000	80×10^3	No	No	512.00
Wireless network	10	2.82×10^{-3}	Yes	Yes	15.00

time-division multiple access (TDMA) communication protocol and operates on one of the 16 frequency channels in the 2.4 GHz frequency band at any given time [54]. In addition, a designated sink node receives all packets transmitted by sensor nodes and is connected to a laptop, where RSS data are stored and processed.

To further demonstrate the advantages of the wireless network, we compare it with a GPR system [41] and a computed tomography (CT) system [38], both of which are used for sensing below-ground roots. As shown in Table 1, first, the wireless network used in our framework is low-cost, with each sensor node costing about USD 10. This cost advantage enables the deployment of additional sensor nodes with minimal budget increase, facilitating large-scale data collection and enhancing adaptability to diverse sensing conditions. Second, in terms of energy consumption, each sensor node consumes only 2.82 mW, which is substantially lower than 13.3 W and 80 kW required by the GPR and CT systems, respectively. The low-power design allows for battery operation, enhancing mobility and extending runtime compared with the GPR system under equivalent battery conditions. Third, in the wireless network with 16 nodes, each RSS sample used for training and evaluation is approximately 15 KB, which is significantly smaller than those of the GPR and CT systems. The smaller sample size accelerates data transmission and reduces storage requirements. These results demonstrate the superiority of our framework and highlight its practicality for real-world deployment. Note that we have tested using a subset of the 16 sensor nodes in our evaluations, as discussed in Section 5.5.

4.1.2 Through-soil sensing toolkit. Collecting large amounts of wireless sensing data is labor-intensive, particularly for root tubers buried in soil. To address this challenge, we employ a set of tools, including a rotating platform and "plug-and-play" containers, to facilitate data collection. First, we use two types of containers to minimize frequent soil digging and tuber burying. As shown in Fig. 5, the larger container holds soil with predefined locations for inserting one or two smaller containers. Potato tubers of different dimensions are placed in the smaller containers with soil. For example, in our experiment, 26 potato tubers are placed in small containers, allowing for easy replacement of the containers during data collection. Second, by placing the "nesting containers" on a rotating platform and rotating it through various predefined angles, potato tubers can be located at different positions and orientations within the sensing area. The idea of using a rotating platform is inspired by the data augmentation method [19], which is commonly employed to generate large amounts of data after the acquisition stage. More details can be found in [51].

4.2 Data Collection Campaign

To estimate the biomass of below-ground tubers, we perform a data collection campaign using multiple potato tubers with varying dimensions and weights. Specifically, sensor nodes are deployed on a 72 cm × 72 cm rack, creating a sensing area for tubers buried in the soil. RSS data are collected from 26 potato tubers buried in a 40 × 40 × 40 plastic container, with tuber depths ranging from 11 cm to 13.5 cm. Additional characteristics of the tubers are provided in Table 2. During data collection, the platform is rotated 32 times for each tuber, positioning it at different angles and locations within the sensing area. In total, we collect RSS data for 832 unique tuber-position pairs, each corresponding to 40 seconds of RSS measurements.

Table 2. The characteristics of potato tubers and the corresponding number of RSS measurements collected for these tubers.

Weight Distribution	Number	Weight Range(g)	Length × Width × Thickness(cm)	RSS
< 100 g	9	66.4 ~ 98.1	L:5.3-7.4 W:4.5-5.2 T:4.0-5.0	246,848
100 g ~ 150 g	8	103.2 ~ 146.1	L:5.9-7.4 W:5.4-6.4 T:5.0-5.6	216,928
> 150 g	9	156.7 ~ 209.8	L:7.8-9.5 W:6.1-6.5 T:4.7-6.4	246,768
Total	26	66.4 ~ 209.8	L:5.3-9.5 W:4.5-6.5 T:4.0-6.4	710,544

Table 3. Parameters used in our DNN model.

Parameter description	Default Value
Learning rate of the deep neural network model.	$5e^{-4}$
Input dimension of a training or testing sample.	$16 \times 16 \times 15$
Input and output dimensions of the first linear layer.	512, 64
Input and output dimensions of the second linear layer.	64, 1
Kernel size of the spatial attention in convolution blocks.	7×7
Kernel size of the channel attention in convolution blocks.	5
Number of output channels of the first and second blocks.	64, 64
Number of output channels of the third and fourth blocks.	64, 128
Number of output channels of the fifth and sixth blocks.	128, 256
Number of output channels of the seventh and eighth blocks.	256, 512

4.3 Dataset

In this study, the dataset contains RSS data from 26 potato tubers with varying dimensions and weights, ensuring tuber diversity. Specifically, nine (34.62%) tubers have weights less than 100 g, eight (30.76%) tubers have weights between 100 g and 150 g, and nine (34.62%) tubers have weights exceeding 150 g, covering different growth stages of tubers. The maximum and minimum weights of the tubers are 209.8 g and 66.4 g, respectively. Furthermore, since tuber positions are not fixed during growth, we use a rotating platform to augment position variations during data collection. In total, we collect RSS data from 832 tuber-position pairs, generating 710,544 RSS measurements for below-ground tuber biomass estimation. Note that our dataset is publicly available on https://zenodo.org/records/15000852.

5 Evaluation and Discussion

We perform extensive evaluations and use various metrics to assess the performance of our DNN model for root tuber BGB estimation. In addition, we compare its performance with that of other SOTA baseline models. Further details are provided in the following sections.

5.1 Metrics and Algorithm Parameters

5.1.1 Evaluation Metrics. We evaluate the performance of our model using four metrics: MAPE [28], mean absolute error (MAE) [53], root mean squared error (RMSE) [11], and coefficient of determination (R^2) [63]. First, MAE quantifies the average absolute difference between estimated and label values to evaluate the performance of regression models, with lower MAE indicating better estimation accuracy. Second, RMSE measures the square

Table 4. Performance of our model and baseline models on randomly positioned tubers. We mark the best and second-best results using bold and underlined text, respectively.

Method	Evaluation Metrics			
	MAE (g) ↓	RMSE (g) ↓	R^2 ↑	MAPE (%) ↓
MLP [59]	9.09	14.34	0.89	7.36
ResNet [32]	<u>5.75</u>	<u>11.94</u>	<u>0.92</u>	<u>4.66</u>
Transformer [23]	6.98	13.86	0.90	5.67
CNN-GRU [47]	12.37	18.29	0.82	10.02
CNN-Transformer [18]	10.79	15.91	0.87	8.76
Ours	**3.47**	**10.22**	**0.94**	**2.81**

root of the mean squared difference between estimated and label values, commonly used in regression evaluation due to its sensitivity to outliers [10]. A smaller RMSE value indicates better performance, with a value of zero representing perfect estimation. Third, MAPE expresses MAE as a proportion of the label value, providing a more intuitive measure of model performance. Lower MAPE values suggest higher estimation accuracy. Fourth, R^2 is computed as one minus the ratio of the sum of squared differences between the estimated and label values to the total variance of the label values. This metric assesses the goodness of fit of a regression model and ranges from $-\infty$ to 1, with 1 indicating perfect estimation.

5.1.2 Model Parameters. After data imputation, the sequential RSS data is divided into individual samples for training and testing, each with dimensions of $16 \times 16 \times 15$, covering data from 16 frequency channels and 16 sensor nodes. Then, we propose a novel DNN model for extracting discriminative features to accurately estimate BGB of root tubers. Specifically, the model consists of eight convolution blocks, each containing two convolution layers with a 3×3 kernel size and 1×1 padding, along with a channel attention mechanism and a spatial attention mechanism. In the first, third, fifth, and seventh blocks, the first convolution layer and the residual convolution layer use a stride of 2 to reduce the feature map dimensions, while the second convolution layer uses a stride of 1 to maintain consistent dimensions. The channel attention mechanism is implemented using a one-dimensional convolution layer with a kernel size of 5, a stride of 1, and a padding of 2. The spatial attention mechanism is implemented with a convolution layer that uses a 7×7 kernel, a 1×1 stride, and a 1×1 padding. Additionally, we use a convolution layer with a kernel size of 7 and a stride of 2 at the beginning of the model, and set the temperature parameter to 0.5 for computing the contrastive loss. Further details of the model parameters are provided in Table 3.

5.2 Baselines

For comprehensive comparisons, we choose five SOTA DNN models as baselines, which can be categorized into two groups: single-architecture models and hybrid-architecture models. First, for single-architecture models, the MLP model used in [59] demonstrates efficacy and ease of implementation in estimating AGB of potato plants. The ResNet model [32] is an end-to-end model consisting of multiple convolution layers and residual connections for plant biomass estimation. The transformer model used in [23] employs a self-attention mechanism to improve estimation performance and demonstrates efficacy in accurately estimating wheat yield. Second, for hybrid-architecture models, the CNN-GRU [47] model combines a convolution neural network with gated recurrent unit (GRU) [13] for feature extraction, followed by densely connected layers for estimation. The CNN-Transformer

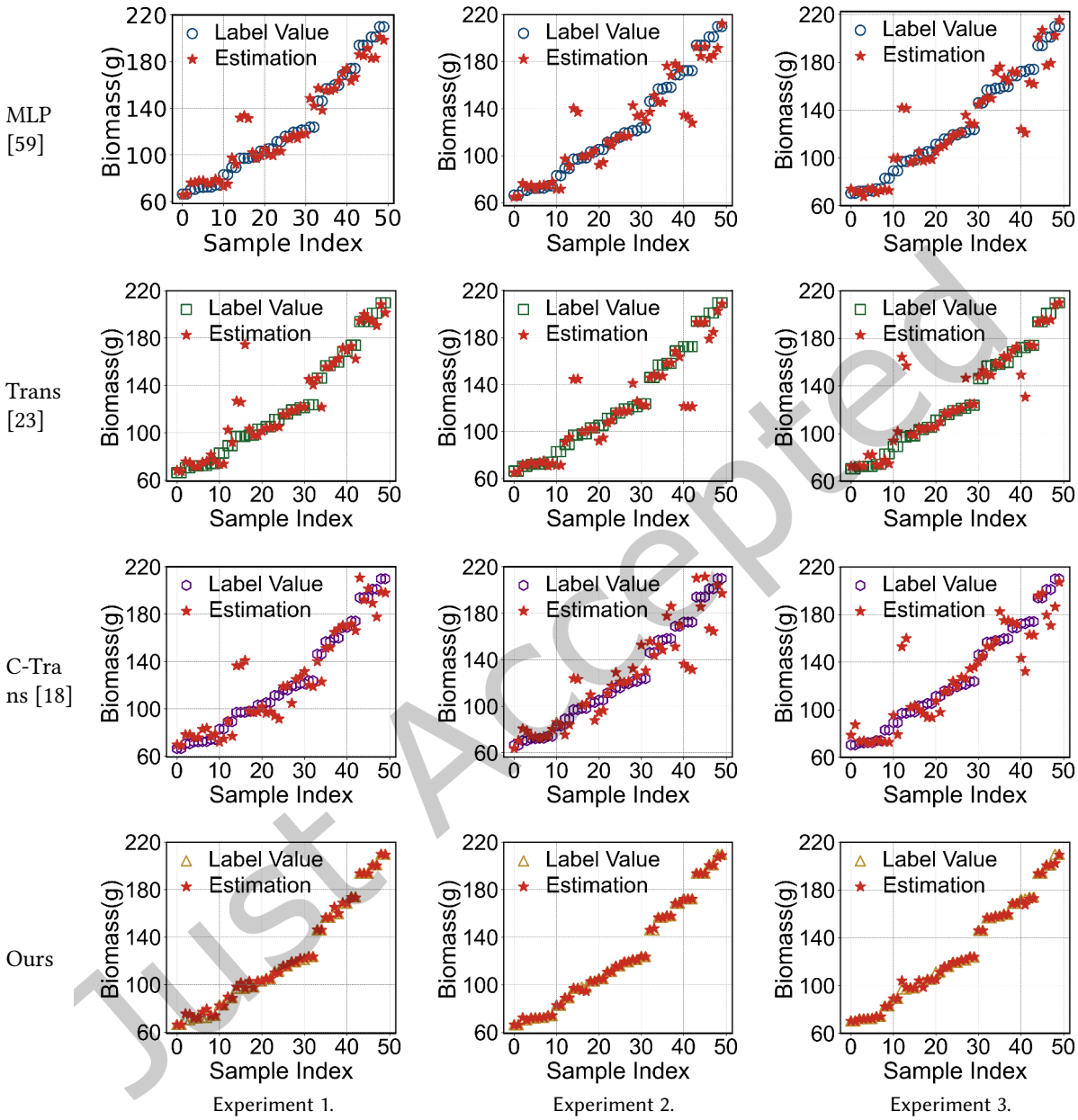


Fig. 6. Visualization of estimation accuracy for various models. "Trans" and "C-Trans" represent the transformer model [23] and the CNN-Transformer model [18], respectively.

model [18] first employs a convolution neural network to extract local features from the input data, which are then fed into a transformer network for global feature extraction and estimation.

5.3 Evaluation on Random Positions

First, we evaluate the average biomass estimation performance of our model at randomly positioned tubers. Specifically, RSS data are collected from 26 tubers placed at 32 positions, and the tuber-position pairs are randomly split in a 9:1 ratio for training and testing. The data splitting process is repeated three times, and four evaluation metrics are used to assess the performance of our model and baselines. As shown in Table 4, we present the average results of four evaluation metrics for all models. Our model achieves an average MAE of 3.47, outperforming all compared models, the best of which yields a value of 5.75. Furthermore, our model achieves an average RMSE value of 10.22, showing improvements of 14.41% and 26.27% over ResNet [32] and Transformer [23], respectively, which are the second and third best models. In addition, our model achieves average R^2 and MAPE values of 0.94 and 2.81%, respectively, outperforming the baseline models. These results confirm the efficacy of our model in comparison with the baseline models. Specifically, among single-architecture models, the MLP model, composed of simple linear layers, is insufficient to extract discriminative features from RSS data. The ResNet model, which uses only convolution layers with residual connections, fails to capture the multi-frequency and multi-sensor characteristics of RSS data, resulting in suboptimal biomass estimation performance. The transformer model employs a self-attention mechanism to focus on global feature extraction, but has limited capacity to capture fine-grained local features and requires large amounts of training data to perform effectively [57]. Furthermore, hybrid-architecture models, which combine different neural networks, increase the complexity of the DNN models, raising the risk of overfitting on limited-size datasets and degrading model performance. In contrast, our model integrates a convolution neural network with residual connections and two attention mechanisms, which not only prevent gradient explosion and vanishing but also facilitate discriminative feature extraction from RSS data. In addition, we introduce a contrastive learning method that aligns features with similar biomass labels, further enhancing the generalizability and performance of the DNN model [62].

Second, to enable visual comparison, we plot the estimated values and corresponding biomass labels to illustrate the performance of different models. For each evaluation experiment, 50 samples are randomly selected and visualized. As shown in Fig. 6, our model achieves more accurate estimations compared to the baseline models. For example, in the first column of Fig. 6, all models achieve relatively accurate estimation in Experiment 1. However, the baseline models exhibit noticeable variability in some samples, whereas our model produces more stable estimates for these cases. The second and third columns present the results for Experiments 2 and 3, respectively, with our model consistently outperforming baseline models in estimation accuracy. These results further demonstrate the superiority of our model and indicate that the proposed DNN architecture and contrastive learning strategy enhance the stability of BGB estimation for potato tubers.

Third, to illustrate the error distributions of different models, the MAPE values of all test tubers are grouped into six bins. Fig. 7 presents histograms of these values across three evaluation experiments for each model. As shown in the first column of Fig. 7, our model achieves an average MAPE value of 4.03%, outperforming the baseline models, which achieve 9.63%, 9.06%, and 5.39%, respectively. In the second column, the maximum MAPE value of our model remains below 20%, whereas those of the baseline models exceed 30%. In the third column, most MAPE values from our model are below 8%, while those from the baseline models are below 24%, 26%, and 18%, respectively. These results further highlight the superior performance of our model. On the one hand, the combination of convolution layers and attention mechanisms enables automatic extraction of discriminative features, thereby improving estimation accuracy. On the other hand, the contrastive learning approach enhances the robustness and generalizability of our model by reducing estimation bias.

5.4 Evaluation on Random Tubers

In this section, we perform leave-k-out evaluations using RSS data from 26 tubers to assess the performance of our model on randomly selected test tubers, where “k” denotes the number of test tubers and is set to 1 and 2.

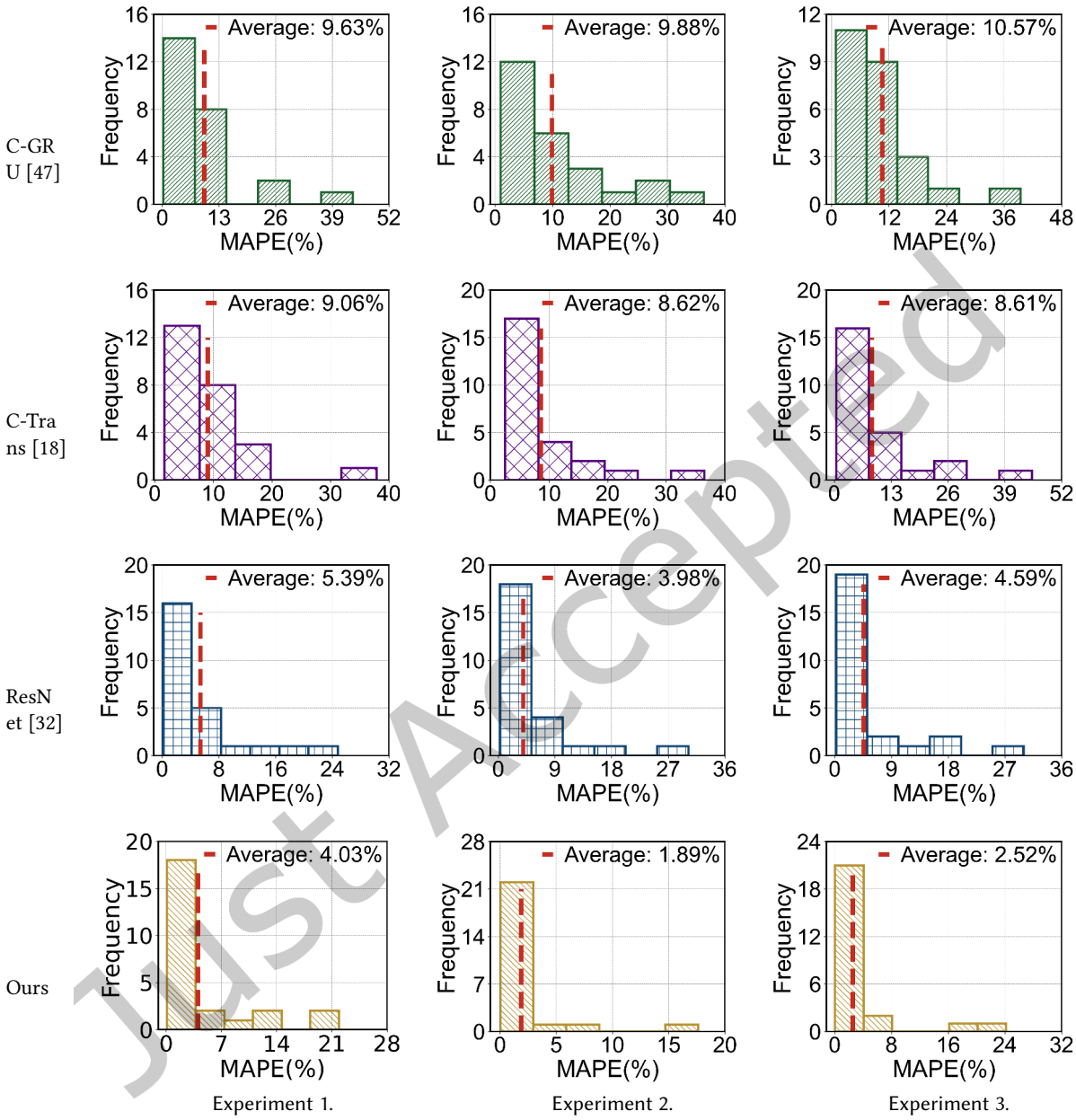
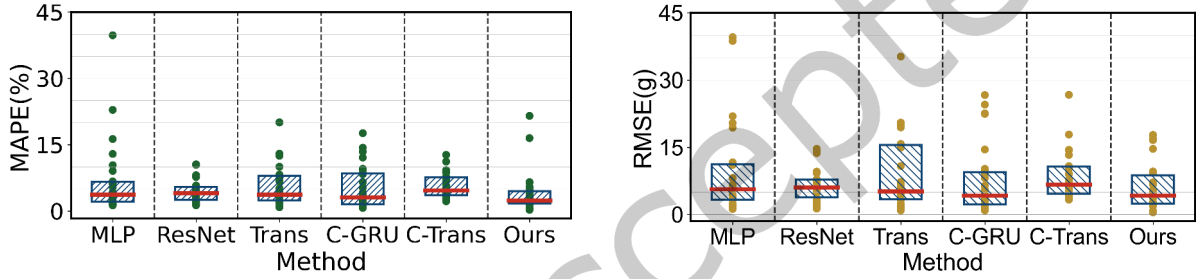


Fig. 7. Histogram of MAPE values for different models across various experiments. "C-GRU" and "C-Trans" denote the CNN-GRU [47] and CNN-Transformer [18] models, respectively.

Specifically, we repeat the leave-1-out experiments 26 times to cover all tubers in the dataset, and repeat the leave-2-out experiments 50 times to obtain statistical results and average performance. As shown in Table 5, our

Table 5. Leave-k-out performance of different models, where "k" represents the number of test tubers. The best and second-best results are indicated in bold and underlined text, respectively.

Method \ Case	Leave-1-out				Leave-2-out			
	MAE (g) ↓	RMSE (g) ↓	R^2 ↑	MAPE (%) ↓	MAE (g) ↓	RMSE (g) ↓	R^2 ↑	MAPE (%) ↓
MLP [59]	8.79	9.54	0.89	6.88	8.64	10.42	0.86	6.81
ResNet [32]	<u>5.41</u>	<u>6.69</u>	**0.97**	<u>4.53</u>	8.38	10.60	0.85	6.60
Transformer [23]	8.29	9.13	0.92	6.24	<u>8.07</u>	<u>9.89</u>	<u>0.87</u>	<u>6.40</u>
CNN-GRU [47]	6.93	7.63	<u>0.94</u>	5.33	9.90	12.25	0.80	7.87
CNN-Transformer [18]	7.31	9.00	<u>0.94</u>	5.90	8.79	10.81	0.83	7.15
Ours	**4.65**	**6.12**	**0.97**	**4.08**	**5.99**	**7.75**	**0.90**	**4.77**

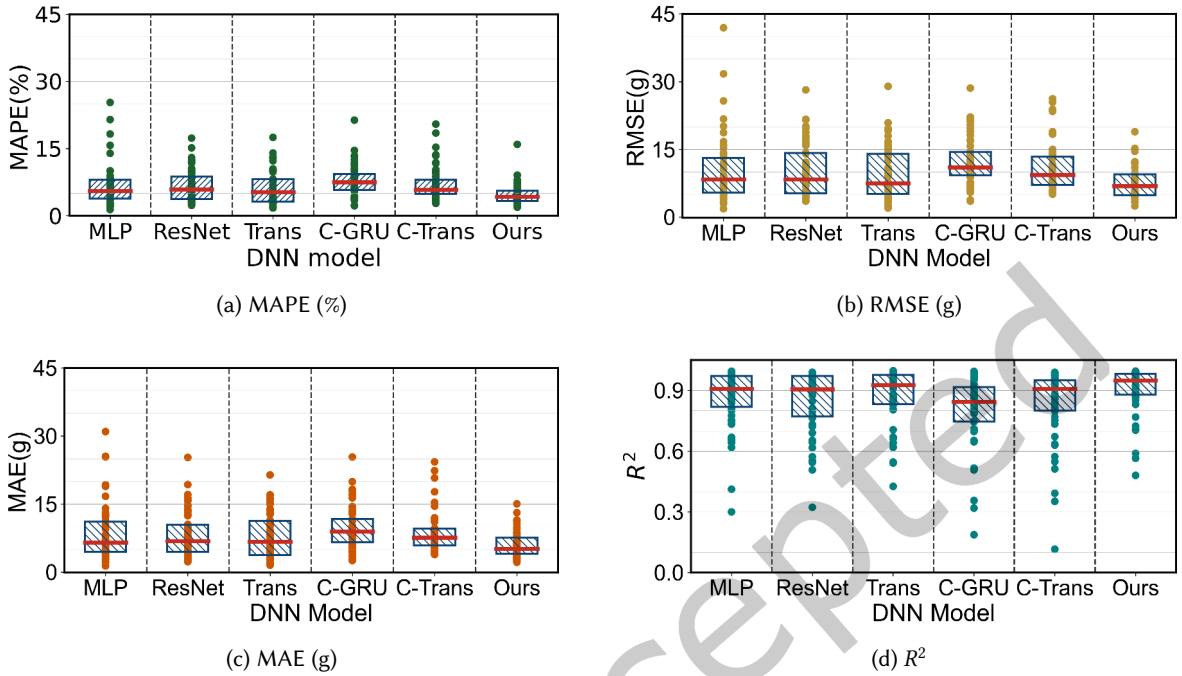


(a) MAPE (%) (b) RMSE (g)

Fig. 8. MAPE and RMSE values for different tubers obtained from various models in leave-1-out experiments. Points represent the MAPE or RMSE values for individual tubers, and the boxes represent the lower quartiles, medians, and upper quartiles of these values. "Trans", "C-GRU", and "C-Trans" represent Transformer [23], CNN-GRU [47], and CNN-Transformer [18], respectively.

model consistently achieves average MAE, RMSE, and MAPE values below 6.0, 8.0, and 5.0%, respectively, across different configurations, outperforming all baseline models. Moreover, our model achieves superior R^2 values in each configuration compared to the baseline models, reaching 0.97 and 0.90, respectively. Additionally, while the performance of all models decreases as the number of test tubers increases, our model consistently demonstrates the best performance in each configuration. These results not only demonstrate the efficacy of our model in estimating below-ground tuber biomass but also highlight its robustness and generalizability under data-limited conditions, further confirming the superiority of both the network design and the contrastive learning method.

To provide an intuitive visualization for the error distributions, Fig. 8 presents scatter plots of MAPE and RMSE values for individual tubers obtained from different models in leave-1-out experiments, with boxes indicating lower quartiles, medians, and upper quartiles. As shown in Fig. 8a, the median and upper quartile of the MAPE values of our model are lower than those of the baseline models, indicating that our model provides more accurate estimates for most tubers. As shown in Fig. 8b, the RMSE distribution of our model is more compact compared to MLP [59], Transformer [23], CNN-GRU [47], and CNN-Transformer [18], indicating that our model generalizes better across most tubers than the baseline models. Furthermore, although the RMSE distribution of ResNet [32]



(a) MAPE (%)

(b) RMSE (g)

(c) MAE (g)

(d) R^2

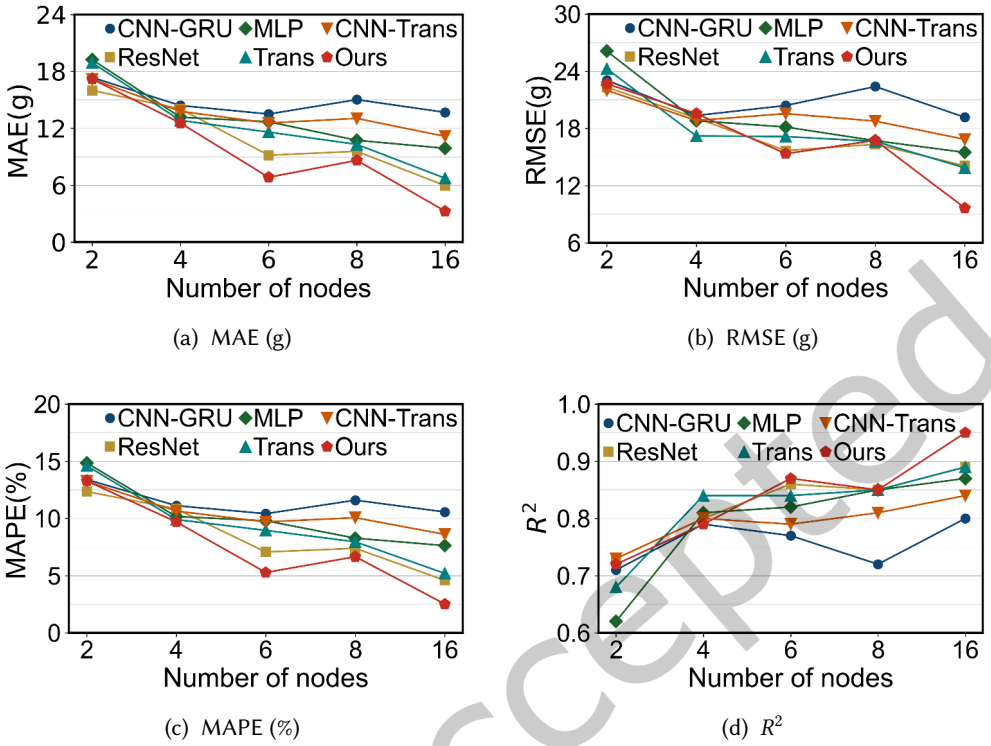
Fig. 9. MAPE, RMSE, MAE, and R^2 values obtained from various models in leave-2-out experiments. Points represent the MAPE, RMSE, MAE, or R^2 values for each evaluation, and the boxes represent the lower quartiles, medians, and upper quartiles of these values. "Trans", "C-GRU", and "C-Trans" represent Transformer [23], CNN-GRU [47], and CNN-Transformer [18] models, respectively.

is similar to that of our model, our model achieves a lower median, indicating more accurate estimates for at least half of the tubers. This further confirms the efficacy of our model in accurately estimating BGB of root tubers.

Furthermore, we present scatter plots of MAPE, RMSE, MAE, and R^2 values for each evaluation in leave-2-out experiments, as shown in Fig. 9. As shown in Fig. 9a and 9c, the lower quartiles, medians, and upper quartiles of MAPE and MAE values of our model are notably lower than those of baseline models, demonstrating the efficacy of our model for non-destructively estimating BGB of root tubers. As shown in Fig. 9b, the RMSE values of our model are not only more compactly distributed but also exhibit lower values compared to those of baseline models, further verifying the efficacy of our model. In addition, Fig. 9d demonstrates the R^2 values from different models, clearly showing that our model outperforms the baseline models. These results highlight the accuracy, stability, and generalizability of our model in estimating BGB of root tubers, especially for data-limited conditions.

5.5 Evaluation on Random Sensor Nodes

In this section, we evaluate the performance of our model under scenarios with varying numbers of sensor nodes. Specifically, we split the tuber-position pairs in the dataset using a 9:1 ratio and assess performance under five configurations: 2, 4, 6, 8, and 16 sensor nodes. Fig. 10 presents the performance of different models across these sensor configurations. First, all models show improved estimation accuracy as the number of sensor nodes increases. For example, the average MAPE value of our model is 13.28% with two sensors, 6.66% with eight sensors,



(a) MAE (g)

(b) RMSE (g)

(c) MAPE (%)

(d) R^2

Fig. 10. MAE, RMSE, MAPE, and R^2 performance of using different numbers of sensor nodes. "Trans" and "CNN-Trans" represent Transformer [23] and CNN-Transformer [18] models, respectively.

and 2.52% with sixteen sensors. Second, our model outperforms the baseline models when the number of sensors is limited. For example, when using 8 sensors for estimation, the MAE value of our model is 8.62 g, whereas the MAE value of ResNet [32] is 9.57 g, which is the best result among all baseline models. Third, as the number of sensors increases, the performance gap between our model and the baselines becomes larger. For example, as shown in Fig. 10d, the difference in R^2 values between our model and CNN-GRU [47] is zero with four sensors, increases to 0.13 with eight sensors, and reaches 0.15 with sixteen sensors. These results demonstrate that our model provides more accurate estimations than the baseline models, particularly in sensor-limited scenarios. The combination of convolution layers and attention mechanisms enables the model to learn effective features from limited wireless signals for BGB estimation. Furthermore, contrastive learning reduces estimation bias and enhances estimation robustness and accuracy.

5.6 Evaluation on Computational Cost

To assess the feasibility of the proposed model for real-time, low-power deployment, we evaluate its model size, floating-point operations (FLOPs), inference time, and energy consumption on three computing platforms: an NVIDIA RTX 3090 GPU, an Intel 10875H CPU, and an NVIDIA Jetson Nano edge device. Specifically, a wireless network sample comprising RSS measurements is used for evaluation, and the results are averaged over three repeated runs. As shown in Table 6, inference on RTX 3090 GPU is the fastest, requiring 0.11 seconds per sample

Table 6. Computational cost of the proposed model on different computing platforms, evaluated on a single wireless network sample and averaged over three runs.

	Parameters (MB)	FLOPs (M)	Power (W)	Energy (J)	Inference Time (S)
3090 GPU	42.92	18.27	73.12	8.09	0.11
10875H CPU	42.92	18.27	45.28	5.45	0.12
Jetson Nano	42.92	18.27	1.74	0.75	0.43

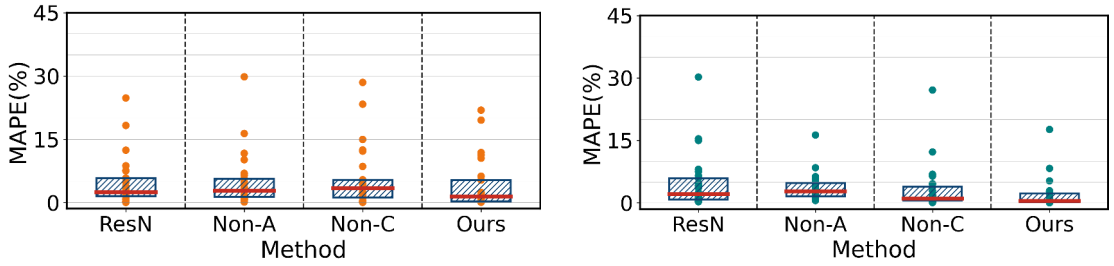
Table 7. Performance of different models in ablation experiments. "Non-A" and "Non-C" refer to the residual network without attention mechanisms and contrastive learning, respectively. The best and second-best results are highlighted in bold and underlined text, respectively.

Case / Method	Experiment 1				Experiment 2			
	MAE (g) ↓	RMSE (g) ↓	R^2 ↑	MAPE (%) ↓	MAE (g) ↓	RMSE (g) ↓	R^2 ↑	MAPE (%) ↓
MLP [59]	9.67	15.68	0.87	7.87	7.71	11.85	0.93	6.57
ResNet [32]	6.62	**11.96**	**0.92**	5.39	4.67	9.76	<u>0.95</u>	3.98
Transformer [23]	8.13	15.30	0.88	6.62	6.08	12.42	0.92	5.18
CNN-GRU [47]	11.82	18.54	0.82	9.63	11.59	17.14	0.85	9.88
CNN-Transformer [18]	11.12	16.15	0.86	9.06	10.11	14.73	0.89	8.62
Non-A	<u>6.02</u>	13.69	<u>0.90</u>	<u>4.91</u>	4.49	<u>9.37</u>	**0.96**	3.82
Non-C	6.74	12.53	**0.92**	5.49	<u>3.67</u>	10.55	0.94	<u>3.13</u>
Ours	**4.95**	<u>12.29</u>	**0.92**	**4.03**	**2.21**	**8.72**	**0.96**	**1.89**

and demonstrating the feasibility for real-time deployment. Inference on Jetson Nano is the slowest, taking 0.43 seconds per sample. However, Jetson Nano exhibits significantly lower power and energy consumption than RTX 3090 GPU, demonstrating the feasibility for low-power deployment. In practice, although the inference time on Jetson Nano exceeds 0.1 seconds per sample, it remains significantly faster than the growth rate of potato tubers, confirming the practicality of our model for real-world deployment. In addition, the inference time and energy consumption on Intel 10875H CPU fall between those of RTX 3090 GPU and Jetson Nano, highlighting the versatility of the proposed model across computing platforms.

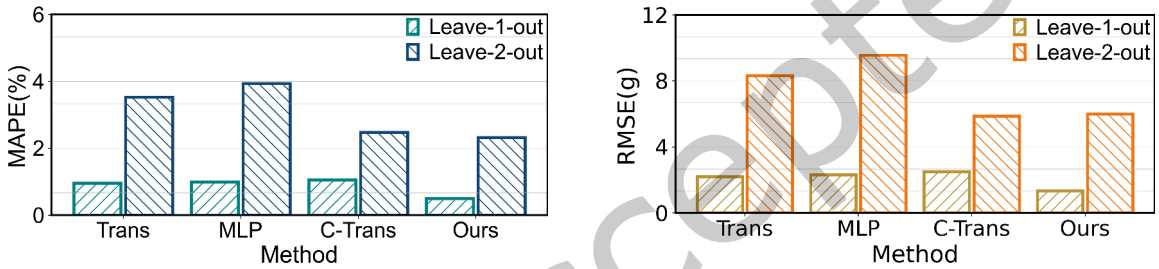
5.7 Ablation Study

In this section, we perform ablation studies to assess the contributions of attention mechanisms and contrastive learning to our model. Specifically, we split the tuber-position pairs of the dataset in a 9:1 ratio and perform evaluations twice. For simplification, we define "Non-A" and "Non-C" as the residual networks without attention mechanisms and contrastive learning, respectively. As shown in Table 7, Non-A and Non-C achieve average MAE values of 6.02 and 6.74 in Experiment 1, and 4.49 and 3.67 in Experiment 2, respectively. These values are higher than those obtained by the complete model. Furthermore, Non-A and Non-C achieve average RMSE values of 11.53 and 11.54 in the two experiments, respectively, which are higher than those of the complete model. These results confirm the contributions of attention mechanisms and contrastive learning in improving estimation accuracy and model robustness.



(a) Experiment 1. (b) Experiment 2.

Fig. 11. MAPE performance of different models in the ablation studies. "Non-A" and "Non-C" refer to the residual networks without attention layers and contrastive learning, respectively. "ResN" represents the ResNet [32] model.



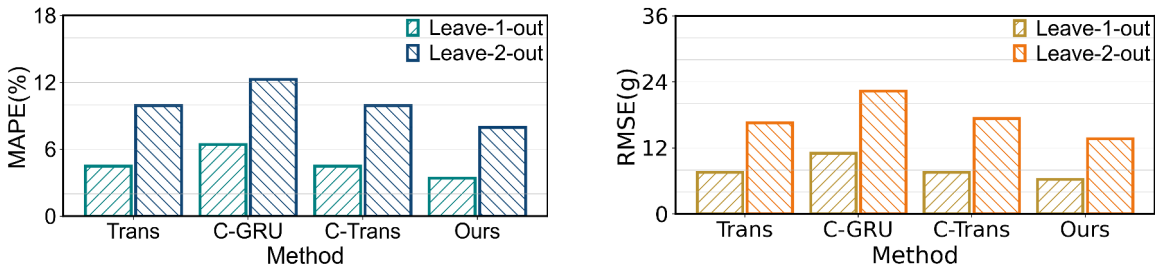
(a) MAPE (%). (b) RMSE (g).

Fig. 12. MAPE and RMSE of different models under leave-1-out and leave-2-out experiments in a multi-tuber scenario. "Trans" and "C-Trans" denote Transformer [23] and CNN-Transformer [18], respectively.

Fig. 11 presents the distribution of MAPE values for each tuber across different models and experiments, where the boxes indicate the lower quartiles, medians, and upper quartiles. First, the error distributions produced by Non-A are more compact than those of Non-C, indicating that contrastive learning effectively reduces estimation bias. Second, the lower quartiles of MAPE for Non-C are smaller than those of Non-A, demonstrating that attention mechanisms further improve BGB estimation accuracy. Third, the lower quartiles, medians, and upper quartiles of MAPE values of our model are notably lower than those of both Non-A and Non-C, further confirming the contributions of attention mechanisms and contrastive learning to BGB estimation. In addition, we compare our model with ResNet [32], a baseline model that uses a convolution neural network with residual connections as its backbone. Our model achieves superior performance in terms of error distribution and estimation accuracy compared with ResNet [32]. This further demonstrates the efficacy of the attention mechanisms and contrastive learning in enhancing the convolution neural network, thereby enabling accurate estimation of BGB of tubers.

5.8 Discussion

Expand the dataset in the future. In this study, due to practical constraints such as sensor deployment, measurement time, and other experimental costs, RSS data are collected from only 26 tubers. This limited number may not fully capture all possible tuber shapes and biomass, potentially affecting the generalizability of DNN models. For example, as shown in Table 5, the performance of all models in leave-2-out experiments is lower than



(a) MAPE (%). (b) RMSE (g).

Fig. 13. MAPE and RMSE of different models under leave-1-out and leave-2-out experiments in outdoor environments. "Trans", "C-GRU", and "C-Trans" denote Transformer [23], CNN-GRU [47], and CNN-Transformer [18], respectively.

that in leave-1-out experiments, indicating that removing even a single tuber from the training set can reduce inference performance. These results highlight the need to increase the number of tubers in the dataset to improve estimation accuracy. Furthermore, as measurements are collected over time for each tuber, RSS samples obtained by segmenting link measurements along the temporal dimension are inherently correlated. These correlations can exacerbate data imbalance and potentially limit the generalizability of DNN models. In future work, we plan to expand the dataset by including more tubers of varying shapes, biomass, and growth conditions. This expansion is expected to enhance dataset diversity, reduce sample correlations, and improve the robustness and generalizability of the proposed framework. In addition, we plan to determine the optimal number of tubers according to the target application scenario.

Extend biomass estimation to multiple below-ground tubers and other crop types. In this study, we focus on biomass estimation in the single-tuber scenario. To further validate its feasibility in multi-tuber settings, we perform additional experiments using 35 different combinations of potato tubers in a greenhouse, with each combination containing one, two, or three tubers. RSS data from these combinations are used in leave-k-out evaluations, with k set to 1 and 2. Fig. 12 presents the MAPE results and the RSME results of various models in both leave-1-out and leave-2-out experiments. Our model outperforms baseline models, with average MAPE and RMSE values of 1.41% and 3.66 g, respectively. These results demonstrate the feasibility of our framework in multi-tuber scenarios. In future work, we plan to extend the framework to more complex multi-tuber scenarios by increasing the number and variety of tubers in each combination, as well as incorporating diverse multi-tuber sensing environments, thereby enhancing its practical applicability. In addition, this work focuses on sensing potato tubers, representing the first step towards non-destructive root tuber biomass sensing using a wireless network. In the future, we plan to expand our work to include a wider range of root tuber types.

Extend to outdoor biomass sensing. In this study, the proposed framework is suitable for deployment in indoor environments, e.g., below-ground biomass (BGB) estimation in a greenhouse. In our recent work [49], we deploy our framework in outdoor environments to validate its feasibility. Using RSS data collected from 18 potato tubers, we perform leave-1-out and leave-2-out evaluations, and Fig. 13 presents the MAPE and the root mean squared error (RMSE) results of various models. Compared with baseline models, our model achieves the best estimation performance. For example, across two evaluation experiments, the average MAPE and RMSE values of our model are 5.69% and 9.96 g, demonstrating the efficacy of our framework for real-world deployment. However, estimation performance in the outdoor environment is lower than that in the indoor environment. In practice, sensing performance is often degraded by environmental noise [3], including other plants in the sensing area and weather conditions that can alter soil properties, such as wind, rain, and sunlight. To further enhance outdoor

sensing performance, we first plan to replace the omni-directional antennas with directional ones for sensor nodes, thereby improving the quality of wireless link observations [52]. Second, we plan to collect RSS data from a larger number of tubers for training, as increasing the diversity of training samples can further improve the generalizability of data-driven models [61]. Third, we plan to explore transfer learning methods [39] to enable a model trained in indoor environments to adapt to outdoor environments. Finally, we can increase the number of sensor nodes to further improve the estimation performance by providing richer measurement information.

Extend to other agricultural indicators. In this study, our framework treats biomass as a scalar value and directly estimates it using RSS measurements. In practice, tuber shape and density influence wireless signal propagation, thereby affecting BGB estimation. Meanwhile, they also serve as important indicators in agricultural applications. Our recent work [51] demonstrates that RSS measurements from a low-cost wireless network can be used to reconstruct the maximum cross-sections of root tubers. In future work, we plan to use these reconstructions as priors to enhance BGB estimation. Furthermore, we plan to deploy a multi-layer wireless network to capture 3D volumetric shapes of root tubers and estimate their density, providing additional indicators for agricultural applications and further improving BGB estimation performance.

6 Related Work

Crop Biomass Estimation. As an important phenotypic trait, biomass plays a crucial role in each stage of crop growth. With the advancement of sensing technology, numerous methods have been developed for crop biomass estimation [6, 20, 33, 42]. For example, [22] combines hyperspectral remote sensing with deep learning models to estimate potato biomass, demonstrating the efficacy of remote sensing for AGB estimation. Furthermore, [44] develops a novel sensing system for wheat biomass and yield estimation, in which two popular multispectral cameras are mounted on a UAV for data acquisition. Extensive experiments performed under varying sun angles and phenological stages demonstrate its efficacy, providing a reliable approach for the quantitative assessment of crop traits in the field. Although these methods achieve satisfactory performance for AGB estimation, accurately estimating BGB of root tubers remains challenging due to their below-ground nature [21].

To achieve non-destructive estimation of BGB, recent studies focus on using GPR [2, 41, 65, 66], which transmits radio frequency signals that penetrate soil to detect below-ground targets. For instance, [2] employs GPR for non-destructive assessment of cassava root biomass, highlighting its potential in cassava breeding. Meanwhile, [66] uses GPR to estimate the spatial distributions and biomass of roots, showing its efficacy in inferring both spatial patterns and quantitative traits. Although these GPR-based methods hold potential for BGB sensing, their broader applicability is constrained by high costs, substantial power requirements, and the large physical size of GPR systems. In this study, we propose an alternative approach based on a wireless network for BGB estimation of root tubers. Its signals can penetrate soil for sensing, and compared with GPR, its lower cost, reduced power consumption, and compact size further enhance its feasibility for real-world deployment.

Wireless Network Sensing. Recent research on wireless networks is shifting from purely communication-oriented designs [40, 55] to integrated sensing and communication (ISAC) [60], providing a new paradigm for wireless sensing [4, 7, 27, 34, 45, 51]. For example, [4] proposes a noteworthy contribution, using a wireless network to assess moisture levels in stored rice. The study introduces an approach that combines RSS data with regression-based machine learning to provide a non-invasive, contactless method for obtaining a 3D volumetric distribution of moisture content within stored rice grains. Another significant contribution is presented in [51], which proposes a wireless network-based solution to reconstruct maximum cross-section images of below-ground root tubers. Furthermore, DNNs have been widely applied in wireless network sensing. For example, [27] and [1] propose to use DNN models to improve the performance and robustness of wireless network-based sensing. In contrast to previous works, we extend the sensing scenario to BGB estimation of root tubers, offering a new crop

trait sensing capability. In addition, we design a novel DNN model that incorporates attention mechanisms and contrastive learning to accurately estimate BGB from RSS data.

Contrastive Learning. Contrastive learning is a pairwise representation learning technique that separates semantically divergent samples while pulling together similar samples in feature space [15, 24, 29, 30, 62]. In recent years, numerous contrastive learning methods have been proposed to enhance the performance and robustness of DNN models. For example, MoCo [24] is proposed as an unsupervised contrastive method that performs contrastive learning using a dynamic queue and a momentum encoder. SupCon [30] is a supervised contrastive learning method that leverages label information to guide representation learning. However, SupCon tends to learn a feature space biased toward the majority of samples and therefore introduces a k-positive contrastive loss to mitigate this bias. In addition, RNC [62] ranks samples and contrasts them based on their relative rankings to enforce continuous representations suitable for regression tasks. Inspired by these works, we introduce a contrastive learning method into our model to reduce estimation bias for high-frequency labels, thereby enhancing its robustness and generalizability in BGB estimation.

7 Conclusion

This paper proposes a novel, non-destructive framework for sensing root tuber BGB. It constructs a new BGB dataset with more than 700,000 RSS measurements collected from a low-cost wireless network, using both frequency and spatial diversities of the measurements to investigate tuber biomass sensing. Furthermore, it proposes a novel DNN model that integrates a convolution neural network and attention mechanisms to extract discriminative features from RSS data for accurate BGB estimation. In addition, it introduces a contrastive learning method to further reduce estimation bias and enhance generalizability of the DNN model. Extensive evaluation results demonstrate the efficacy of the framework in various BGB sensing scenarios.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62350710797.

References

[1] Sayed Saad Afzal, Atsutse Kludze, Subhajit Karmakar, Ranveer Chandra, and Yasaman Ghasempour. 2023. AgriTera: accurate non-invasive fruit ripeness sensing via sub-terahertz wireless signals. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–15.

[2] Afolabi Agbona, Brody Teare, Henry Ruiz-Guzman, Iliyana D Dobreva, Mark E Everett, Tyler Adams, Osval A Montesinos-Lopez, Peter A Kulakow, and Dirk B Hays. 2021. Prediction of root biomass in cassava based on ground penetrating radar phenomics. *Remote Sensing* 13, 23 (2021), 4908.

[3] Cesare Alippi, Maurizio Bocca, Giacomo Boracchi, Neal Patwari, and Manuel Roveri. 2016. RTI Goes Wild: Radio Tomographic Imaging for Outdoor People Detection and Localization. *IEEE Transactions on Mobile Computing* 15, 10 (2016), 2585–2598.

[4] Abd Alazeez Almaleeh, Ammar Zakaria, Latifah Munirah Kamarudin, Mohd Hafiz Fazalul Rahiman, David Lorater Ndzi, and Ismahadi Ismail. 2022. Inline 3D volumetric measurement of moisture content in rice using regression-based ML of RF tomographic imaging. *Sensors* 22, 1 (2022), 405.

[5] Gamil Alsharahi, A Mostapha, Ahmed Faize, and Abdellah Driouach. 2016. Modelling and simulation resolution of ground-penetrating radar antennas. *Journal of electromagnetic engineering and science* 16, 3 (2016), 182–190.

[6] Bikram Pratap Banerjee, German Spangenberg, and Surya Kant. 2021. CBM: An IoT enabled LiDAR sensor for in-field crop height and biomass measurements. *Biosensors* 12, 1 (2021), 16.

[7] Jan Bauer and Nils Aschenbruck. 2020. Towards a low-cost rssi-based crop monitoring. *ACM Transactions on Internet of Things* 1, 4 (2020), 1–26.

[8] Lokesh Borawar and Ravinder Kaur. 2023. ResNet: Solving vanishing gradient in deep networks. In *Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022*. Springer, 235–247.

[9] John R Butnor, JA Doolittle, Kurt H Johnsen, L Samuelson, T Stokes, and L Kress. 2003. Utility of ground-penetrating radar as a root biomass survey tool in forest systems. *Soil Science Society of America Journal* 67, 5 (2003), 1607–1615.

[10] Tianfeng Chai and Roland R Draxler. 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development* 7, 3 (2014), 1247–1250.
[11] Samprit Chatterjee and Ali S Hadi. 2015. *Regression analysis by example*. John Wiley & Sons.
[12] Xingcan Chen, Yi Zou, Chenglin Li, and Wendong Xiao. 2024. A deep learning based lightweight human activity recognition system using reconstructed WiFi CSI. *IEEE Transactions on Human-Machine Systems* 54, 1 (2024), 68–78.
[13] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).
[14] Haitao Da, Yaxin Li, Le Xu, Shuai Wang, Limin Hu, Zhengbang Hu, Qiaorong Wei, Rongsheng Zhu, Qingshan Chen, Dawei Xin, and Zhenqing Zhao. 2025. Advancing soybean biomass estimation through multi-source UAV data fusion and machine learning algorithms. *Smart Agricultural Technology* 10 (2025), 100778.
[15] Gaole Dai, Huatao Xu, Hyungjun Yoon, Mo Li, Rui Tan, and Sung-Ju Lee. 2024. ContrastSense: domain-invariant contrastive learning for in-the-wild wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 4 (2024), 1–32.
[16] Wei Dai, Furong Shi, Xinyu Wang, Haixia Xu, Liming Yuan, and Xianbin Wen. 2024. A multi-scale dense residual correlation network for remote sensing scene classification. *Scientific Reports* 14, 1 (2024), 22197.
[17] Álvaro G Dieste, Francisco Argüello, Dora B Heras, Paul Magdon, Anja Linstädter, Olena Dubovyk, and Javier Muro. 2024. ResNeTS: a ResNet for Time Series Analysis of Sentinel-2 Data Applied to Grassland Plant-Biodiversity Prediction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2024).
[18] Jiangli Du, Yue Zhang, Pengxin Wang, Kevin Tansey, Junming Liu, and Shuyu Zhang. 2025. Enhancing Winter Wheat Yield Estimation With a CNN-Transformer Hybrid Framework Utilizing Multiple Remotely Sensed Parameters. *IEEE Transactions on Geoscience and Remote Sensing* (2025).
[19] Lisa Dunlap, Alyssa Umino, Han Zhang, Jiezhi Yang, Joseph E Gonzalez, and Trevor Darrell. 2023. Diversify your vision datasets with automatic diffusion-based augmentation. *Advances in neural information processing systems* 36 (2023), 79024–79034.
[20] Salah Elsayed, Salah El-Hendawy, Mosaad Khadr, Osama Elsherbiny, Nasser Al-Suhaibani, Majed Alotaibi, Muhammad Usman Tahir, and Waleed Darwish. 2021. Combining thermal and RGB imaging indices with multivariate and data-driven modeling to estimate the growth, water status, and yield of potato under different drip irrigation regimes. *Remote Sensing* 13, 9 (2021), 1679.
[21] Yiguang Fan, Yang Liu, Jibo Yue, Xiuliang Jin, Riqiang Chen, Mingbo Bian, Yanpeng Ma, Guijun Yang, and Haikuan Feng. 2024. Estimation of potato yield using a semi-mechanistic model developed by proximal remote sensing and environmental variables. *Computers and Electronics in Agriculture* 223 (2024), 109117.
[22] Haikuan Feng, Yiguang Fan, Jibo Yue, Mingbo Bian, Yang Liu, Riqiang Chen, Yanpeng Ma, Jiejie Fan, Guijun Yang, and Chunjiang Zhao. 2025. Estimation of potato above-ground biomass based on the VGC-AGB model and deep learning. *Computers and Electronics in Agriculture* 232 (2025), 110122.
[23] Fengwei Guo, Pengxin Wang, Kevin Tansey, Yue Zhang, Mingqi Li, Junming Liu, and Shuyu Zhang. 2024. A novel transformer-based neural network under model interpretability for improving wheat yield estimation using remotely sensed multi-variables. *Computers and Electronics in Agriculture* 223 (2024), 109111.
[24] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9729–9738.
[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
[26] Bao Huy, Nguyen Quy Truong, Krishna P. Poudel, Hailemariam Temesgen, and Nguyen Quy Khiem. 2024. Multi-output deep learning models for enhanced reliability of simultaneous tree above- and below-ground biomass predictions in tropical forests of Vietnam. *Computers and Electronics in Agriculture* 222 (2024), 109080.
[27] Sijie Ji, Xuanye Zhang, Yuanqing Zheng, and Mo Li. 2023. Construct 3d hand skeleton with commercial wifi. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*. 322–334.
[28] Fred Collopy J.Scott Armstrong. 1992. Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting* 8, 1 (1992), 69–80.
[29] Mahsa Keramati, Lili Meng, and R. David Evans. 2024. ConR: Contrastive Regularizer for Deep Imbalanced Regression. arXiv:2309.06651 https://arxiv.org/abs/2309.06651
[30] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. *Advances in neural information processing systems* 33 (2020), 18661–18673.
[31] Mo Li and Yunhao Liu. 2007. Underground structure monitoring with wireless sensor networks. In *Proceedings of the 6th international conference on Information processing in sensor networks*. 69–78.
[32] Tao Liu, Tianle Yang, Shaolong Zhu, Nana Mou, Weijun Zhang, Wei Wu, Yuanyuan Zhao, Zhaosheng Yao, Jianjun Sun, Chen Chen, Chengming Sun, and Zujian Zhang. 2024. Estimation of wheat biomass based on phenological identification and spectral response. *Computers and Electronics in Agriculture* 222 (2024), 109076.

[33] Yang Liu, Haikuan Feng, Jibo Yue, Xiuliang Jin, Zhenhai Li, and Guijun Yang. 2022. Estimation of potato above-ground biomass based on unmanned aerial vehicle red-green-blue images with different texture features and crop height. *Frontiers in Plant Science* 13 (2022), 938216.

[34] Yutong Liu, Landu Jiang, Linghe Kong, Qiao Xiang, Xue Liu, and Guihai Chen. 2021. Wi-Fruit: See through fruits with smart devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–29.

[35] Yibo Liu, Jie Pei, Yaopeng Zou, Shaofeng Tan, Yinan He, Xiaopo Zheng, Tianxing Wang, Huajun Fang, Li Wang, and Jianxi Huang. 2025. A Novel Hybrid-DCNN-Based Framework for Enhanced Rice Aboveground Biomass Estimation Under Limited Samples. *IEEE Transactions on Geoscience and Remote Sensing* 63 (2025), 1–16.

[36] Shanjun Luo, Xueqin Jiang, Yingbin He, Jianping Li, Weihua Jiao, Shengli Zhang, Fei Xu, Zhongcai Han, Jing Sun, Jinpeng Yang, et al. 2022. Multi-dimensional variables and feature parameter selection for aboveground biomass estimation of potato based on UAV multispectral imagery. *Frontiers in Plant Science* 13 (2022), 948249.

[37] Michał Maj, Tomasz Rymarczyk, Michał Styła, Tomasz Cieplak, Damian Pliszczuk, and Jakub Pizoń. 2025. Poster: Optimizing Radio Tomography with Edge Computing: A Low-Latency Approach for Human Detection. In *Proceedings of the 23rd ACM Conference on Embedded Networked Sensor Systems*. 656–657.

[38] Jean-Pascal Matteau, Paul Celicourt, Elnaz Shahriarina, Philipe Letellier, Thiago Gumiere, and Silvio J Gumiere. 2022. Relationship between irrigation thresholds and potato tuber depth in sandy soil. *Frontiers in Soil Science* 2 (2022), 898618.

[39] Cong T. Nguyen, Nguyen Van Huynh, Nam H. Chu, Yuris Mulya Saputra, Dinh Thai Hoang, Diep N. Nguyen, Quoc-Viet Pham, Dusit Niyato, Eryk Dutkiewicz, and Won-Joo Hwang. 2021. Transfer Learning for Future Wireless Networks: A Comprehensive Survey. arXiv:2102.07572

[40] Soumyadeep Paty, Arindam Biswas, Sonia Djebali, Guillaume Guerard, and Supreeti Kamilya. 2025. IoT-Enabled Methane Monitoring and LSTM-Based Forecasting System for Enhanced Safety in Underground Coal Mining. *ACM Transactions on Internet of Things* 6, 1 (2025), 1–29.

[41] Henry Ruiz-Guzman, Tyler Adams, Afolabi Agbona, Matthew Wolfe, Mark Everett, Jean-Francois Chamberland, and Dirk B Hays. 2025. Thresholding and continuous wavelet transform (CWT) analysis of Ground Penetrating Radar (GPR) data for estimation of potato biomass. *Computers and Electronics in Agriculture* 232 (2025), 110114.

[42] Arun Kumar Sangaiah, Jayakrishnan Anandakrishnan, Venkatesan Meenakshisundaram, Mohd Amiruddin Abd Rahman, Padmapriya Arumugam, and Mrinali Das. 2024. Edge-IoT-UAV adaptation toward precision agriculture using 3d-lidar point clouds. *IEEE Internet of Things Magazine* (2024).

[43] Hemant Servia, Sajid Pareeth, Claire I Michailovsky, Charlotte de Fraiture, and Poolad Karimi. 2022. Operational framework to predict field level crop biomass using remote sensing and data driven models. *International Journal of Applied Earth Observation and Geoinformation* 108 (2022), 102725.

[44] Sahameh Shafiee, Tomasz Mroz, Ingunn Burud, and Morten Lillemo. 2023. Evaluation of UAV multispectral cameras for yield and biomass prediction in wheat under different sun elevation angles and phenological stages. *Computers and Electronics in Agriculture* 210 (2023), 107874.

[45] Cong Shi, Jian Liu, Hongbo Liu, and Yingying Chen. 2021. WiFi-enabled user authentication through deep learning in daily activities. *ACM Transactions on Internet of Things* 2, 2 (2021), 1–25.

[46] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034* (2013).

[47] Jie Wang, Pengxin Wang, Huiren Tian, Kevin Tansey, Junming Liu, and Wenting Quan. 2023. A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables. *Computers and Electronics in Agriculture* 206 (2023), 107705.

[48] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. 2020. ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 11534–11542.

[49] Tao Wang. 2025. *Received Signal Strength Measurements for Below- ground Potato Tuber Biomass Estimation in Outdoor Environments*. doi:10.5281/zenodo.18375100

[50] Tao Wang, Yang Zhao, Jie Liu, and Yujie Zhuang. 2024. Demo Abstract: Underground Potato Root Tuber Sensing via a Wireless Network. In *2024 23rd ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 251–252.

[51] Tao Wang, Yang Zhao, Jinghua Wang, Zhibin Huang, Jie Liu, and Qiaorong Wei. 2025. See-Through Soil: Underground Root Tuber Sensing With RF Sensor Networks. *IEEE Transactions on Geoscience and Remote Sensing* 63 (2025), 1–15.

[52] Bo Wei, Ambuj Varshney, Neal Patwari, Wen Hu, Thiemo Voigt, and Chun Tung Chou. 2015. dRTI: directional radio tomographic imaging. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks* (Seattle, Washington) *(IPSN '15)*. Association for Computing Machinery, New York, NY, USA, 166–177.

[53] Cort J. Willmott and Kenji Matsuura. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30, 1 (2005), 79–82.

[54] Joey Wilson and Neal Patwari. 2010. Radio tomographic imaging with wireless networks. *IEEE Transactions on Mobile Computing* 9, 5 (2010), 621–632.
[55] Damien Wohwe Sambo and Anna Förster. 2023. Wireless underground sensor networks: A comprehensive survey and tutorial. *Comput. Surveys* 56, 4 (2023), 1–44.
[56] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*. 3–19.
[57] Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang. 2021. Cvt: Introducing convolutions to vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*. 22–31.
[58] Tomoaki Yamaguchi and Keisuke Katsura. 2024. A novel neural network model to achieve generality for diverse morphologies and crop science interpretability in rice biomass estimation. *Computers and Electronics in Agriculture* 218 (2024), 108653.
[59] Sen Yang, Quan Feng, Wanxia Yang, and Xueze Gao. 2023. Simple, low-cost estimation of potato above-ground biomass using improved canopy leaf detection method. *American Journal of Potato Research* 100, 2 (2023), 143–162.
[60] Zhiyun Yao, Kai Niu, Xuanzhi Wang, Rong Zheng, Junzhe Wang, Duo Zhang, and Daqing Zhang. 2025. WiCaliper: Simultaneous Material and 3D Size Sensing for Everyday Objects Using WiFi. *IEEE Journal on Selected Areas in Communications* (2025).
[61] Yu Yu, Shahram Khadivi, and Jia Xu. 2022. Can Data Diversity Enhance Learning Generalization?. In *Proceedings of the 29th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 4933–4945.
[62] Kaiwen Zha, Peng Cao, Jeany Son, Yuzhe Yang, and Dina Katabi. 2023. Rank-n-contrast: Learning continuous representations for regression. *Advances in Neural Information Processing Systems* 36 (2023), 17882–17903.
[63] Dabao Zhang. 2017. A coefficient of determination for generalized linear models. *The American Statistician* 71, 4 (2017), 310–316.
[64] Fuchang Zhang, Tengfei Bao, Hong Huo, and Tao Fang. 2025. TransformBioRegressor: A Transformer-based Grassland Biomass Estimator using Multi-source Data. In *2025 4th Asia Conference on Algorithms, Computing and Machine Learning (CACML)*. 1–8.
[65] Luyun Zhang and Xihong Cui. 2024. A Simple Method for Estimating Root Spatial Distribution and Root Biomass of Shrubs Based on Ground-Penetrating Radar. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 6488–6491.
[66] Luyun Zhang, Zheng Zhang, Li Guo, Xihong Cui, John R Butnor, Shupeng Li, Xin Cao, and Xuehong Chen. 2024. A simple method for estimating the coarse lateral root biomass of shrubs using ground-penetrating radar: Validation by Caragana microphylla Lam. in Inner Mongolia. *Science of the Total Environment* 919 (2024), 170897.
[67] Qiming Zhang, Jing Zhang, Yufei Xu, and Dacheng Tao. 2024. Vision transformer with quadrangle attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 5 (2024), 3608–3624.
[68] Shaohua Zhang, Xinghui Qi, Jianzhao Duan, Xinru Yuan, Haiyan Zhang, Wei Feng, Tiancai Guo, and Li He. 2024. Comparison of attention mechanism-based deep learning and transfer strategies for wheat yield estimation using multisource temporal drone imagery. *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024), 1–23.

Received 21 November 2025; revised 26 January 2026; accepted 12 February 2026