Supplementary Materials

1. Actual pictures of tomatoes in the middle and late stages of growth are used to prove that all tomatoes do not suffer from significant nutrient stress (yellowing, chlorosis, plant failure, etc.) and are all in a state of mild nutrient stress.

Figure S1. Real photos of tomatoes in the middle and late stages of growth

2. Experiment details:

As described in Section 2.1, tomato seedlings were transplanted into rock wool blocks on March 15, 2024, and irrigation with nutrient solutions of varying concentrations was initiated. As outlined in Section 2.2, spectral data collection from the tomato plants commenced on March 21, 2024, and continued until June 21, 2024, spanning a three-month period, with a total of 11 rounds of data collection. During the initial collection, the tomato plants were at the first inflorescence stage (no visible fruit). The final collection occurred during the fruit ripening stage of the 3rd to 4th clusters (growth rates varied among different groups). At this time, the average daytime temperature in the greenhouse had exceeded 30°C, approaching the critical temperature for normal tomato growth. Spectral data were collected at intervals of approximately 10 days.

In Section 3.4, the RF and CNN+LSTM models were validated using all spectral data collected throughout the sampling period.

3. Spectral data collection method

Figure S2. Spectral data collection method

4. Operations to mitigate the effects of overfitting

For the validation of model performance under the complete training set in Section 3.3, we compared the effects of different data optimization methods on mitigating overfitting. The measures employed in this study to address overfitting included model simplification, early stopping, L2 regularization, Dropout, and dynamic learning rate. Among these, model simplification, early stopping, and Dropout showed limited effectiveness in alleviating overfitting. After applying these methods, the printed model prediction accuracy curves, as shown in Figure S3, exhibited noticeable "spikes," which are a clear indication of overfitting. When dynamic learning rate optimization was implemented, the model prediction accuracy curves, depicted in Figure S4, demonstrated significant improvement in overfitting. The most effective reduction in overfitting was achieved with L2 regularization, as illustrated by the model prediction accuracy curves in Figure S5.

Regarding the Leave-One-Group-Out validation in Section 3.4, we believe that the primary reason for the model's inability to correctly fit the data is the insufficient number of control groups, which goes far beyond simply addressing overfitting to improve prediction accuracy. Taking research on predicting tomato leaf nutrient content as an example: such studies typically use sample sizes ranging from 400 to 600. However, in this experiment, there are only 16 samples (despite having a large number of replicates per sample, the sample categories are not diverse).

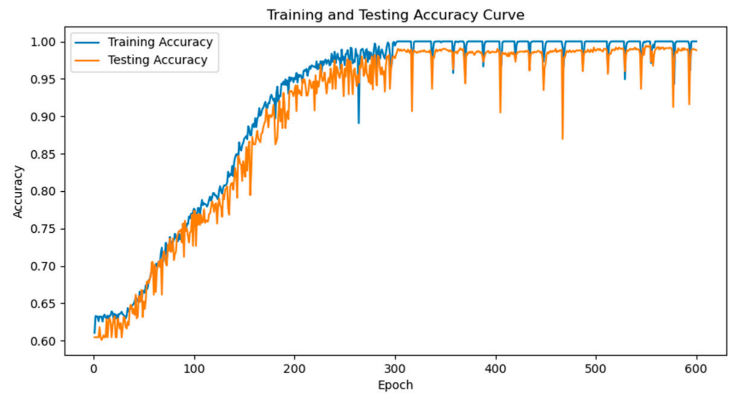

Figure S3. Unoptimized model prediction accuracy curve

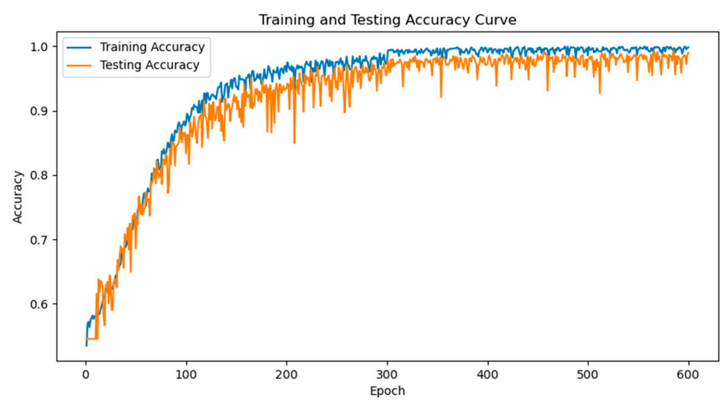

Figure S4. Model prediction accuracy curve after dynamic learning rate optimization

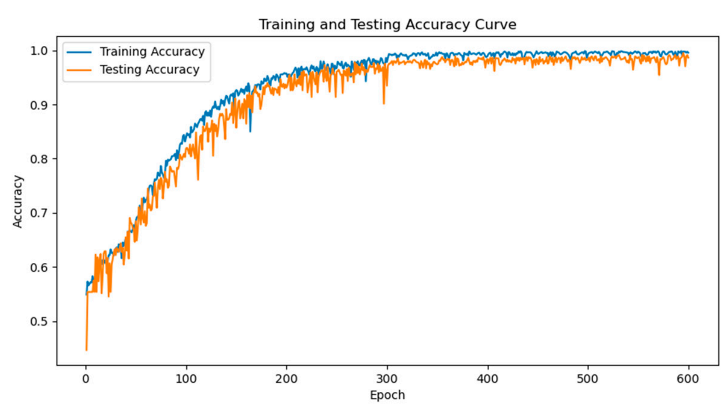

Figure S5. Model prediction accuracy curve after L2 regularization optimization