horticulturae

MDPI

Article

Non-Destructive Prediction of Soluble Solid Content in Kumquats Using a Multi-Scale Convolutional Neural Network

Xian Liu [1], Ling Zhang [2,3], Xiaoxin Shi [2,3], Weibin Tong [2,3] and Juan Lin [2,3,*]

[1] Institute of Digital Agriculture, Fujian Academy of Agricultural Sciences, Fuzhou 350001, China
[2] College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350008, China
[3] Smart Agriculture and Forestry Key Laboratory of Fujian University, Fujian Agriculture and Forestry University, Fuzhou 350008, China
* Correspondence: juan.lin@fafu.edu.cn

Highlights

What are the main findings?

- A multi-scale convolutional neural network was developed to predict the soluble solid content of kumquats using near-infrared spectra collected over 900–1700 nm.
- The proposed MS-CNN achieved an R^2 of 0.88, an RMSE of 0.62, and an MAE of 0.51 on the prediction set.
- The MS-CNN showed better predictive performance than the evaluated conventional machine-learning and deep-learning models.

What are the implications of the main findings?

- The proposed method provides a rapid and nondestructive approach for evaluating the internal quality of kumquats.
- The method has potential for practical fruit grading and quality-control applications.

Abstract

Traditional methods for detecting the soluble solid content (SSC) of kumquats are often destructive, time-consuming, and inefficient. In this study, a multi-scale convolutional neural network (MS-CNN)-based method is proposed for the rapid and non-destructive prediction of kumquat SSC. By integrating near-infrared spectroscopy (900–1700 nm) with deep learning, 424 spectral samples of kumquats were collected and modeled using the MS-CNN framework. The proposed model adopts a multi-scale feature extraction structure inspired by the Inception architecture, which effectively enhances the representation of spectral features and reduces overfitting. Experimental results showed that the MS-CNN achieved an R_p^2 of 0.88, an RMSEP of 0.62 °Brix, and an MAEP of 0.51 °Brix on the internal prediction set. Among the evaluated models, the MS-CNN achieved the highest R_p^2, while its RMSEP was comparable to that of PLSR and lower than those of SVR, BP, CNN, and BiLSTM. The proposed approach enables fast, accurate, and non-destructive prediction of kumquat SSC, providing a novel technical solution for fruit quality assessment. This work holds significant theoretical and practical value, and future efforts will focus on expanding the dataset, optimizing the network structure, exploring multi-index joint prediction, and promoting its real-world application.

Keywords: near-infrared spectroscopy; kumquat; soluble solids; regression prediction

Academic Editor: Riccardo Lo Bianco

Received: 27 June 2026
Revised: 13 July 2026
Accepted: 16 July 2026
Published: 19 July 2026

1. Introduction

Kumquat (also known as Jin Gan) is a specialty agricultural product of Youxi County, Sanming City, Fujian Province, China, and is recognized as a national geographical indication product. Youxi kumquat is characterized by an attractive shape, bright color, thin peel, and few seeds, and its nutritional value is among the highest in citrus fruits. It also exhibits certain medicinal properties [1]. Although a national standard for kumquat was established in this region in 2008, harvesting and sorting remain primarily manual, limiting production efficiency and scalability. New technical solutions are therefore needed to achieve intelligent and standardized production at the production source. Soluble Solid Content (SSC) is a key indicator that directly reflects fruit sweetness. It has become one of the critical parameters for evaluating fruit quality. With increasing consumer demand for superior flavor, accurate SSC detection has become increasingly vital for kumquat grading and premium fruit screening [2]. Traditional SSC detection methods mostly involve destructive sampling. They are time-consuming, inefficient, and often result in sample loss. These limitations hinder the adoption of rapid, non-destructive, and high-throughput detection in modern industrial production. Therefore, developing a rapid, accurate, and non-destructive method for SSC prediction based on near-infrared spectroscopy (NIRS) is essential for intelligent kumquat sorting and improved quality control.

In recent years, non-destructive sensing techniques, including hyperspectral and multispectral imaging as well as electronic noses, have been widely used in fruit quality assessment. Among these techniques, near-infrared spectroscopy has been widely applied in fruit quality analysis. This is due to its advantages of rapid detection, minimal sample preparation, and the ability to simultaneously analyze multiple components. When combined with both traditional and advanced machine learning methods, NIR spectroscopy has further enhanced its utility in fruit quality assessment [3,4].

Earlier studies showed that traditional modeling methods, such as Partial Least Squares Regression (PLSR) and Support Vector Regression (SVR), could achieve satisfactory predictive performance with relatively small datasets. For example, Hayato et al. [5] developed an SSC prediction model for white strawberries based on visible and near-infrared spectroscopy, achieving coefficients of determination for the calibration and prediction sets (R_c^2 and R_p^2) of 0.89 and 0.85, respectively. Yan Yu et al. [6] combined chemometric methods with a hybrid wavelength selection strategy to establish quantitative calibration models for the prediction of SSC and firmness in 'Nanguo' pears. Their proposed Si-GA-PLS model achieved R_c^2 and R_p^2 values of 0.9406 and 0.9119, respectively. Comparable results have been reported for apples [7], pears [8], and citrus [9] with prediction accuracy generally exceeding 0.90. However, studies on kumquat are relatively limited. Li Xinrong [10] established a PLSR calibration model for predicting the SSC in 'Huapi' kumquat using spectra collected over the range of 720–920 nm and preprocessed using Savitzky–Golay (SG) smoothing, achieving a correlation coefficient (r) of 0.9582 between predicted and reference SSC values, together with a root mean square error (RMSE) of 51.87% and a residual predictive deviation (RPD) of 3.24. Nevertheless, this approach used a relatively narrow spectral range and therefore may not have fully exploited the information available across a broader spectrum. Because fruits have complex and heterogeneous compositions, and their components may interact with one another, information from a single spectral band may not adequately characterize variations in SSC, which may lead to prediction errors. Moreover, traditional methods rely on manual feature extraction, which makes them susceptible to noise, baseline drift, and peak overlap and may limit their ability to characterize the nonlinear relationships present in spectral data.

In recent years, deep learning methods have shown great potential in spectral analysis due to their ability to learn features directly from data in an end-to-end manner.

One-dimensional convolutional neural networks (1D-CNNs) can automatically extract hierarchical features directly from full-spectrum data, thereby enhancing model robustness and prediction accuracy. For example, Xiang et al. [11] combined near-infrared spectroscopy with a 1D-CNN to predict the SSC of Satsuma mandarins, achieving a coefficient of determination for the prediction set (R_p^2) of 0.8655 and a root mean square error of prediction (RMSEP) of 0.3339. These results demonstrated the feasibility of applying deep learning to fruit quality assessment. Wu et al. [12] addressed the problem of limited sample availability in cherry tomatoes by using a deep convolutional generative adversarial network (DCGAN) for data augmentation. When combined with a 1D-CNN regression model, this method substantially improved prediction accuracy and generalization, achieving an R_p^2 of 0.9807. Mishra et al. [13] employed a six-layer CNN to predict the dry matter content of mangoes and reported a lower RMSEP than that obtained using conventional PLSR models. For pear SSC prediction, the model also reduced RMSE by approximately 13% compared with PLSR. Huang et al. [14] proposed an LSTMED-MLP fusion model that combines temporal features with global dependencies for apple SSC prediction. The model achieved coefficients of determination of 0.95 and 0.90 for the calibration and prediction sets, respectively, with corresponding RMSE values of 0.77 and 0.84. These results demonstrated the effectiveness of deep fusion architectures in modeling spectral and sequential data. In addition, Jun Li et al. [15] developed an in situ SSC grading model for longan using an improved BP neural network. When used with a portable visible–near-infrared spectroscopic device, the model achieved a prediction accuracy of 83.33%, demonstrating its potential for rapid, non-destructive, on-site assessment.

Despite significant advances in deep learning for fruit SSC prediction, existing methods generally suffer from single-scale feature extraction, complex network structures, and strong dependence on sample size. Traditional 1D-CNNs generally employ convolutional kernels with a fixed size, resulting in a relatively limited receptive field at each layer. Small convolutional kernels are more sensitive to local variations and narrow absorption features, whereas larger receptive fields are more suitable for capturing broad spectral trends and long-range dependencies. Therefore, a single-scale convolutional structure may not effectively represent both narrow spectral peaks and broadband information simultaneously.

To address these limitations, this study proposes a Multi-Scale 1D-CNN (MS-CNN) model. The model employs convolutional kernels with different receptive fields to extract multi-level features in parallel. This allows simultaneous modeling of local details and global trends, improving robustness against noise, peak shifts, and band overlap, while enhancing prediction accuracy. The proposed approach offers a new technical solution for rapid, non-destructive detection of kumquat SSC using near-infrared spectroscopy.

2. Experimental Section

2.1. Materials and Instruments

The main experimental equipment included the GO-IQF1700 Near-Infrared Spectrometer (Ocean Insight, Orlando, FL, USA), the H1-2000-HP-FHSA Tungsten Halogen Light Source (Ocean Insight, USA), and the PAL-1 Portable Digital Fruit Brix Meter (ATAGO, Tokyo, Japan). The spectral acquisition equipment configuration is shown in Figure 1, and the portable fruit brix meter is presented in Figure 2.

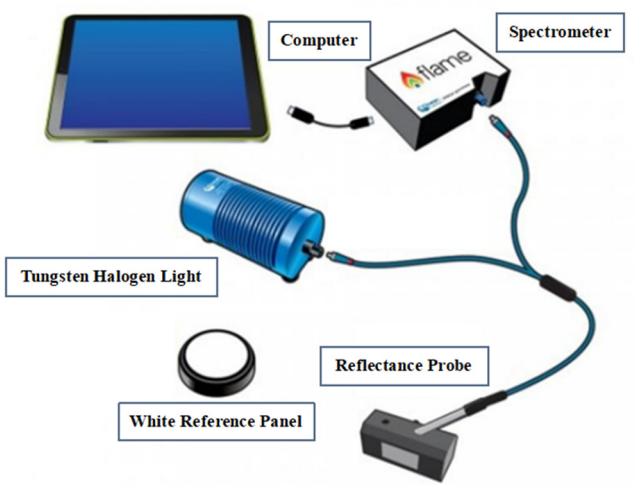

Figure 1. Spectral acquisition equipment.

Figure 2. Refractometer.

2.2. Sample Collection Methods

2.2.1. Sample Preparation

Kumquat samples were obtained from the Guanqian Town Kumquat Base in Youxi County, Sanming City, Fujian Province, China. A total of 450 samples were initially collected, and after data cleaning and the exclusion of noisy measurements, 424 valid samples were retained. The harvested kumquats were washed, sequentially numbered, and stored under laboratory conditions at 20 °C and 60% relative humidity for 24 h. Spectral measurements were performed after the samples had equilibrated to room temperature (20 °C). During spectral acquisition, areas with visible defects, including bruises and scars, were carefully avoided.

2.2.2. Near-Infrared Spectral Acquisition

The surfaces of the samples were gently wiped with tissue paper, and spectral acquisition was performed only after the fruits had equilibrated to room temperature. Before spectral acquisition, the tungsten halogen lamp was preheated for 30 min to ensure stable illumination. An expanded polystyrene (EPS) plate was used as the reference for background correction. Each sample was positioned over the probe connected to the NIR spectrometer, and its reflectance spectrum was acquired from the equatorial region over the wavelength range of 900–1700 nm. The spectrometer was configured for automatic reference correction, and 32 scans were averaged for each measurement. High-frequency noise was reduced using Savitzky–Golay (SG) smoothing with a polynomial order of 1. All spectral data were saved sequentially in the order of acquisition. The raw reflectance spectra are shown in Figure 3. Each curve represents the reflectance spectrum of one

kumquat sample and comprises reflectance values measured at 512 wavelength points over the range of 900–1700 nm.

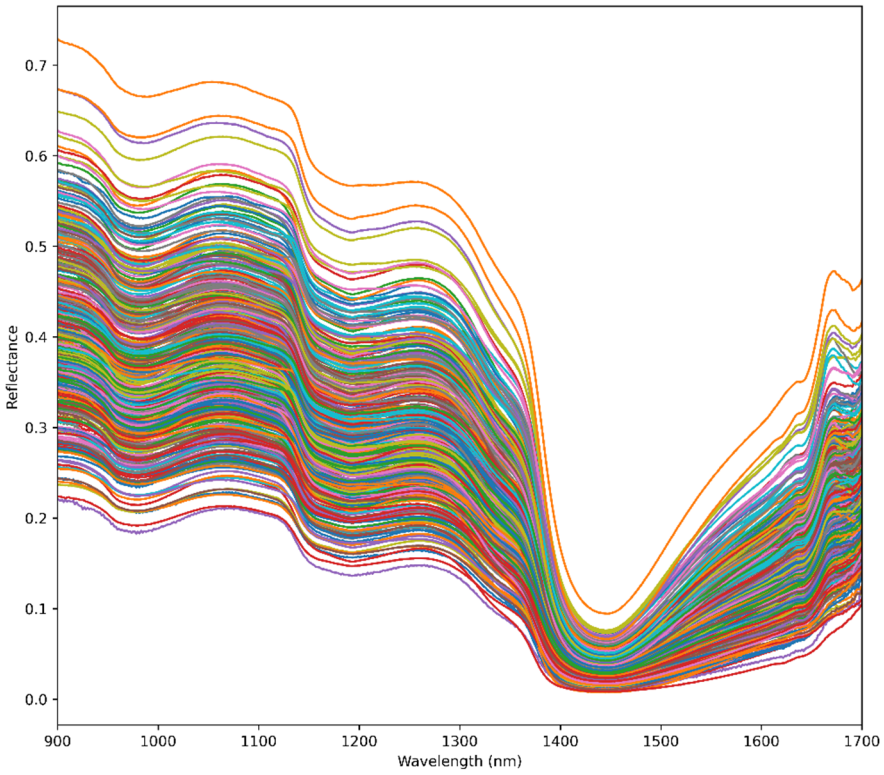

Figure 3. Near-infrared spectroscopy of kumquats. Each colored line represents the reflectance spectrum of one kumquat sample measured at 512 wavelength points over the range of 900–1700 nm.

Following spectral acquisition, drops of juice were extracted from the top, middle, and bottom regions of each fruit and placed on the measurement window of the PAL-1 portable digital refractometer. Each sample was measured in triplicate, and the average value was used as the reference soluble solid content (SSC).

2.3. Data Processing

Sample partitioning was performed using sample set partitioning based on joint X–Y distances (SPXY). This approach ensures that the samples are evenly distributed across both the spectral feature space and the target variable space, enhancing the representativeness of the resulting subsets. The dataset was split at a ratio of 7:3, yielding 296 calibration samples and 128 prediction samples.

2.4. Evaluation Metrics

To comprehensively evaluate the performance of the proposed network architecture, several commonly used regression metrics were adopted, including the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). The definitions and formulas of these metrics are presented below. The coefficient of determination (R^2) is used to measure the goodness of fit between predicted and measured values. R^2 has an upper bound of 1 and may take negative values when the model performs worse than a baseline prediction based on the mean of the

measured values. A value closer to 1 indicates better model performance, while a lower value indicates poorer model performance. The calculation of R^2 is given in Equation (1):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \overline{y_i})^2}{\sum_{i=1}^{n} (\overline{y_i} - \overline{y})^2} \quad (1)$$

where y_i represents the measured value, $\overline{y_i}$ denotes the predicted value from the model, and $\overline{y}$ is the mean value of the measured values. The mean squared error (MSE) is the average of the squared differences between the predicted values and the measured values. A smaller MSE indicates better predictive accuracy. The calculation of MSE is given in Equation (2):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y_i})^2 \quad (2)$$

The root mean squared error (RMSE) is the square root of the MSE. It is commonly used to quantify the magnitude of the differences between predicted values and measured values. A smaller RMSE indicates better predictive performance of the model. The calculation of RMSE is given in Equation (3):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y_i})^2} \quad (3)$$

Here, RMSEC and R_c^2 denote the metrics for the calibration set, while RMSEP and R_p^2 represent those for the prediction set.

The mean absolute error (MAE) is defined as the average of the absolute differences between the predicted values and the measured values. A smaller MAE indicates better predictive accuracy. The calculation of MAE is given in Equation (4):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \overline{y_i}| \quad (4)$$

where MAEC is the metric for the calibration set and MAEP is for the prediction set.

Collectively, these metrics provide a comprehensive assessment of the model's regression performance and serve as a basis for subsequent model comparison and optimization.

2.5. Experimental Environment

The hardware environment of this experiment consisted of an Intel Core i7-11700K CPU (Intel Corporation, Santa Clara, CA, USA), an NVIDIA GeForce GTX 1660 Ti GPU (NVIDIA Corporation, Santa Clara, CA, USA), and 16 GB of RAM. For the software environment, the PyTorch framework was employed, along with CUDA 11.8 as the computing platform. The programming language used was Python 3.9.

3. Multi-Scale 1D-CNN

3.1. Model Overall Architecture

To fully leverage multi-scale features in one-dimensional near-infrared spectral signals and improve quantitative regression performance, a multi-scale 1D-CNN (MS-CNN) based on the Inception architecture was developed. The model is designed for spectral quantitative regression tasks and structurally integrates multi-scale feature extraction with residual connections. This design maintains a lightweight architecture while enabling efficient feature fusion and stable training of deep networks.

The overall network architecture consists primarily of an initial convolutional layer, a multi-scale feature extraction module (Inception module), a residual fusion structure, and

a fully connected regression layer. The raw spectral signals are first processed by the initial convolutional layer for preliminary feature extraction and are subsequently passed through the multi-scale module to capture features at different receptive-field scales. The extracted features are then fused through residual connections and passed to the fully connected regression layer to produce the SSC predictions. The overall model framework is illustrated in Figure 4 and can be formally described as Equation (5):

$$\hat{y} = f_\theta(X) = \mathrm{FC}\Big(\mathrm{Res}\Big(\mathrm{Inception}\big(\mathrm{Conv1D}(X)\big)\Big)\Big) \tag{5}$$

where $X \in \mathbb{R}^{B\times 1\times L}$ represents the input spectral signal (with B denoting the batch size and L denoting the spectral length), $f_\theta(\cdot)$ denotes the model's non-linear mapping function, and $\hat{y}$ is the predicted SSC value of the kumquat.

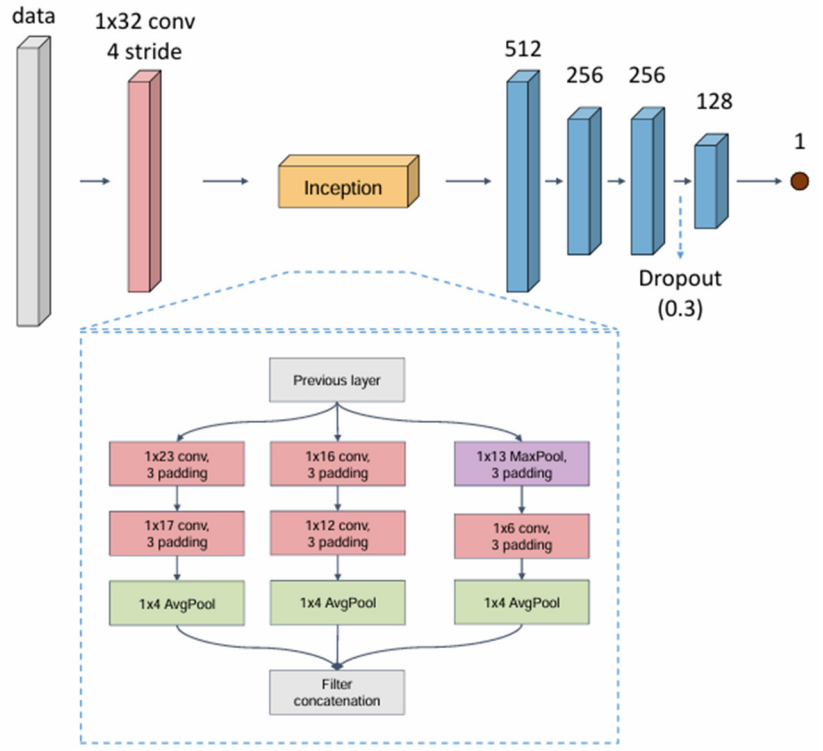

Figure 4. Framework of MS-CNN. The orange block represents the Inception module. In the enlarged Inception module, the pink, purple, and green boxes denote one-dimensional convolution, max-pooling, and adaptive average-pooling operations, respectively, while the gray boxes represent the input feature and feature-concatenation operation. The blue dashed box and connecting dashed lines indicate an enlarged view of the three-branch Inception module. The outputs of the three branches are concatenated, projected through a 1×1 convolution, and fused with the shortcut connection before being passed to the regression head.

3.2. Network Architecture

The Inception module was originally proposed by Google researchers as part of the GoogLeNet architecture [16]. Its core idea is to use parallel convolutional kernels of different sizes within the same layer to extract multi-scale features, thereby enhancing the network's ability to represent complex patterns. Traditional 1D-CNNs generally rely on single-scale convolution, which makes it challenging to capture narrow spectral features and broadband trends in spectral signals. To address this limitation, the Inception concept was adapted for spectral analysis, and a multi-scale Inception module specifically designed for near-infrared spectroscopy was developed. The module comprises three parallel branches:

- Branch 1: employs two large convolutional kernels in series with adaptive average pooling to extract global trend features using a wide receptive field;
- Branch 2: uses medium-sized kernels with varying strides to capture mesoscopic structural information;
- Branch 3: combines max pooling with small kernels to strengthen the learning of local details and narrow spectral features.

The outputs of all branches are aligned to the same length using adaptive pooling, concatenated along the channel dimension, and subsequently processed by a 1×1 convolution for channel integration and dimensionality reduction. This module effectively captures both local and global features, thereby enhancing the model's robustness to noise, baseline drift, and spectral-band interactions.

3.3. Residual Connection Mechanism

In deep convolutional networks, increasing network depth may lead to vanishing gradients, thereby hindering effective model training. To alleviate this problem, a residual connection mechanism is introduced into the multi-scale module. This mechanism uses identity mapping to enable the direct cross-layer propagation of input features, thereby improving gradient flow and preserving feature information [17].

Let the input to the multi-scale module be x and its nonlinear transformation be $F(x)$; then the output of the residual unit is expressed as Equation (6):

$$y = F(x) + x \tag{6}$$

where $F(x)$ represents the nonlinear transformation of the main branch (i.e., the multi-scale convolutional module), and x denotes the identity-mapped input.

In this model, the output features of the Inception module are fused with the input features through a residual path using element-wise addition. Adaptive average pooling is applied to adjust the temporal dimension and align the feature lengths. This design improves gradient propagation and alleviates the vanishing gradient problem during training. It also promotes feature reuse and fusion. As a result, the network maintains a strong feature-representation capability while enabling faster convergence and more stable training.

3.4. Fully Connected Regression Layer and Output Architecture

After multi-scale feature extraction and residual fusion, the high-dimensional feature maps are flattened into a one-dimensional vector and fed into a fully connected regression layer. This regression layer consists of two linear layers separated by a dropout layer to mitigate overfitting. The final prediction of kumquat SSC is given by Equation (7):

$$\hat{y} = W_2\sigma(W_1x + b_1) + b_2 \tag{7}$$

where W_1, W_2 are the weight matrices, b_1, b_2 are the bias terms, and $\sigma(\cdot)$ denotes the ReLU activation function.

This layer maps the high-dimensional feature space onto the regression output space through a nonlinear transformation of the learned spectral features. This process enables the model to learn the quantitative relationship between the complex spectral data and kumquat SSC.

3.5. Model Training Strategy

During model training, the AdamW optimizer is used to update the network parameters while balancing convergence and regularization. To further improve training stability

and convergence efficiency, a ReduceLROnPlateau learning rate scheduler is applied. When the validation loss fails to improve for a specified number of consecutive epochs, the learning rate is automatically reduced to facilitate further optimization. Mean squared error (MSE) is used as the loss function. An early-stopping strategy is employed to terminate training when the validation loss no longer improves and to retain the model with the best validation performance, thereby reducing the risk of overfitting.

In summary, the proposed MS-CNN model first extracts initial spectral features through an initial convolution, captures spectral information across multiple scales via the multi-scale feature-extraction module, and improves information flow and training stability through residual connections. With the incorporation of adaptive learning-rate scheduling and regularization strategies, the model maintains a compact architecture while effectively representing both local details and global trends in the spectral signals, thereby achieving competitive prediction performance in the kumquat SSC regression task evaluated in this study.

4. Model Performance Analysis

To comprehensively evaluate the predictive performance of the MS-CNN, two types of experiments were conducted. First, the MS-CNN was compared with several commonly used models to assess its overall predictive performance. Second, an ablation study was conducted to investigate the effects of key model components, including the Inception module, residual connections, and the max-pooling layer, on prediction accuracy.

During model training, the mean squared error (MSE) was used as the loss function, and the AdamW optimizer was employed to update the network parameters. AdamW is an extension of the Adam optimizer that decouples weight decay from the adaptive gradient-based parameter update. Unlike applying conventional L_2 regularization directly to the loss function, this decoupled weight-decay mechanism allows the regularization strength to be controlled independently of the adaptive learning rate. Consequently, AdamW can constrain excessive parameter growth, improve optimization stability, and reduce the risk of overfitting. This characteristic is particularly suitable for the relatively small spectral dataset used in this study. The initial learning rate was set to 0.005, and the model was trained for a maximum of 500 epochs. An early-stopping strategy was also adopted to terminate training when the validation loss no longer improved, thereby preventing overfitting and retaining the model with the best validation performance.

4.1. Ablation and Comparative Experimental Results

To evaluate the effectiveness of different structural designs within the model, four ablation configurations were evaluated, and the results are summarized in Table 1. Model 1 served as the baseline full-spectrum CNN structure. Model 2 incorporated the Inception module into the CNN. In Model 3, one convolutional layer in the third branch of the Inception module was replaced with a max-pooling operation based on the architecture of Model 2. Model 4 represented the complete MS-CNN architecture proposed in this study.

Table 1. Results of the ablation experiment on the MS-CNN model.

No.	Model	$RMSE_C\downarrow$ (°Brix)	$R_c^2\uparrow$	$MAE_C\downarrow$ (°Brix)	$RMSE_P\downarrow$ (°Brix)	$R_p^2\uparrow$	$MAE_P\downarrow$ (°Brix)
1	CNN	**0.49**	**0.95**	**0.38**	0.73	0.83	0.57
2	Inception + CNN	0.71	0.89	0.56	0.71	0.84	0.58
3	Inception + Max-Pool + CNN	0.63	0.91	0.51	0.63	0.87	0.51
4	MS-CNN	0.63	0.91	0.52	**0.62**	**0.88**	**0.51**

Note: ↑ indicates that a higher value is better, whereas ↓ indicates that a lower value is better. Bold values indicate the best performance in each column.

Table 1 presents the performance metrics of each model on the calibration and prediction sets, including RMSEC, R_c^2, MAE_C, $RMSE_P$, R_p^2 and MAE_P. As shown in Table 1, predictive performance generally improved as the network architecture was progressively optimized. The addition of the Inception module improved the nonlinear representation of spectral features and reduced the discrepancy between calibration and prediction performance. The further incorporation of max pooling, together with a lower dropout rate, contributed to improved predictive performance on the prediction set. The complete MS-CNN model achieved the highest R_p^2 and the lowest RMSEP among the evaluated ablation configurations while retaining a lightweight architecture. These results demonstrate that the MS-CNN can effectively capture multi-scale spectral information while preserving model simplicity.

4.2. Comparison of Kumquat SSC Prediction Models

To further evaluate the prediction performance of the MS-CNN, it was compared with five commonly used modeling methods, including PLSR, SVR, BP, CNN, and BiLSTM. All models were developed using the same spectral dataset, SG preprocessing procedure, and calibration–prediction partition. As shown in Table 2, the CNN achieved the highest calibration-set performance, with an R_c^2 of 0.95 and an RMSEC of 0.48 °Brix. However, its performance decreased on the prediction set, where the R_p^2 and RMSEP were 0.81 and 0.76 °Brix, respectively. In comparison, the MS-CNN achieved a lower calibration-set fit (R_c^2 = 0.91 and RMSEC = 0.63 °Brix) but better prediction-set performance than the CNN, with an R_p^2 of 0.88 and an RMSEP of 0.62 °Brix. These results suggest that, for the present calibration–prediction partition, the CNN fitted the calibration samples more closely, whereas the MS-CNN performed better than the CNN on the held-out prediction set. Nevertheless, the calibration and prediction sets were generated by applying the SPXY algorithm to the same pool of 424 samples rather than by using an independently collected external dataset. Therefore, the present results reflect internal prediction performance rather than independent external validation. Further evaluation using kumquat samples collected from different production areas, harvest periods, and measurement conditions is required to assess the external robustness and generalizability of the MS-CNN.

Table 2. Comparison of training results of different predictive models.

Modeling Method	Preprocessing	$R_c^2\uparrow$	$RMSE_C\downarrow$ (°Brix)	$R_p^2\uparrow$	$RMSE_P\downarrow$ (°Brix)
PLSR	SG	0.91	0.63	0.85	0.61
SVR		0.81	0.94	0.77	0.84
BP		0.81	0.94	0.73	0.91
CNN		**0.95**	**0.48**	0.81	0.76
BiLSTM		0.94	0.51	0.76	0.86
MS-CNN		0.91	0.63	**0.88**	**0.62**

Note: ↑ indicates that a higher value is better, whereas ↓ indicates that a lower value is better. Bold values indicate the best performance in each column.

5. Discussion

5.1. Comparison with Previous Fruit SSC Prediction Studies

To place the performance of the proposed MS-CNN in the context of previous fruit SSC prediction studies, representative results reported in the literature were compared, as summarized in Table 3. Except for the LSTMED-MLP model used for apple SSC prediction based on electrical parameters, the other studies mainly employed near-infrared spectroscopy combined with traditional chemometric or deep-learning methods. The proposed MS-CNN achieved an R_p^2 of 0.88 and an RMSEP of 0.62 °Brix for kumquat SSC

prediction. Compared with traditional PLSR models, its R_p^2 was higher than those reported for white strawberry (R_p^2 = 0.85), Gong pear (R_p^2 = 0.77–0.82), and citrus (R_p^2 = 0.72), and was comparable to that reported for red strawberry (R_p^2 = 0.89). However, higher R_p^2 values have been reported for 'Nanguo' pear, apple, and 'Huapi' kumquat, indicating that the proposed model does not outperform all existing models under every experimental condition. Compared with deep-learning methods, the prediction performance of the MS-CNN was comparable to that of the 1D-CNN models developed for mandarin orange and cherry tomato, both of which achieved R_p^2 values of approximately 0.87, and was close to that of the LSTMED-MLP model for apple (R_p^2 = 0.90). The training and prediction results of the MS-CNN model are shown in Figure 5.

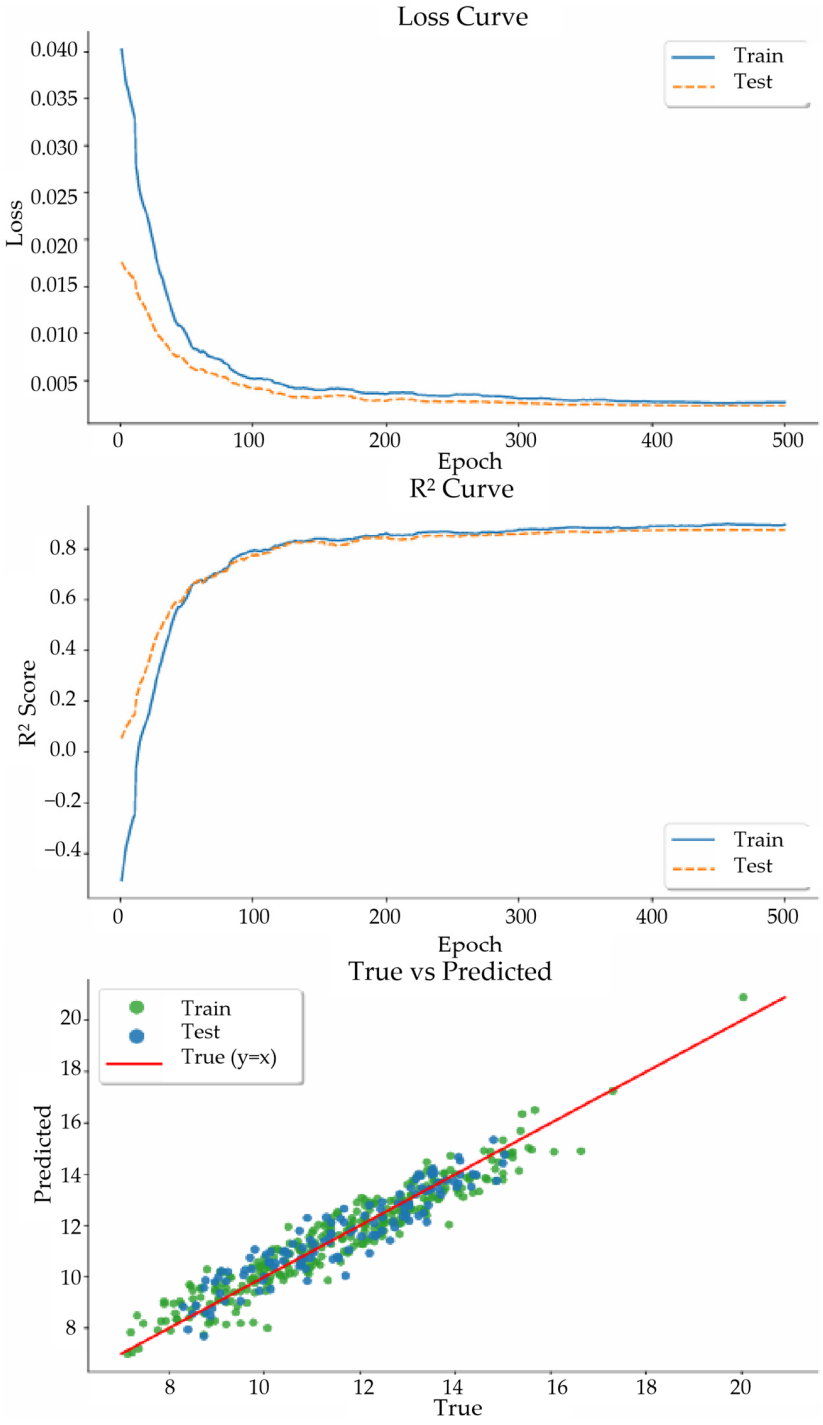

Figure 5. Training and prediction results of the MS-CNN model.

Table 3. Comparison of different fruits and their predictive model performance.

Model	Preprocessing	Fruit	R_c^2↑	$RMSE_C$↓ (°Brix)	R_p^2↑	$RMSE_P$↓ (°Brix)
PLSR	SG	White Strawberry	0.84	0.53	0.85	0.43
		Red Strawberry	0.88	0.42	0.89	0.36
PLSR	None	'Nanguo' Pear	0.96	0.16	0.94	**0.17**
PLSR	SG + MSC	Apple	0.95	0.71	**0.96**	0.55
PLSR	GFS	Gong Pear	0.83	0.42	0.77~0.82	0.45~0.62
PLSR	D1	Citrus	0.75	0.52	0.72	0.57
PLSR	SG	'Huapi' Kumquat	0.95	0.64	0.95	0.65
1D-CNN	SNV	Mandarin Orange	0.90	0.26	0.87	0.33
1D-CNNR	SG + SNV	Cherry Tomato	**0.99**	**0.11**	0.87	0.52
LSTMED-MLP	None	Apple	0.95	0.77	0.90	0.84
MS-CNN	SG	Kumquat	0.91	0.63	0.88	0.62

Note: ↑ indicates that a higher value is better, whereas ↓ indicates that a lower value is better. Bold values indicate the best performance in each column.

The competitive performance of the MS-CNN may be attributed to its parallel convolutional branches with different receptive fields, which enable the simultaneous extraction of narrow local absorption features and broad spectral trends, while the residual connections improve feature reuse and training stability.

Nevertheless, the results in Table 3 were obtained from independent studies involving different fruit species, sample sizes, SSC distributions, spectral instruments, wavelength ranges, preprocessing methods, and dataset-partitioning strategies. Therefore, these numerical results should be regarded as a contextual comparison rather than a direct ranking of model performance, and they cannot alone demonstrate the transferability of the MS-CNN across fruit species. Further external validation using independent datasets from different kumquat varieties, production regions, harvest periods, and measurement conditions is required to confirm the robustness and generalization ability of the proposed model.

5.2. Computational Complexity and Inference Efficiency

The computational complexity and single-sample processing efficiency of PLSR and MS-CNN were further evaluated using a 512-point spectrum as input. The MS-CNN contains 206,673 trainable parameters and requires approximately 1.81 million multiply–accumulate operations per prediction, whereas PLSR employs a substantially simpler linear latent-variable projection and therefore has lower computational complexity. To ensure a fair comparison, both models were evaluated on the same CPU using a batch size of one and a single CPU thread. After 100 warm-up runs, each model was repeatedly evaluated 5000 times. The average model-only inference times were 0.094 ms for PLSR and 0.445 ms for MS-CNN. When SG preprocessing and the model-specific feature transformations were included, the average single-sample processing times were 0.325 ms for PLSR and 0.902 ms for MS-CNN, corresponding to approximately 3076 and 1108 samples per second, respectively. Although the MS-CNN required greater computational resources than PLSR, its average total processing time remained below 1 ms per spectrum, supporting its potential for real-time SSC prediction. However, further system-level validation is required to evaluate the overall processing time under actual production-line conditions.

6. Conclusions

This study proposed a rapid and non-destructive method for predicting the soluble solid content (SSC) of kumquats based on a multi-scale one-dimensional convolutional neural network. By integrating near-infrared spectroscopy with deep learning, the proposed method was designed to address the limitations of conventional approaches in extracting

informative features from high-dimensional spectral data. The main contributions of this study are as follows: (1) A multi-scale convolution-based spectral analysis framework was developed to adaptively capture spectral features at different scales. (2) A feature extraction module combining the Inception structure and residual connections was designed to improve spectral feature representation and training stability. (3) A deep learning prediction model for kumquat SSC was established to achieve accurate, rapid, and non-destructive detection. On the internal prediction set, the MS-CNN achieved the highest R_p^2 among the evaluated models. Its RMSEP was comparable to that of PLSR and lower than those of SVR, BP, CNN, and BiLSTM. These findings demonstrate the potential of the multi-scale architecture for kumquat SSC prediction under the experimental conditions of this study and suggest that the proposed model could support standardized and intelligent grading in the kumquat industry.

Future research may focus on several directions: (1) expanding the sample dataset to include kumquat samples of various varieties and origins to enhance the model's robustness and generalizability; (2) incorporating an attention mechanism or self-supervised learning to further optimize the model architecture and strengthen feature representation; (3) exploring multi-indicator joint modeling approaches for comprehensive assessment of quality parameters, including SSC, acidity, and firmness; and (4) integrating portable or online near-infrared spectroscopy devices to develop real-time detection systems for practical production lines, enabling intelligent grading and monitoring of fruit quality.

Overall, the proposed MS-CNN model has significant theoretical and practical implications. The model introduces a novel approach for the non-destructive detection of kumquat SSC. It also provides a potentially useful framework for the non-destructive prediction of quality attributes in other fruits and agricultural products, although further external validation is required.

Author Contributions: Methodology, X.L., L.Z. and X.S.; investigation, X.L., L.Z. and W.T.; data curation, L.Z. and X.S.; software, X.L., L.Z. and W.T.; formal analysis, X.L. and X.S.; resources, J.L. and X.L.; supervision, J.L.; writing—original draft preparation, X.L. and L.Z.; writing—review and editing, J.L. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 11901094), the 2024 Fujian Provincial Competitive Public Welfare Project (Grant No. 2024R1085), the 2023 Putian Science and Technology Project (Grant No. 2023GM03), the Science and Technology Innovation Fund of Fujian Agriculture and Forestry University (Grant No. KFB24046), and the Key Project of Education and Scientific Research for Young and Middle-aged Teachers in Fujian Province (Grant No. JZ23001).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Li, X.; Meenu, M.; Xu, B. Recent Development in Bioactive Compounds and Health Benefits of Kumquat Fruits. *Food Rev. Int.* **2023**, *39*, 4312–4332. [CrossRef]
2. Paz, P.; Sánchez, M.-T.; Pérez-Marín, D.; Guerrero, J.E.; Garrido-Varo, A. Instantaneous Quantitative and Qualitative Assessment of Pear Quality Using near Infrared Spectroscopy. *Comput. Electron. Agric.* **2009**, *69*, 24–32. [CrossRef]
3. Ma, T.; Zhao, J.; Inagaki, T.; Su, Y.; Tsuchikawa, S. Rapid and Nondestructive Prediction of Firmness, Soluble Solids Content, and pH in Kiwifruit Using Vis–NIR Spatially Resolved Spectroscopy. *Postharvest Biol. Technol.* **2022**, *186*, 111841. [CrossRef]
4. Nagy, M.M.; Wang, S.; Farag, M.A. Quality Analysis and Authentication of Nutraceuticals Using near IR (NIR) Spectroscopy: A Comprehensive Review of Novel Trends and Applications. *Trends Food Sci. Technol.* **2022**, *123*, 290–309. [CrossRef]

5. Seki, H.; Murakami, H.; Ma, T.; Tsuchikawa, S.; Inagaki, T. Evaluating Soluble Solids in White Strawberries: A Comparative Analysis of Vis-NIR and NIR Spectroscopy. *Foods* **2024**, *13*, 2274. [CrossRef] [PubMed]
6. Yu, Y.; Yao, M. A Portable NIR System for Nondestructive Assessment of SSC and Firmness of Nanguo Pears. *LWT* **2022**, *167*, 113809. [CrossRef]
7. Zeng, S.; Zhang, Z.; Cheng, X.; Cai, X.; Cao, M.; Guo, W. Prediction of Soluble Solids Content Using Near-Infrared Spectra and Optical Properties of Intact Apple and Pulp Applying PLSR and CNN. *Acta A Mol. Biomol. Spectrosc.* **2024**, *304*, 123402. [CrossRef] [PubMed]
8. Yan-De, L.; Jun, L.; Bin, L.; Xiao-Gang, J.; Ming-Wang, Z.; Jin-Liang, Y.; Qiu, W. Robustness of global model of soluble solids in Gongli pear based on near-infrared spectroscopy. *Spectrosc. Spectr. Anal.* **2022**, *42*, 2781. [CrossRef]
9. Kim, M.-J.; Yu, W.-H.; Song, D.-J.; Chun, S.-W.; Kim, M.S.; Lee, A.; Kim, G.; Shin, B.-S.; Mo, C. Prediction of Soluble-Solid Content in Citrus Fruit Using Visible–Near-Infrared Hyperspectral Imaging Based on Effective-Wavelength Selection Algorithm. *Sensors* **2024**, *24*, 1512. [CrossRef] [PubMed]
10. Li, X.R. Non-destructive testing technology of sugar content in Huapi kumquat by near infrared spectroscopy. *J. South. Agric.* **2019**, *50*, 838. [CrossRef]
11. Zhiyong, X.; Yubin, M. Study on Prediction Model of Citrus SSC Based on CNN and near Infrared Spectroscopy. *J. Chin. Agric. Mech.* **2024**, *45*, 139. [CrossRef]
12. Wu, Z.; Liu, F.; Li, Z.; Chen, H. Spectroscopic method for detection of soluble solid content in cherry tomato using deep convolutional generative adversarial network-based data augmentation. *Food Sci.* **2025**, *46*, 214. [CrossRef]
13. Mishra, P.; Passos, D. Multi-Output 1-Dimensional Convolutional Neural Networks for Simultaneous Prediction of Different Traits of Fruit Based on near-Infrared Spectroscopy. *Postharvest Biol. Technol.* **2022**, *183*, 111741. [CrossRef]
14. Huang, L. Prediction of the soluble solid contents for apple fruit using electrical parameters. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 252. [CrossRef]
15. Li, J.; Zhang, M.; Wu, K.; Chen, H.; Ma, Z.; Xia, J.; Huang, G. Field Grading of Longan SSC via Vis-NIR and Improved BP Neural Network. *Agriculture* **2024**, *14*, 2297. [CrossRef]
16. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Boston, MA, USA, 2015; pp. 1–9.
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Las Vegas, NV, USA, 2016; pp. 770–778.