agriculture

MDPI

Article

Deep Learning-Enabled Nondestructive Prediction of Moisture Content in Post-Heading Paddy Rice (*Oryza sativa* L.) Using Near-Infrared Spectroscopy

Ha-Eun Yang [1,2], Hong-Gu Lee [1], Jeong-Eun Lee [1], Jeong-Yong Shin [1], Wan-Gyu Sang [3], Byoung-Kwan Cho [4] and Changyeun Mo [1,5,*]

1 Interdisciplinary Program in Smart Agriculture, Kangwon National University, Chuncheon 24341, Republic of Korea; haeun001209@kangwon.ac.kr (H.-E.Y.); hgl@kangwon.ac.kr (H.-G.L.); wjddms523@kangwon.ac.kr (J.-E.L.); kvhffh@kangwon.ac.kr (J.-Y.S.)
2 Smart Agriculture Division, Korea Agriculture Technology Promotion Agency, Iksan 54667, Republic of Korea
3 Department of Crop Production and Physiology, National Institute of Crop Science, Rural Development Administration, Wanju 55365, Republic of Korea; wg_sang@korea.kr
4 Department of Smart Agriculture Systems, College of Agricultural and Life Science, Chungnam National University, Daejeon 34134, Republic of Korea; chobk@cnu.ac.kr
5 Department of Biosystems Engineering, College of Agriculture and Life Sciences, Kangwon National University, Chuncheon 24341, Republic of Korea
* Correspondence: cymoh100@kangwon.ac.kr; Tel.: +82-33-250-6494

Abstract

Rapid non-destructive evaluation of the moisture content of freshly harvested paddy rice in the field is essential for determining the optimal harvest timing, ensuring high-quality rice production and energy savings. This study developed a non-destructive prediction model for the moisture content of paddy rice using near-infrared (NIR) spectroscopy combined with machine learning and deep learning techniques. Rice samples were collected weekly during the ripening period after heading, and NIR reflectance spectra were acquired in the range of 950–2200 nm. Seven spectral preprocessing techniques were applied; and the prediction models developed, using partial least squares regression, support vector regression, deep neural network, and one-dimensional convolutional neural networks (1D-CNNs) based on VGGNet and EfficientNet architectures. Among these, the EfficientNet-based 1D-CNN combined with Savitzky–Golay 1st order derivative preprocessing showed the highest performance, achieving an R_p^2 of 0.999 and an $RMSE_P$ of 0.001 (Friedman test, $p < 0.001$; Kendall's $W = 0.97$), significantly outperforming previous traditional machine learning models. The results demonstrate that the proposed prediction model enables highly accurate estimation of moisture content in freshly harvested paddy rice without requiring drying or milling. The proposed approach can be implemented across various agricultural operations, enabling optimal harvest timing, quality control during storage, energy efficient drying, and real-time monitoring via on-combine sensor systems.

Keywords: paddy rice; moisture content prediction; near-infrared spectroscopy (NIRS); deep learning; 1D convolutional neural network (1D-CNN); harvest timing decision

Check for updates

Academic Editor: Jerry Gao

Received: 16 December 2025
Revised: 10 March 2026
Accepted: 12 March 2026
Published: 17 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Rice (*Oryza sativa* L.) is a major staple crop that provides essential calories and nutrients to more than half of the world's population [1]. In the context of increasing food demand driven by climate change, natural disasters, and population growth, the adoption of

precision agriculture has gained considerable attention for ensuring food security and maintaining crop quality [2]. In this context, the development of technologies capable of quantitatively evaluating and managing rice quality is recognized as a critical task for improving agricultural productivity and economic efficiency.

The Moisture content (MC) is a key indicator that strongly influences harvest timing, post-harvest drying quality, and overall economic profitability of rice. After heading, MC gradually decreases and typically reaches 22–24% (wet basis, w.b.) at the optimal harvest stage, which is widely regarded as a critical threshold for precision harvesting [3]. When MC exceeds approximately 22%, drying costs increase substantially, whereas MC below this level elevates the risk of kernel breakage during milling, leading to quality deterioration and economic loss [4]. Moreover, a significant portion of the total energy consumed during rice processing is concentrated in the drying stage, highlighting the importance of appropriate MC management for reducing drying costs and maintaining product quality [5–7]. Nevertheless, in actual field practice, harvest timing is often determined based on the number of days after heading, a method that relies heavily on farmers' experience. This approach fails to capture real-time variations in MC, underscoring the need for more objective and accurate technologies to support harvest timing decisions.

Traditionally, grain moisture content has been measured using the oven-drying method or electrical resistance-based methods. While the oven-drying method serves as a reference technique with high accuracy, it is time-consuming and unsuitable for real-time field applications. Electrical resistance methods allow faster measurements but suffer from lower precision and sensitivity to measurement conditions [8]. These limitations restrict their direct use for real-time moisture management and harvest decision-making in the field.

To overcome these constraints, various non-destructive moisture-sensing technologies have been proposed, including microwave sensors, capacitive sensors, image-based analysis, and near-infrared (NIR) spectroscopy [9–11]. However, microwave and capacitive sensors are highly influenced by sample characteristics such as particle size and bulk density, while image-based methods are sensitive to illumination conditions and environmental variability, and are limited in assessing internal moisture [12]. Consequently, there remains a strong demand for more stable and quantitative approaches to moisture measurement.

Among non-destructive techniques, NIR spectroscopy has emerged as a promising alternative due to its ability to rapidly acquire moisture-related information without damaging the sample. NIR spectroscopy enables simultaneous analysis of multiple chemical and physical components in agricultural products and has been widely adopted as an environmentally friendly technique for quality assessment [13–15]. In the NIR region (800–2500 nm), absorption bands arise from overtone and combination vibrations of molecular functional groups, primarily associated with O–H, C–H, N–H, and S–H bonds. These spectral characteristics allow simultaneous quantification of key components such as moisture (O–H), lipids (C–H), and proteins (N–H, S–H) in agricultural samples [16]. In addition, NIR spectral information has been shown to be closely associated with plant physiological properties such as internal cell structure and water status, and has therefore been widely applied in studies monitoring crop growth conditions in rice cultivation [17–21].

Because NIR spectral data are high-dimensional and exhibit nonlinear characteristics, various machine learning (ML) techniques have been applied to interpret spectral information effectively. Partial least squares regression (PLSR), support vector regression (SVR), PLS-CARS, and artificial neural networks (ANNs) have been widely used for agricultural quality prediction, particularly for moisture content estimation [22–24]. These methods have contributed to quantitative modeling of relationships between spectral features and target variables.

More recently, deep learning (DL) approaches, including convolutional neural networks (CNNs), have been increasingly applied to NIR spectral analysis and have demonstrated strong potential for effectively learning complex spectral patterns due to their superior nonlinear feature representation capabilities compared with conventional machine learning methods [25–27].

However, most existing NIR-based studies on rice have primarily focused on post-harvest grain samples such as brown rice, polished rice, or powdered rice flour. These studies mainly aim to evaluate compositional traits under relatively stable physiological conditions after crop maturation, such as predicting static components including protein content [28–30]. In contrast, paddy rice during the post-heading growth stage still contains the husk, and its physiological properties continuously change as grain development and maturation progress. These characteristics introduce spectral interference and dynamic moisture variations, making accurate moisture prediction more challenging.

Furthermore, many previous studies have analyzed samples collected at a single maturity stage, which limits their applicability to practical agricultural operations such as harvest timing decisions. In real agricultural environments, the moisture content of rice gradually decreases from heading to harvest, and accurately monitoring these dynamic changes is essential for determining the optimal harvest timing. Nevertheless, studies that continuously predict the moisture content of paddy rice throughout the growth period remain limited.

To address this research gap, this study proposes a non-destructive moisture prediction approach for paddy rice by combining near-infrared (NIR) spectroscopy with machine learning and deep learning techniques. In this study, moisture prediction models were constructed using NIR spectral data obtained from paddy rice samples collected throughout the growth period from heading to harvest, and the predictive performances of various machine learning and deep learning algorithms were systematically compared.

The main contributions of this study are summarized as follows:

1. Dynamic moisture prediction throughout the growth period
   Unlike previous studies that mainly focused on post-harvest grain samples, this study analyzes the NIR spectral characteristics of paddy rice samples collected after heading and proposes an approach to continuously predict the moisture content of rice throughout the growth period from heading to harvest.
2. Systematic comparison of machine learning and deep learning models
   Moisture prediction models were constructed using conventional machine learning models (PLSR and SVR) as well as deep learning models (DNN and 1D-CNN), and their predictive performances were systematically compared to identify suitable models for NIR-based moisture prediction.
3. Development of a lightweight deep learning model for practical field applications
   A lightweight deep learning architecture trained using freshly collected paddy rice samples obtained weekly from heading to the optimal maturity stage is proposed, demonstrating the potential for practical non-destructive moisture prediction in real agricultural environments.

By incorporating the moisture dynamics of paddy rice during the growth period, the proposed prediction model can support harvest timing decisions and contribute to reducing drying costs and post-harvest quality losses.

2. Materials and Methods

2.1. Materials

In this study, rice (*Oryza sativa* L. subsp. *japonica*) samples of the cultivar Sindongjin were used. The samples were harvested in 2023 by the National Institute of Crop Science

in the Republic of Korea (Figure 1). The samples were not collected from a single plant but were instead obtained from multiple individual plants within the same experimental field in order to reflect field-level variability. Sindongjin is a mid-late maturing cultivar primarily cultivated in Jeollabuk-do Province and is one of the most widely grown rice varieties in Korea [31].



Figure 1. Map of the sampling site. The sampling site is located at the National Institute of Crop Science (NICS) in Wanju, Republic of Korea (35°50′ N, 127°02′ E).

Typically, the optimal harvest period for mid-late maturing rice cultivars is 55 to 60 days after heading. For Sindongjin, heading occurs around mid-August, and the optimal harvest time falls in early October [32].

To ensure a wide range of moisture content during the growth and harvest periods, rice samples were harvested at five different time points: the 5th (21 September), 6th (28 September), and 7th (4 October) weeks after heading (representing the growth stage), and the 8th (12 October) and 9th (19 October) weeks after heading (representing the harvest stage). Harvested samples were sealed in zipper bags after field collection and transported to the laboratory to minimize moisture loss. Samples were stored in sealed conditions for approximately 1–2 days and were opened only immediately before initial weighing and NIR spectral acquisition. Each sample was placed in a Petri dish (diameter: 55 mm, height: 15 mm), with 20 samples prepared for each week, resulting in a total of 100 samples. To ensure consistency during NIR measurements, the samples were evenly distributed in individual Petri dishes. Representative samples from each week are shown in Figure 2.



Figure 2. Paddy rice samples collected near the harvest stage after grain formation following heading.

2.2. Near-Infrared (NIR) Spectral Measurement

2.2.1. NIR Spectral Measurement System

The near-infrared spectral measurement system used to acquire the NIR reflectance spectra of paddy rice is shown in Figure 3. The system consisted of a near-infrared spectrometer (SM304, Korea Spectral Products, Seoul, Republic of Korea), a 100 W tungsten-halogen lamp as the light source (ASBN-W100, KSP, Seoul, Republic of Korea), and a sample rotation unit driven by a stepper motor (28BYJ-48-5V, FSXSEMI, Shanghai, China).

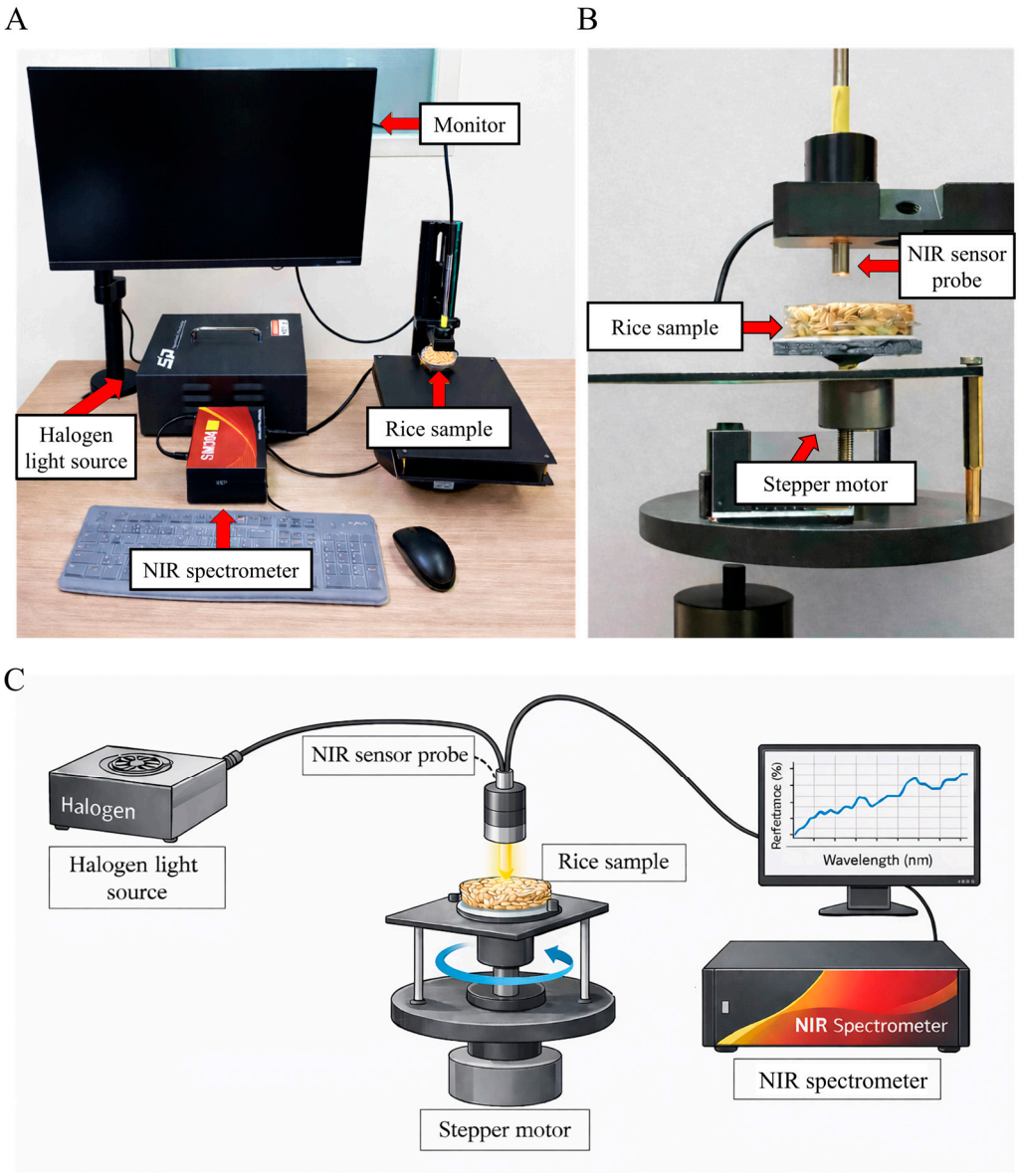

Figure 3. (**A**) Configuration of the NIR spectroscopic measurement system; (**B**) sample rotation unit; (**C**) schematic diagram of the near-infrared (NIR) spectroscopy system.

During spectral acquisition, light was delivered perpendicularly (at 90°) onto the surface of the rice samples using an optical fiber. The distance from the probe to the sample was maintained at 10 mm. The Petri dish containing the paddy rice was rotated by 36° increments using the stepper motor, and reflectance spectra were acquired in a stationary manner at each position to capture spatial variability across the sample surface.

2.2.2. Acquisition of Near-Infrared Spectral Data

The reflectance spectra of the paddy rice samples were measured using a near-infrared spectrometer with a spectral resolution of 3.8 nm and an exposure time set to 80 ms. Spectral data were acquired over the wavelength range of 900–2500 nm, and only the wavelength region from 900 to 2200 nm, where stable signal quality was observed, was used for subsequent analysis. Noisy spectral regions at both ends were excluded from the analysis. To minimize measurement errors and to capture representative signals from the entire surface area of each sample, ten spectra were collected from different positions on each sample. Each spectrum was acquired by rotating the Petri dish and performing close-range measurements at different spatial positions, such that the illuminated regions and the rice kernels contributing to the measured signal varied across measurements, even within the same sample. This approach was intended to minimize repeated measurement of identical rice kernel compositions and to reduce potential data overlap between the training and prediction sets.

To obtain a wide range of moisture content data for model training, 20 paddy rice samples were collected and measured over a period of 5 weeks. These samples were naturally air-dried at room temperature (25 °C) to gradually reduce their moisture content. Spectra were measured twice per day over three-day intervals at a total of 6 different drying stages for each sample. Through this systematic process, a total of 6000 spectra were acquired according to the following formula: 6000 spectra = 20 samples × 5 weeks × 10 measurements per sample × 6 drying stages.

This methodology ensured that the dataset represented a broad range of moisture content and captured spatial variability, thereby enhancing the generalizability of the developed models.

The reflectance values of the measured samples were corrected using both white and dark references. A 50% diffuse reflectance standard (Labsphere Inc., North Sutton, NH, USA) was used as the white reference, and the dark reference was obtained by completely blocking the light path during measurement. The reflectance was calculated using Equation (1).

$$R_{cal} = \frac{R_s - R_{dark}}{2.0R_{white} - R_{dark}} \quad (1)$$

Here, R_{white}, R_{dark}, and R_s denote the reflectance values of the white reference, dark reference, and sample, respectively. The correction factor of 2.0 represents the scaling required to convert the measured signal of the 50% diffuse white reference to an equivalent 100% reflectance level.

2.3. Moisture Content Analysis Using the Oven-Drying Method

The moisture content of the paddy rice samples was determined using the oven-drying method. Before acquiring spectral data, the initial weight of each sample was measured. Following spectral acquisition, the samples were oven-dried at 135 °C for 24 h to determine their dry weight [33]. Moisture content on a wet basis was calculated using the initial and final weights, as shown in Equation (2):

$$MC = \frac{W_w}{W_w + W_d} \times 100 \quad (2)$$

Here, W_w is the weight of water obtained by subtracting the dry weight from the initial sample weight, and W_d is the dry weight of the sample.

2.4. Spectral Data Preprocessing

To eliminate spectral distortions caused by external environmental factors, light scattering, and noise, and to improve overall model performance, spectral preprocessing techniques were applied. To compare model performance with and without preprocessing, the following preprocessing methods were evaluated: Savitzky–Golay 1st order derivative, Savitzky–Golay second derivative, maximum normalization, mean normalization, range normalization, standard normal variate (SNV), and multiplicative scatter correction (MSC). To ensure the reproducibility of the Savitzky–Golay derivative preprocessing, the specific configurations, including window size and polynomial order, are detailed in Appendix A.4 (Table A4). The performance of each moisture prediction model was assessed accordingly. Spectral preprocessing was performed using Unscrambler v9.7 (CAMO Process AS, Oslo, Norway).

2.5. Model Development for Moisture Prediction

2.5.1. Data Processing and Partitioning

For model development, the entire dataset (i.e., a total of 6000 spectra obtained through repeated measurements across different drying stages from 100 physical samples) was split into calibration (training) set and prediction (test) set at a ratio of 90:10. At this stage, the continuous moisture content values were divided into predefined intervals, and stratified sampling was applied to ensure similar moisture content distributions between the calibration and prediction datasets. Ninety percent of the data was used for model training and cross-validation, while the remaining 10% served as an independent prediction set that was not involved in model development, used to evaluate final model performance.

To ensure generalization and prevent overfitting during model training, 10-fold cross-validation was employed [34]. The calibration dataset was divided into 10 equal-sized subsets using stratified sampling based on moisture content to ensure a similar moisture distribution across all folds. In each iteration, one subset was used for validation while the remaining nine were used for model calibration. This process was repeated ten times, and the average performance across the ten folds was used as the final result.

2.5.2. Development of Moisture Prediction Models for Post-Heading Paddy Rice

Moisture content prediction models were developed for paddy rice samples harvested from the 5th to the 9th week after heading (i.e., from post-heading to the optimal harvest stage). Two deep learning models, the DNN and the 1D-CNN, and two machine learning models, PLSR, and SVR, were constructed and compared. In all models, the input data consisted of NIR spectra, and the output data were the corresponding moisture content values. A total of 6000 data points were used for model development. For the deep learning models (DNN and 1D-CNN), key hyperparameters such as learning rate, batch size, and dropout rate were optimized through a systematic manual tuning process. Various candidate ranges were iteratively evaluated, and the final configurations were selected based on the lowest validation RMSE to ensure optimal generalization and stability. Detailed information regarding the optimized hyperparameters and specific configurations for all models, including the Savitzky–Golay preprocessing parameters, is provided in Appendix A (Tables A1–A4).

PLSR Model

PLSR is a classical linear modeling technique widely used for spectral data analysis, capable of capturing linear relationships between input and output variables [35,36]. It has been extensively applied in industries where spectroscopy is utilized for regression

modeling [37]. In this study, PLSR was used to model the linear correlation between the spectral data and the moisture content of the rice samples.

SVR Model

SVR is a machine learning algorithm capable of modeling nonlinear relationships between input features and target variables by projecting data into a high-dimensional space [38,39]. SVR performs reliably even on small datasets and is well suited for capturing nonlinear patterns when using kernel functions such as the Radial Basis Function (RBF) [38,40]. In this study, SVR was employed to capture the nonlinear relationship between spectral data and moisture content, with the RBF kernel used to project the input data into a higher-dimensional feature space. The key hyperparameters of the SVR model, including the penalty parameter C, epsilon (ϵ) in the loss function, and gamma (γ) of the RBF kernel, were optimized through a grid-search procedure.

DNN Model

The DNN is a deep learning model capable of learning complex and nonlinear patterns in data. It has significantly advanced state-of-the-art performance across a wide range of applications, including agriculture [41]. DNNs are based on a multilayer neural network architecture and are effective in capturing global patterns from continuous data such as spectral information [42]. In this study, a DNN model was employed to learn the global relationship between near-infrared spectral data and the moisture content of paddy rice after heading.

To predict the moisture content of paddy rice during the post-heading period, a multilayer perceptron (MLP) architecture, which is commonly used in deep learning, was adopted and optimized to fit the structure of the spectral data. Figure 4 illustrates the architecture of the DNN model, which is based on a Multi-Layer Perceptron (MLP) structure designed to accommodate the characteristics of the spectral data.

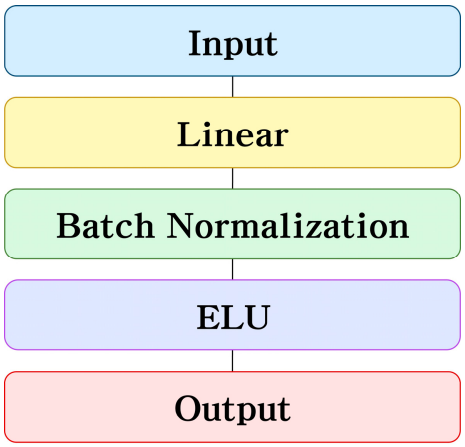

Figure 4. Architecture of the proposed DNN model based on MLP structure.

The input shape was set to (32, 304), representing 32 samples per batch and 304 spectral bands. The model consisted of six fully connected layers, including five hidden layers and one output layer. Batch normalization and an ELU (Exponential Linear Unit) activation function were applied after each layer to introduce nonlinearity. The number of units in the hidden layers gradually decreased as follows: 200, 100, 50, and 25, eventually producing a single output value representing the predicted moisture content. The model was trained using the Adam optimization algorithm, and the loss function was defined as the Root Mean Squared Error (RMSE). The batch size was set to 32, and training was performed

for a total of 100 epochs. An early stopping mechanism with a patience of 20 epochs was implemented to prevent overfitting (Table 1).

Table 1. Hyperparameters used in DNN.

Hyperparameter	Value
Learning Rate	0.001
Batch Size	32
Number of Epochs	100
Hidden Layer	5
Weight Decay	0.0000001
Loss Function RMSE	(Root Mean Square Error)
Optimizer	Adam

1D-CNN Model

A deep learning model based on 1D-CNN was developed to predict moisture content of paddy rice from post-heading to the optimal harvest period. 1D-CNN is a variant of convolutional neural networks adapted for sequential or continuous data, originally designed for image processing but modified for one-dimensional inputs. By applying convolution operations, 1D-CNN effectively captures local features within input data, making it suitable for spectral data that consist of continuous values [43,44]. This study aimed to learn local patterns between spectral data and moisture content using 1D-CNN.

Two different CNN architectures with distinct feature extraction approaches were selected and compared for this purpose. The first architecture was based on VGG-19, widely utilized in agricultural product quality evaluation research [45–47]. VGG-19 is known for its hierarchical feature learning through deep convolutional layers, demonstrating excellent performance in image classification tasks [48]. The simple and intuitive structure of VGG-19 also facilitates interpretability in spectral analysis. Therefore, VGG-19 was adopted as a baseline model for systematic learning of continuous patterns in spectral data. However, despite its depth and simplicity, the VGG-based model expands network depth and width in an unstructured manner, leading to limited efficiency improvements relative to increased computational complexity.

To overcome these limitations, a lightweight model with enhanced computational efficiency, EfficientNet architecture, was selected as the second structure. EfficientNet is a modern architecture optimized to effectively capture information at various scales by uniformly scaling network depth, width, and resolution through a compound scaling method [40]. It was chosen as a comparative model capable of more efficiently and precisely capturing the complex features of spectral data. Both architectures were modified to fit the one-dimensional spectral data structure, and their predictive performances were compared to identify the optimal model.

Figure 5 illustrates the VGG-19-based 1D-CNN architecture designed for one-dimensional spectral data. The input shape is (32, 1, 304), representing 32 samples per batch, 1 channel, and 304 spectral bands. Each convolutional layer operates within blocks with a fixed kernel size of 3. Max pooling with a size of 2 is used for downsampling. Exponential Linear Unit (ELU) activation functions were applied in hidden layers. Dropout was set to 0.2 to prevent overfitting. The model was trained using the Adam optimizer and RMSE loss function. The batch size was set to 32, with 100 epochs of training, and early stopping was applied with a patience of 20 epochs (Table 2). The total number of trainable parameters for the VGG-19-based 1D-CNN model is approximately 1.66 M.

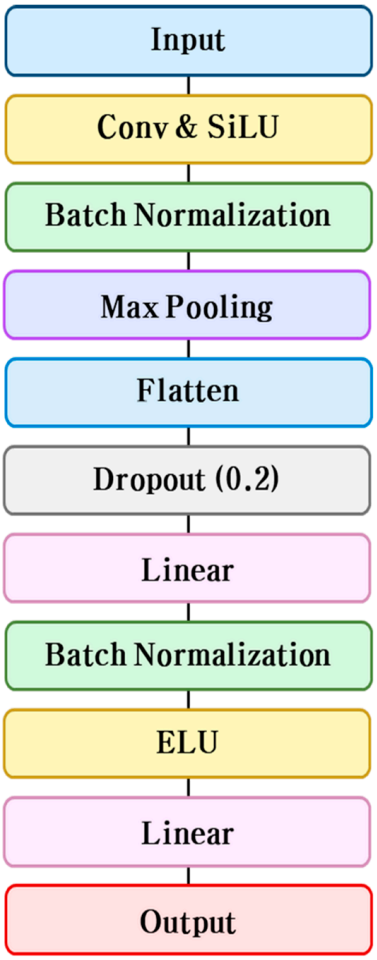

Figure 5. Architecture of 1D-CNN model based on VGG-19.

Table 2. Hyperparameters used in 1D-CNN.

Hyperparameter	VGG 19	EfficientNet
Learning Rate	0.001	0.0005
Batch Size	32	32
Number of Epochs	100	100
Hidden Layer	15	8
Weight Decay	0.0000001	0
Loss Function RMSE	RMSE	RMSE
Optimizer	Adam	Adam

EfficientNet improves both efficiency and accuracy by balancing depth, width, and resolution through the compound scaling method [49]. The lightweight block structure and scaling strategy of EfficientNet's most basic model, EfficientNet-B0, were adapted by replacing the original 2D convolutions with 1D convolutions and reconstructing the MBConv blocks to fit one-dimensional spectral data. Figure 6 presents the EfficientNet-based 1D-CNN architecture tailored for one-dimensional spectral data. The model consists of eight main layers, including seven MBConv blocks and one head layer. Each block incorporates batch normalization and SiLU activation functions, and skip connections are employed to minimize information loss. The input shape remains (32, 1, 304). Training parameters were consistent with the VGG-19-based model, using the Adam optimizer, RMSE loss, batch size of 32, 100 epochs, and early stopping with patience set to 20 (Table 2).

The EfficientNet-based 1D-CNN model is a more lightweight architecture, consisting of approximately 0.64 M trainable parameters.

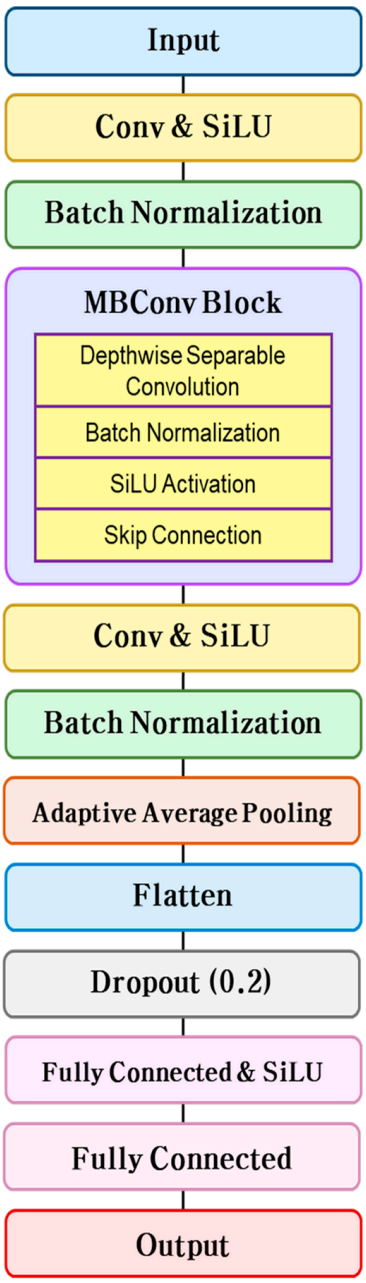

Figure 6. Architecture of 1D-CNN model based on EfficientNet.

2.5.3. Performance Evaluation of Paddy Rice Moisture Content Prediction Models

The performance of the paddy rice moisture content prediction models was evaluated by comparing the predicted values with the actual measured values during the calibration, cross-validation, and prediction phases.

To assess model performance, the coefficient of determination for calibration (R_c^2) and root mean square error of calibration ($RMSE_C$) were used to indicate how well the model fit the training data. The model's generalization ability was evaluated using the coefficient of determination for validation (R_v^2) and root mean square error of validation ($RMSE_V$), obtained through cross-validation. Most importantly, the model's predictive ability on new, unknown data was assessed using the coefficient of determination for prediction (R_p^2) and root mean square error of prediction ($RMSE_P$). These metrics are the key indicator for

demonstrating the model's real-world applicability. The optimal model was selected based on the highest R_p^2 and the lowest $RMSE_P$.

R_c^2, R_v^2, and R_p^2 were used to represent the linearity between measured and predicted values during calibration, cross-validation, and for unknown samples not used in model development, respectively. Prediction accuracy was evaluated using $RMSE_C$, $RMSE_V$, and $RMSE_P$.

The equations for R^2 and RMSE are defined as follows and RMSE values are expressed as a percentage (%):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2}, \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad (4)$$

where y_i is the reference moisture content of the paddy rice sample (measured value), $\hat{y}_i$ is the predicted moisture content, $\overline{y}_i$ is the mean of the reference values, and n is the number of samples.

2.5.4. Statistical Analysis

To rigorously evaluate whether the differences in predictive performance among the models were statistically significant, a multi-step statistical analysis was performed based on the RMSE values obtained from the 10-fold cross-validation.

First, the Friedman test, a non-parametric alternative to the repeated measures ANOVA, was employed to determine if there were significant differences in performance across the five models (PLSR, SVR, DNN, 1D-CNN based on VGG-19, 1D-CNN based on EfficientNet) [50]. This non-parametric approach was selected as it does not assume a normal distribution of the fold-wise results, providing a more robust evaluation for the 10-fold cross-validation datasets.

Upon identifying significant differences ($p < 0.05$), the Conover post hoc test was conducted for pairwise comparisons to identify specific model pairs with significant performance gaps [51]. To minimize the risk of Type I errors arising from multiple comparisons, the Bonferroni correction was applied to adjust the p-values.

Furthermore, Kendall's coefficient of concordance (W) was calculated to assess the effect size and the degree of consistency in the model rankings across the 10-fold cross-validation. A W value close to 1 indicates that the models' performance rankings are highly consistent across different data splits. All statistical computations and the generation of the significance heatmap were performed using the scipy.stats and scikit-posthocs libraries in Python (version 3.11), with the statistical significance level set at $\alpha = 0.05$.

3. Results

3.1. Analysis of Paddy Rice Moisture Content According to Weeks After Heading

The initial moisture content analysis results of paddy rice samples, from post-heading to the optimal harvest period, are presented in Table 3 and Figure 7A. The initial moisture content of the entire sample set ranged from 21.60% to 40.71%. As time elapsed after heading, the moisture content of the paddy rice gradually decreased, showing a declining trend as the samples approached the optimal harvest period. During weeks 8 to 9 after heading, which are considered the optimal harvest period for the Sindongjin cultivar, the average moisture content was measured at approximately 21% to 23%. This range corresponds well with the reported optimal moisture content for harvest, which is typically between 21% and 24% [11].

Table 3. Initial moisture content (%) of 'Sindongjin' rice. (N = 100).

Weeks After Heading	5 Weeks	6 Weeks	7 Weeks	8 Weeks	9 Weeks
Number of sample	20	20	20	20	20
Average moisture content (%)	35.25	30.33	24.70	23.68	21.19
Minimum moisture content (%)	30.01	26.45	23.81	23.27	20.35
Maximum moisture content (%)	40.71	32.15	25.48	24.44	21.60
Standard deviation	1.77	1.19	0.48	1.71	0.34

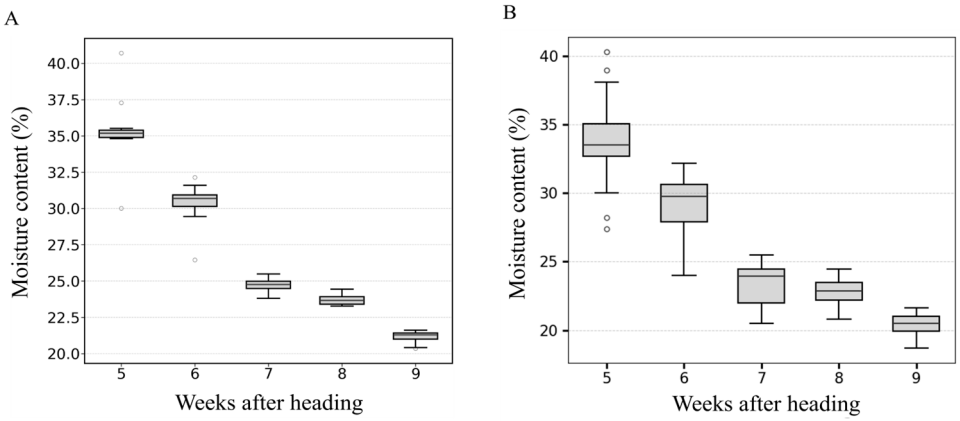

Figure 7. Moisture content distribution in paddy rice from heading to optimal harvest: (**A**) observed initial moisture content; (**B**) extended moisture content range obtained through repeated measurements during natural air-drying for model development.

Specifically, the moisture content decreased by about 14%, from 35.25% at week 5 to 21.19% at week 9. The rate of moisture loss was relatively rapid between weeks 5 and 7, and then slowed down from weeks 8 to 9. This pattern aligns with the general physiological characteristic of rice, where the rate of moisture evaporation diminishes during the later stages after heading [52].

Additionally, analysis of the standard deviation for each week indicated that, to validate model performance across a wider moisture content range, additional measurements were conducted by naturally drying the paddy rice samples, as described in Section 2.2.2. As shown in Figure 7B, the extended moisture content distribution ranged broadly from 17% to above 40%, contributing to enhanced model generalizability during training.

3.2. Analysis of NIR Spectral Characteristics of Paddy Rice According to Weeks After Heading

Near-infrared (NIR) reflectance spectra were measured for a total of 100 paddy rice samples collected from heading to harvest maturity. Figure 8 presents the raw spectra and the average spectra for each week after heading.

Throughout the period from heading to harvest maturity, the spectral profiles of paddy rice samples maintained similar patterns within the wavelength range of 950–2150 nm. As the weeks progressed after heading and the moisture content of the samples decreased, an increasing trend in reflectance was observed.

In particular, the spectral reflectance tended to increase overall as the samples progressed from early post-heading (weeks 5 and 6) towards harvest maturity. This phenomenon is attributed to the higher absorption of near-infrared light by samples with higher moisture content, resulting in lower reflectance, especially in wavelength regions where water exhibits strong absorption. Therefore, higher moisture content in the sample leads to greater light absorption in related wavelength regions, thereby reducing reflectance.

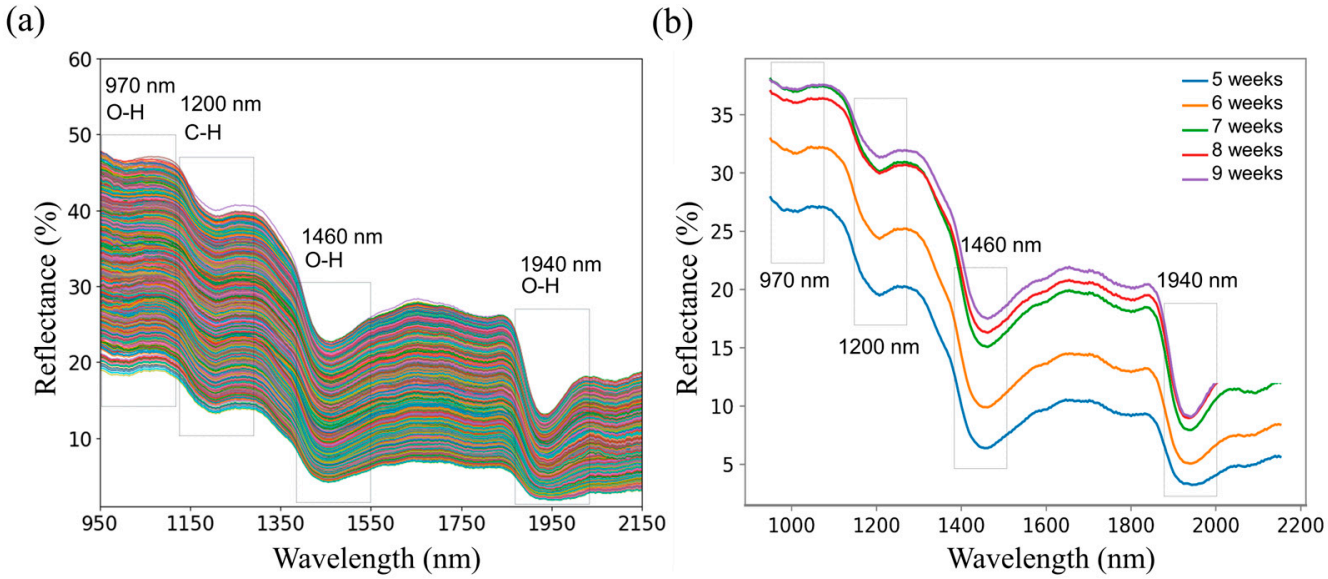

Figure 8. Spectra of Paddy Samples from Heading to Optimal Harvest Time: (**a**) Raw spectra, (**b**) weekly average spectra.

In the NIR region, absorption peaks around 1460 and 1940 nm are associated with the bending and stretching vibrations of the O–H bonds in water [53]. The absorption peak near 970 nm is primarily attributed to a combination overtone of O–H bond vibrations from carbohydrates and water [54]. Absorption around 1200 nm is due to the second overtone of C–H stretching in carbohydrates [55], while absorption near 1450 nm is mainly related to the first overtone of O–H stretching, a primary water absorption band [54,55]. The absorption peak near 1940 nm corresponds to the stretching vibration of water's O–H bonds [56].

Accordingly, the increasing reflectance observed in these spectral valleys as the samples approached harvest maturity is considered to reflect a decrease in moisture content in the paddy rice samples.

Figure 9 illustrates the reflectance spectra of paddy rice after applying optimal preprocessing techniques (1st and 2nd order derivatives), visualized by different moisture content ranges. The samples were categorized and compared across six moisture content intervals: lowest moisture content range after air-drying (17–20%), optimal harvest stage (21–24%), and maturation stages (25–28%, 29–33%, and 34–37%).

When 1st and 2nd order derivative preprocessing was applied, spectral differences became clearly distinguishable across moisture content levels, particularly in the regions near 1150, 1900, and 2000 nm, which are known to be sensitive to water content. Among these, the spectral differences around 1900 nm were the most prominent. This region corresponds to the fundamental vibration band of water (O–H stretch + bend), which exhibits strong absorption and is largely unaffected by other components. Therefore, the observed sensitivity in this region is likely due to its specificity to water absorption.

In contrast, the 1450 nm region, which is commonly known as a representative water absorption band, exhibited relatively less distinct separation between moisture levels. This is likely because the paddy rice samples used in this study were measured with husks intact. The spectral region around 1450 nm may be influenced not only by water but also by overlapping absorption from other components in plant tissues, such as cellulose, starch, and proteins, leading to potential spectral interference.

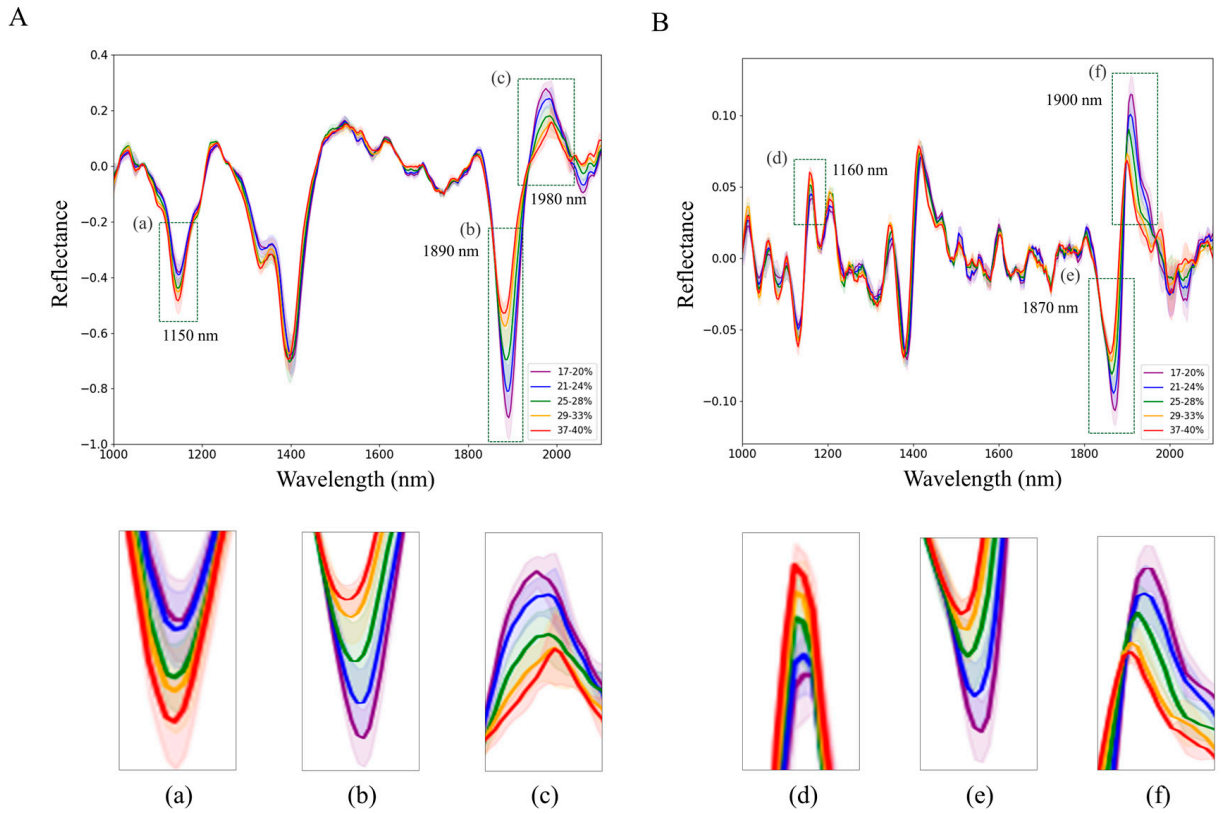

Figure 9. Mean and standard deviation of reflectance spectra of paddy rice samples according to moisture content. (**A**) First-derivative preprocessed spectra; (**B**) second-derivative preprocessed spectra Distinct spectral peaks and valleys associated with moisture content at (**a**) 1150 nm; (**b**) 1890 nm; (**c**) 1980 nm; (**d**) 1160 nm; (**e**) 1870 nm; and (**f**) 1900 nm.

3.3. Results of PLSR Model Development

A total of eight partial least squares regression (PLSR) models were developed to predict the moisture content of paddy rice samples collected from heading to harvest maturity. The performance of these models, developed using various spectral preprocessing techniques, is summarized in Table 4.

Table 4. PLSR model performance for predicting the moisture content of paddy rice.

Model Type	Preprocessing	Calibration		Validation		Prediction		F*
		R_c^2	$RMSE_C$	R_v^2	$RMSE_V$	R_p^2	$RMSE_P$	
PLSR	Raw	0.941	0.012	0.942	0.012	0.920	0.028	7
	Mean Normalization	0.923	0.014	0.923	0.014	0.921	0.014	3
	Range Normalization	0.923	0.014	0.922	0.014	0.911	0.112	4
	Maximum Normalization	0.937	0.013	0.936	0.013	0.932	0.016	6
	1st order D* (gap size = 22.8 nm)	0.940	0.013	0.940	0.013	0.941	0.012	6
	2nd order D* (gap size = 7.6 nm)	0.936	0.013	0.936	0.013	0.937	0.013	5
	MSC	0.914	0.015	0.914	0.015	0.912	0.015	3
	SNV	0.927	0.014	0.927	0.014	0.926	0.014	4

Note. D*: Derivative, F*: Factor.

Among the PLSR models developed for predicting moisture content in paddy rice, the model without any spectral preprocessing showed the highest performance during calibration and cross-validation, with an R_c^2 and $RMSE_C$ of 0.941 and 0.012, respectively, and an R_v^2 of 0.942 and an $RMSE_V$ of 0.012. However, when tested on an independent (unseen) sample set, the model achieved an R_p^2 of 0.920 and an $RMSE_P$ of 0.028. Conversely, among

the models developed using various preprocessing techniques, the one employing Savitzky–Golay 1st order derivative preprocessing demonstrated the overall best performance on the independent prediction set, achieving an R_p^2 of 0.941 and an $RMSE_P$ of 0.012.

Figure 10 presents the calibration and prediction results of the optimal PLSR model using 1st order derivative preprocessing. Spectral preprocessing contributed to minimizing the gap between calibration and prediction performance, leading to enhanced model robustness and generalization capability.

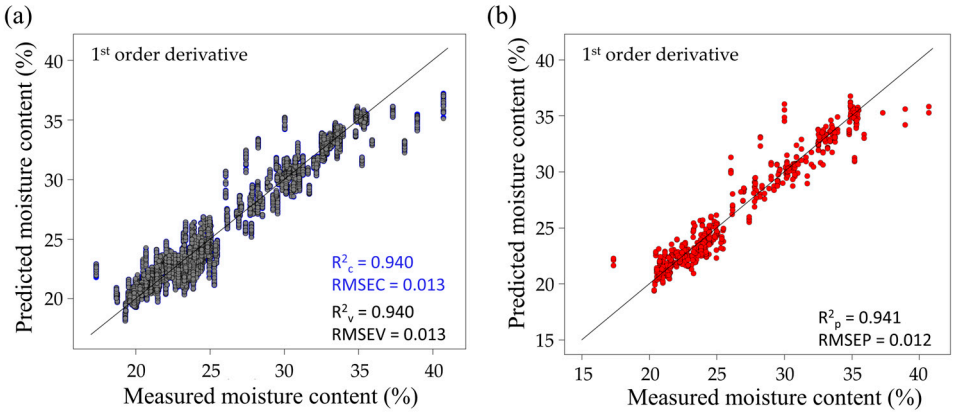

Figure 10. Optimal performances of PLSR models for paddy rice moisture content prediction: (**a**) calibration with 1st-order derivative data, (**b**) prediction with 1st-order derivative data.

Figure 11 presents the regression coefficient plot of the PLSR model utilizing Savitzky–Golay 1st order derivative preprocessing, which illustrates the contribution and relative importance of each wavelength variable to the prediction of moisture content [57]. Wavelengths corresponding to either positive or negative regression coefficients that surpassed the threshold defined by the standard deviation (dashed line) were regarded as significant for capturing variations in moisture-related parameters [58]. Accordingly, significant positive peaks associated with moisture prediction were observed around 980, 1150, 1400, 1900, and 2000 nm, while significant negative peaks were identified near 1120, 1370, and 1930 nm.

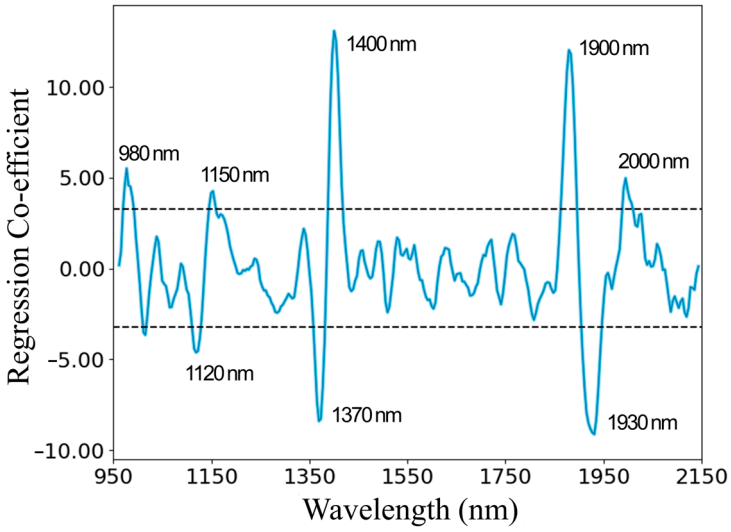

Figure 11. Regression coefficient plot of the optimal PLSR models for predicting moisture content in paddy rice after heading stage. The dashed lines indicate the threshold based on the standard deviation of the regression coefficients.

Notably, the effective wavelengths identified in the regression coefficient plot of the Savitzky–Golay 1st order derivative model (Figure 11) at 1150 and 1900 nm correspond to the wavelengths that showed clear reflectance variations according to moisture content in Figure 9. These findings suggest that the spectral regions near 1150 and 1900 nm played a significant role in moisture content detection of paddy rice during the post-heading to harvest period.

These results align with previous studies that reported the regions around 1200, 1400, 1460, and 1940 nm as critical wavelengths for identifying moisture content [53,59,60].

3.4. Results of SVR Model Development

SVR models were developed to predict the moisture content of paddy rice, and the results obtained using various spectral preprocessing techniques are summarized in Table 5.

Table 5. SVR model performance for predicting the moisture content of paddy rice.

Model Type	Preprocessing	Calibration		Validation		Prediction		Kernel Type
		R_c^2	$RMSE_C$	R_v^2	$RMSE_V$	R_p^2	$RMSE_P$	
SVR	Raw	0.930	0.014	0.929	0.014	0.928	0.014	RBF*
	Mean Normalization	0.945	0.012	0.944	0.012	0.938	0.013	RBF
	Range Normalization	0.941	0.013	0.939	0.013	0.934	0.014	RBF
	Maximum Normalization	0.932	0.014	0.931	0.013	0.929	0.014	RBF
	1st order D* (gap size = 3.8 nm)	0.976	0.008	0.974	0.008	0.978	0.008	RBF
	2nd order D* (gap size = 7.6 nm)	0.972	0.009	0.969	0.009	0.974	0.008	RBF
	MSC	0.955	0.011	0.954	0.011	0.940	0.013	RBF
	SNV	0.956	0.011	0.954	0.011	0.940	0.013	RBF

Note. D*: Derivative, RBF*: Radial Basis Function.

Among the SVR models for predicting the moisture content of paddy rice, the model with Savitzky–Golay 1st order derivative preprocessing exhibited the highest predictive performance. This calibration model achieved a coefficient of determination (R_c^2) of 0.976 and an $RMSE_C$ of 0.008. The cross-validation results showed an R_v^2 of 0.974 and an $RMSE_V$ of 0.008. When validated using unseen samples, the model achieved an R_p^2 of 0.978 and an $RMSE_P$ of 0.008. Figure 12 illustrates the calibration and prediction performance of this optimal SVR model based on the 1st order derivative preprocessing. The optimal SVR model utilized a Radial Basis Function (RBF) kernel with optimized hyperparameters ($C = [1]$, $\epsilon = [0.1]$, $\gamma = [0.003]$).

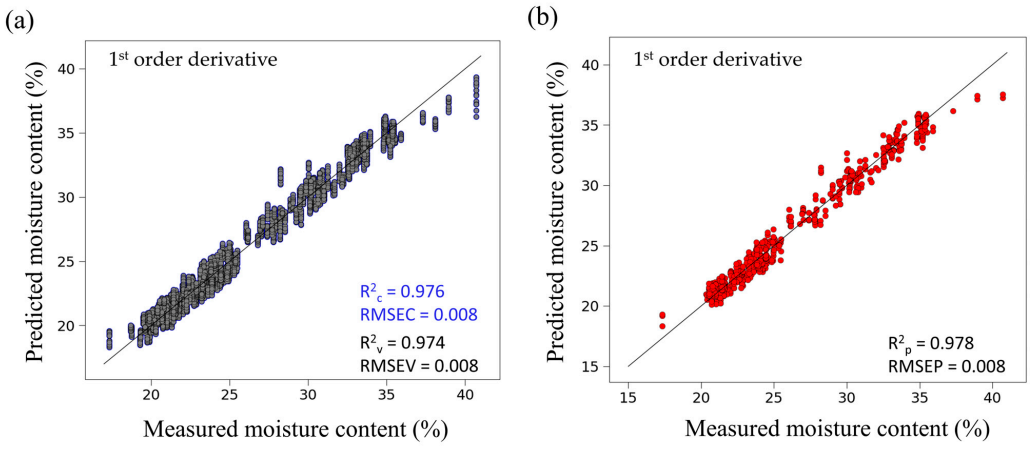

Figure 12. Optimal performances of SVR models for paddy rice moisture content prediction: (**a**) calibration with 1st-order derivative data, (**b**) prediction with 1st-order derivative data.

In general, the application of spectral preprocessing improved the performance of the SVR models for predicting moisture content. Notably, when Savitzky–Golay 1st and 2nd order derivative preprocessing techniques were applied, the SVR models consistently achieved R_p^2 values above 0.970 and $RMSE_P$ values below 0.010.

Overall, the SVR models demonstrated superior predictive performance compared to the PLSR models for estimating moisture content in paddy rice.

3.5. Results of DNN Model Development

A DNN model was developed to predict the moisture content of paddy rice. The results obtained using various spectral preprocessing techniques are summarized in Table 6.

Table 6. DNN model performance for predicting the moisture content of paddy rice.

Model Type	Preprocessing	Calibration		Validation		Prediction	
		R_c^2	$RMSE_C$	R_v^2	$RMSE_V$	R_p^2	$RMSE_P$
DNN	Raw	0.943	0.012	0.938	0.012	0.933	0.013
	Mean Normalization	0.961	0.010	0.959	0.010	0.962	0.010
	Range Normalization	0.959	0.010	0.970	0.009	0.968	0.009
	Maximum Normalization	0.939	0.012	0.931	0.013	0.913	0.015
	1st order D* (gap size = 3.8 nm)	0.992	0.004	0.994	0.004	0.993	0.004
	2nd order D* (gap size = 7.6 nm)	0.992	0.004	0.995	0.003	0.996	0.003
	MSC	0.977	0.008	0.983	0.006	0.980	0.007
	SNV	0.982	0.007	0.985	0.006	0.982	0.007

Note. D*: Derivative.

Among the developed DNN models for moisture content prediction, the model using Savitzky–Golay second-derivative preprocessing demonstrated the best performance. The calibration model with second-derivative preprocessing achieved a coefficient of determination (R_c^2) of 0.992 and an $RMSE_C$ of 0.004. Cross-validation yielded an R_v^2 of 0.995 and an $RMSE_V$ of 0.003, while prediction with unknown samples resulted in an R_p^2 of 0.996 and an $RMSE_P$ of 0.003. Figure 13 displays the calibration and prediction results of this optimal DNN model using second-derivative preprocessing.

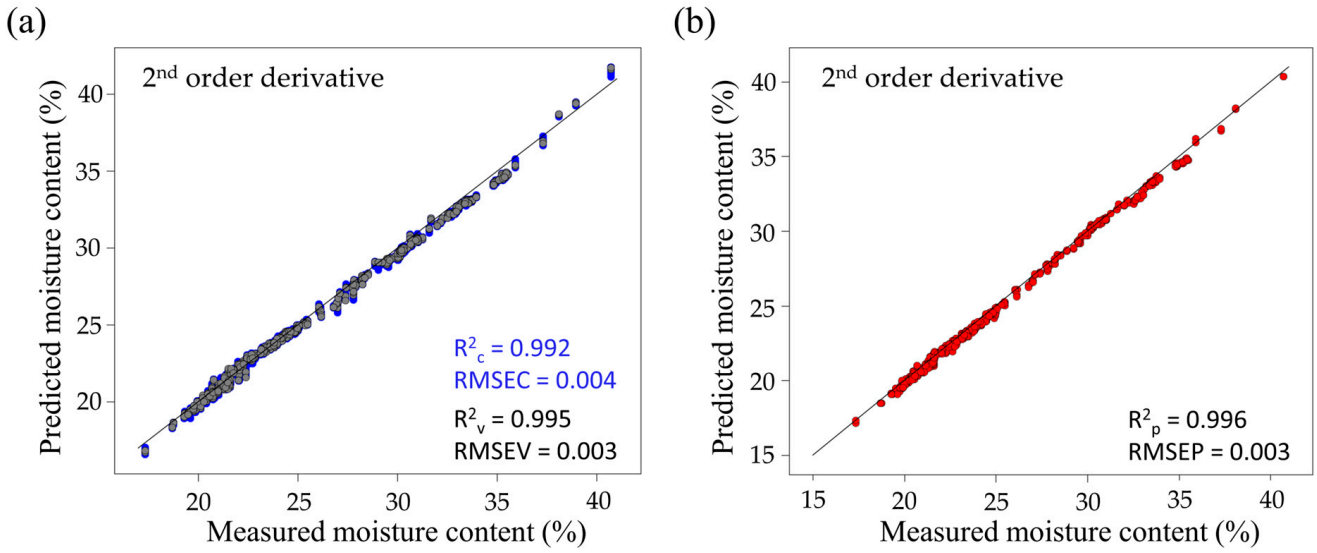

Figure 13. Optimal performances of DNN models for paddy rice moisture content prediction: (**a**) calibration with 2nd-order derivative data, (**b**) prediction with 2nd-order derivative data.

Overall, the application of spectral preprocessing techniques generally improved the predictive performance of the DNN models. In particular, when Savitzky–Golay 1st and 2nd order derivative preprocessing was applied, the DNN models consistently achieved R_p^2 values exceeding 0.990 and $RMSE_P$ values below 0.005.

The DNN models demonstrated superior prediction performance compared to both the PLSR and SVR models in estimating the moisture content of paddy rice.

3.6. Results of 1D-CNN Model Development

1D-CNN models based on VGG Net and EfficientNet architectures were developed to predict the moisture content of paddy rice from the heading stage to harvest maturity. Table 7 summarizes the performance of the 1D-CNN models trained using raw and preprocessed spectral data.

Table 7. 1D-CNN model performance for predicting the moisture content of paddy rice.

Model Type		Preprocessing	Calibration		Validation		Prediction	
			R_c^2	$RMSE_C$	R_v^2	$RMSE_V$	R_p^2	$RMSE_P$
CNN	VGG Net Based	Raw	0.989	0.005	0.994	0.004	0.993	0.004
		Mean Normalization	0.988	0.005	0.995	0.004	0.994	0.004
		Range Normalization	0.990	0.005	0.993	0.004	0.993	0.004
		Maximum Normalization	0.989	0.005	0.993	0.004	0.990	0.005
		1st order D* (gap size = 22.8 nm)	0.990	0.005	0.994	0.004	0.994	0.004
		2nd order D* (gap size = 7.6 nm)	0.989	0.005	0.994	0.004	0.992	0.005
		MSC	0.990	0.005	0.994	0.004	0.991	0.005
		SNV	0.991	0.004	0.995	0.003	0.990	0.005
	Efficient Net Based	Raw	0.998	0.002	0.998	0.002	0.999	0.001
		Mean Normalization	0.998	0.002	0.998	0.002	0.998	0.002
		Range Normalization	0.998	0.002	0.998	0.002	0.998	0.002
		Maximum Normalization	0.998	0.001	0.999	0.001	0.998	0.001
		1st order D* (gap size = 22.8 nm)	0.999	0.002	0.999	0.001	0.999	0.001
		2nd order D* (gap size = 7.6 nm)	0.998	0.002	0.998	0.002	0.998	0.002
		MSC	0.998	0.002	0.999	0.002	0.999	0.001
		SNV	0.998	0.002	0.999	0.001	0.999	0.001

Note. D*: Derivative.

For the VGG Net-based model, the model employing Savitzky–Golay 1st order derivative preprocessing showed the best performance, achieving an R_c^2 of 0.990 and an $RMSE_C$ of 0.005 in calibration, and an R_p^2 of 0.994 and an $RMSE_P$ of 0.004 in prediction.

In the case of the EfficientNet-based model, the model using Savitzky–Golay 1st order derivative preprocessing achieved the highest predictive performance, with an R_c^2 and $RMSE_C$ of 0.999 and 0.001, respectively, and an R_p^2 and $RMSE_P$ of 0.999 and 0.001, respectively.

The model developed using EfficientNet outperformed the VGG Net-based model ($R^2 \approx 0.994$), demonstrating superior performance ($R^2 \approx 0.999$), which highlights the effective applicability of the EfficientNet architecture for detecting moisture content in paddy rice from heading to harvest.

Both 1D-CNN models based on the VGG Net and EfficientNet architectures outperformed the PLSR and SVR models in terms of prediction accuracy. Furthermore, the EfficientNet-based 1D-CNN model exhibited better performance than the DNN model.

The application of spectral preprocessing techniques enhanced model performance, with most models achieving R_p^2 values above 0.990 and $RMSE_P$ values below 0.005. Figure 14 illustrates the calibration and prediction results of the optimal 1D-CNN model for predicting moisture content in paddy rice.

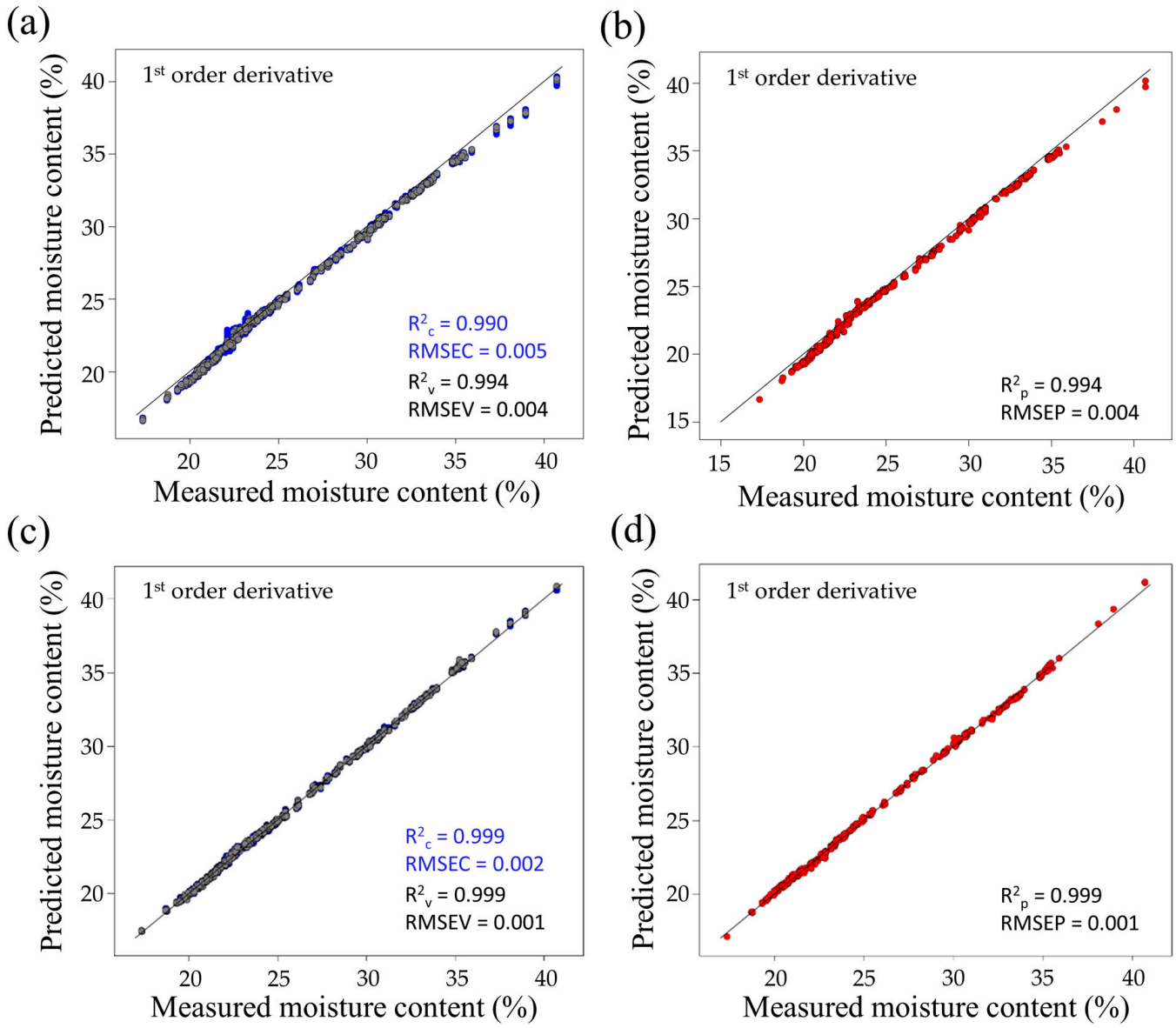

Figure 14. Calibration and prediction results of 1D-CNN models for paddy rice moisture content using 1st-order derivative data. (**a**,**b**) VGG-19-based model; (**a**) calibration, (**b**) prediction. (**c**,**d**) EfficientNet-based model; (**c**) calibration, (**d**) prediction.

3.7. Interpretability of the EfficientNet-Based 1D-CNN Model via SHAP Analysis

To elucidate the internal mechanism of the EfficientNet-based 1D-CNN model and verify the reliability of its learned spectral features, SHAP (SHapley Additive exPlanations) analysis was conducted [61]. Figure 15 presents the global feature importance, visualized by calculating the mean absolute SHAP values for each wavelength across the entire test set.

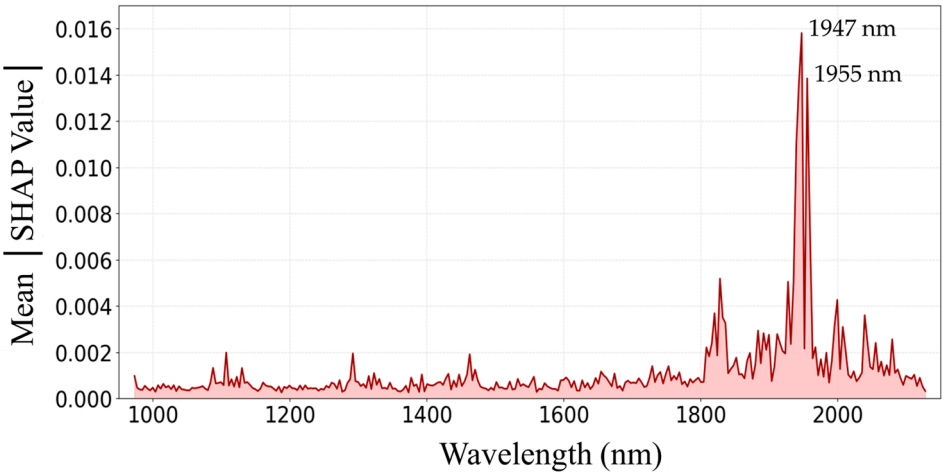

Figure 15. Global feature importance via SHAP analysis (The prominent peak at 1947 nm highlights the model's primary reliance on the O–H absorption band for moisture prediction).

The analysis revealed that within the full spectral range (950–2200 nm), the model exhibited an overwhelmingly high contribution in the specific region of 1940–1950 nm. Notably, the highest peak at 1947 nm aligns precisely with the combination band (stretching + bending modes) of O–H bonds in water (H_2O) molecules, which is a well-established absorption band in NIR spectroscopy [62]. This suggests that the deep learning model does not merely memorize statistical patterns but instead captures physico-chemical vibration characteristics directly related to moisture content.

In contrast to the PLSR model, which identified significant regression coefficients across various water-related bands (e.g., 1400 nm and 1940 nm), the EfficientNet model concentrated its predictive basis on the dominant 1940 nm signal. This phenomenon indicates that the 1D-CNN model effectively excluded redundant information within the multicollinear spectral data and selectively learned the most efficient non-linear features to minimize prediction errors [63]. Consequently, the high explainability confirmed through SHAP analysis demonstrates that the proposed model is a reliable tool with scientific validity for the non-destructive prediction of moisture content in paddy rice.

3.8. Comparison of Machine Learning and Deep Learning Model Performance

To predict the moisture content of paddy rice from the heading stage to harvest maturity, the predictive performance of the developed machine learning and deep learning models was compared. To evaluate the robustness of each model, predictions were validated using unknown samples that were not used during model training. The results are summarized in Table 8 and Figure 16.

Table 8. Optimal performance of machine learning and deep learning models for predicting moisture content in paddy rice.

Model Type	Optimal Preprocessing	Prediction with Unseen Samples	
		R_p^2	$RMSE_P$
PLSR	1st order derivative	0.941	0.012
SVR	1st order derivative	0.978	0.008
DNN	2nd order derivative	0.996	0.003
VGG Net-based 1D-CNN	1st order derivative	0.994	0.004
Efficient Net-based 1D-CNN	1st order derivative	0.999	0.001

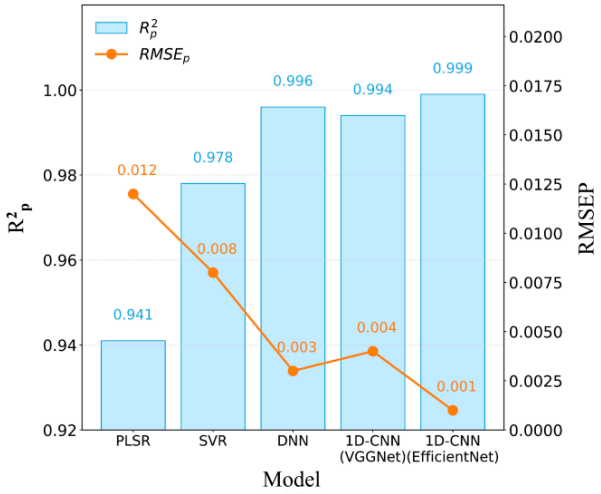

Figure 16. Performance comparison of models for paddy rice moisture prediction.

Five models in total were compared, and for each, the results of the best-performing spectral preprocessing technique were used for analysis. Among all models, the 1D-CNN model based on the EfficientNet architecture demonstrated the highest predictive performance, followed by the DNN model, VGGNet-based 1D-CNN model, SVR, and PLSR models in descending order of accuracy. The best-performing EfficientNet-based model achieved an R_P^2 of 0.999 and an $RMSE_P$ of 0.001.

The DNN-based model, while slightly less accurate than the EfficientNet-based 1D-CNN, outperformed the VGGNet-based 1D-CNN and exhibited significantly better accuracy than the machine learning models (SVR and PLSR).

To further validate the statistical significance of these performance differences, a Friedman test was performed on the RMSE values obtained from the 10-fold cross-validation. The test revealed a statistically significant difference among the five models ($X^2 = 38.80$, $p < 0.001$). Subsequent pairwise comparisons using the Conover post hoc test with Bonferroni correction (Figure 17) confirmed that the 1D-CNN model based on EfficientNet significantly outperformed all other models ($p < 0.001$). Interestingly, while the DNN showed numerically superior results to the 1D-CNN model based on VGG-19, their performance difference was not statistically significant ($p \geq 0.05$). Furthermore, the Kendall's coefficient of concordance (W) was calculated as 0.97, indicating an extremely high level of consistency in the model rankings across the 10-fold cross-validation. This suggests that the structural superiority of the EfficientNet-based architecture is robust and independent of data partitioning.

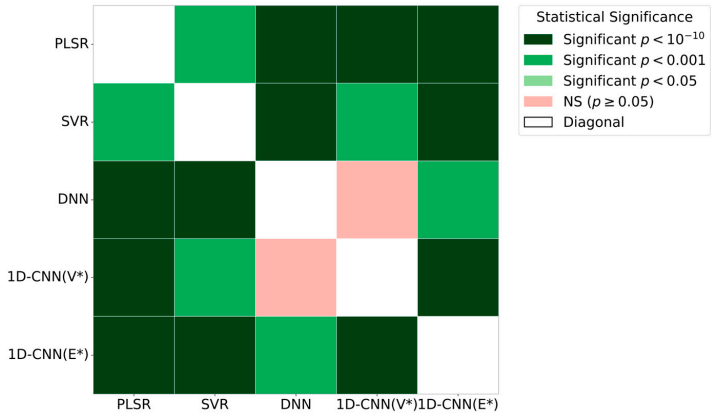

Figure 17. Heatmap of pairwise model comparisons using the Conover post hoc test with Bonferroni correction based on 10-fold cross-validation RMSE. 1D-CNN(V*) and 1D-CNN(E*) represent the 1D-

CNN models based on the VGG-19 and EfficientNet architectures, respectively. Darker green colors indicate higher statistical significance (e.g., $p < 0.001$), while 'NS' denotes no significant difference ($p \geq 0.05$).

4. Discussion

The results of this study demonstrate the potential of combining near-infrared (NIR) spectroscopy with machine learning and deep learning techniques to accurately predict the moisture content of paddy rice. The developed models were applicable to a wide range of moisture content levels (17.32% to 40.71%) in paddy rice samples collected from the heading stage to harvest maturity.

The application of spectral preprocessing techniques generally improved model performance in predicting moisture content. In particular, the use of Savitzky–Golay 1st and 2nd order derivative preprocessing enhanced predictive accuracy in PLSR, SVR, DNN, and 1D-CNN models. This suggests that the Savitzky–Golay derivative preprocessing technique effectively emphasizes meaningful patterns related to moisture content in the spectral signals of paddy rice, while also reducing high-frequency noise [64]. These findings are consistent with previous research reporting the superior effectiveness of derivative preprocessing methods over other techniques [65].

The DNN and 1D-CNN models developed in this study demonstrated overall superior performance compared to traditional machine learning models such as PLSR and SVR. This result can be attributed to the deep neural network architectures of DNN and 1D-CNN, which contain multiple hidden layers and are thus more effective in capturing the nonlinear interactions between the spectral features and the moisture content of paddy rice than the machine learning models, PLSR and SVR.

Notably, the EfficientNet-based 1D-CNN model outperformed the VGGNet-19-based model, achieving a higher predictive accuracy ($R^2 \approx 0.999$, $RMSE_P = 0.001$) compared to the VGGNet model ($R^2 \approx 0.994$, $RMSE_P = 0.004$) ($p < 0.001$). This performance improvement is likely due to the architectural advantages of EfficientNet. EfficientNet employs a compound scaling method to systematically scale network depth, width, and input resolution [40]. Furthermore, it incorporates Mobile Inverted Bottleneck Convolution (MBConv) blocks with depthwise separable convolutions and a Squeeze-and-Excitation mechanism, which dynamically adjusts channel-wise importance and allows the network to more effectively extract and focus on critical features.

In contrast, the VGGNet architecture applies a fixed structure using repeated 3×3 convolutional filters, which may limit its ability to capture features at multiple scales. Therefore, the Squeeze-and-Excitation mechanism embedded in EfficientNet's MBConv blocks may have played a key role in learning important wavelength regions related to moisture content in the spectral data.

Regarding model complexity, the total trainable parameters for the EfficientNet-based 1D-CNN and the modified VGGNet are approximately 0.64M and 1.66M, respectively. These counts are significantly lower than those of standard 2D-CNN architectures, such as VGG16 (~138M) [48] or ResNet50 (~25M) [66], confirming that the proposed models are highly optimized for 1D spectral data rather than being over-parameterized relative to the dataset size. The compact model file sizes, ranging from 2.5 to 6.5 MB, ensure seamless deployment on memory-constrained embedded edge-computing platforms (e.g., Raspberry Pi) and maintain low inference latency for real-time field applications. This balance between high predictive accuracy and model lightness justifies the structural suitability of the models for both the limited dataset scale and practical on-site requirements.

These results suggest that EfficientNet-based architectures can be effectively applied for the development of real-time analytical systems for agri-food quality assessment.

A comparative analysis of moisture prediction performance and application scope between previous studies and the present study is summarized in Table 9. Most previous studies on rice moisture prediction have primarily focused on paddy rice samples collected after harvest, typically during storage or processing stages. Moreover, the moisture content range used for model development in these studies was generally limited to approximately 10–30% (Table 9). In contrast, the present study utilized fresh paddy rice samples collected weekly from the heading stage to the optimal harvest stage, thereby expanding the measurable moisture range to 17.32–40.71%. By incorporating high-moisture rice samples that reflect pre-harvest physiological conditions, this study extends the applicability of NIR-based moisture prediction to growth-stage monitoring and optimal harvest timing decisions in field environments. In terms of prediction accuracy, the deep learning-based models developed in this study demonstrated superior performance compared with previously reported approaches. For instance, Yan et al. (2022) reported a prediction performance of R^2 = 0.969 and RMSEP = 0.785 using an extreme learning machine (ELM) model for paddy rice moisture prediction [67]. Similarly, Lin et al. (2019) developed a PLSR model for rice samples within a moisture range of 13–30%, achieving a prediction accuracy of R^2 = 0.977 [68]. In comparison, the 1D-CNN model proposed in this study achieved R^2 = 0.999 and RMSE = 0.001, indicating substantially improved predictive performance. These results suggest that deep learning-based models can more effectively capture the complex nonlinear relationships between spectral features and moisture content across a wider moisture range, compared with conventional chemometric or machine learning approaches.

Although the proposed model demonstrated low prediction error, several potential sources of error may arise under practical field conditions. Variations in illumination conditions and sample positioning, such as differences in measurement distance and sample tilt, can affect the stability of the reflected spectral signals. In addition, variations in husk thickness and overlapping degree may partially mask moisture-related O–H absorption features, potentially leading to increased prediction errors. These sources of error can be mitigated through stable light intensity control, the use of mechanical guide structures to maintain a constant distance between the light source and the sample, and multi-position averaging measurement strategies.

The findings of this study demonstrate that high-moisture paddy rice, from heading to harvest, can be analyzed in a non-destructive and rapid manner without the need for additional drying. The developed technology offers potential for various practical applications. If implemented as a portable device, the model can support real-time determination of optimal harvest timing directly in the field. As an indoor system, it can be utilized in RPCs to assess moisture content during the purchase or sale of paddy rice. It also holds potential for real-time moisture monitoring during rice drying operations. Furthermore, integration into combine harvesters would enable real-time acquisition of rice quality data during harvesting operations. Depending on the application scenario, system implementation strategies can be designed differently. For field-based and combine harvester applications, on-device inference using lightweight deep learning models may be suitable, while for indoor environments such as Rice Processing Complexes (RPCs), server-based computation for offline or near–real-time analysis may be more appropriate. Furthermore, from a practical implementation perspective, several considerations remain regarding real-time deployment. While the potential for real-time and embedded applications is discussed in this study, the primary focus is on verifying the algorithmic feasibility of NIR-based moisture prediction. Quantitative evaluations of inference latency, memory usage, and hardware resource requirements were not conducted in this work and should be addressed in future studies through benchmarking on actual embedded or edge-computing platforms.

Recent studies have applied NIR spectroscopy in combination with machine learning algorithms to estimate moisture content in a wide range of materials, including grains, dried foods, and biomass fuels, achieving fast and accurate results [69–71]. Such approaches can contribute significantly to maintaining quality in food storage and processing, particularly in terms of food safety and energy efficiency.

In recent years, smart precision agriculture technologies have been actively developed to support data-driven agricultural management. Current state-of-the-art approaches primarily focus on environmental monitoring using IoT sensors, remote sensing based on unmanned aerial vehicles (UAVs) or satellite imagery, and machine learning models for crop growth or yield prediction [41,72–76]. These technologies provide valuable information on crop conditions and field variability, thereby improving the efficiency of agricultural management. However, most existing smart farming systems rely on indirect indicators related to environmental and growth conditions, such as temperature, humidity, soil properties, or vegetation indices, which limit their ability to directly assess the actual quality of crops at the time of harvest [74,76]. In particular, technologies capable of rapidly and non-destructively measuring key quality parameters for determining optimal harvest timing remain limited.

In contrast, the approach proposed in this study enables direct and non-destructive estimation of moisture content in freshly harvested paddy rice by integrating near-infrared (NIR) spectroscopy with machine learning and deep learning techniques. Since moisture content is a critical quality indicator for determining the optimal harvest time, the developed model provides quality-based decision support rather than relying solely on environmental or growth information. Therefore, the proposed method is clearly differentiated from existing smart farming technologies in that it supports direct quality-oriented decision-making. Furthermore, this study presents a practical and field-applicable quality monitoring approach that does not require drying or milling processes. By enabling data-driven decision-making for optimal harvest timing, the proposed method contributes to the advancement of precision agriculture and smart crop management.

Nevertheless, this study has several limitations regarding practical deployment in industrial settings. The proposed models were developed and validated using samples collected from a single rice cultivar (Sindongjin), originating from the same growing region and a single harvest year. Such experimental conditions result in relatively limited sample heterogeneity; therefore, when models are developed using datasets collected across different cultivars, growing regions, or harvest years, predictive performance may deteriorate, as reported in previous studies [77]. To enhance the generalization performance of the proposed models, future research should include additional validation using datasets encompassing diverse japonica and indica cultivars collected from different growing regions and harvest years. Furthermore, to mitigate potential performance degradation under such generalized data conditions, future studies should apply and systematically compare various deep learning architectures, thereby further improving the robustness and practical applicability of paddy rice moisture content prediction models.

Table 9. Comparison of spectroscopic-based rice moisture prediction studies.

Study	Spectroscopic Technique (Wavelength)	Sample Type	Moisture Rangev (% w.b.)	Best Model	Prediction Performance	Application Stage
Lin et al., 2006 [23]	NIR imaging system (870–1014 nm)	milled rice	9.64–17.27	ANN	R^2 = 0.952 SEP = 0.435	Post-harvest quality evaluation
Heman and Hsieh, 2016 [8]	VNIR spectroscopy (350–1000 nm)	Paddy rice	11.5–28.7	PLSR	R^2 = 0.920 SEP = 2.510	Grain moisture measurement

Table 9. *Cont.*

Study	Spectroscopic Technique (Wavelength)	Sample Type	Moisture Rangev (% w.b.)	Best Model	Prediction Performance	Application Stage
Lin et al., 2019 [68]	NIR spectroscopy (950–1650 nm)	Paddy rice	13–30	CARS + PLSR	R^2 = 0.977 PMSEP = 0.930	Post-harvest moisture monitoring
Makky et al., 2019 [78]	SWIR spectroscopy (1000–2500 nm)	Paddy rice	10.5–27	PCA + PLSR	R^2 = 0.968 RMSE = 1.290	Post-harvest moisture detection
Yan et al., 2022 [67]	NIR spectroscopy (950–1650 nm)	Paddy rice	14.2–28	ELM	R^2 = 0.969 RMSE = 0.785	Harvest-time monitoring
Song et al., 2023 [60]	NIR Hyperspectral imaging (900–1700 nm)	Paddy rice	11.01–17.35	SPA + PLSR	R^2 = 0.965 RMSE = 0.003	Post-harvest quality monitoring
Weng et al., 2023 [24]	NIR spectroscopy (350–2500 nm)	brown rice	Not reported	Spectral transformation + PLSR	R^2 = 0.7376 RMSE = 0.314	Grain quality assessment
This study	NIR spectroscopy	Fresh paddy rice	17.32–40.71	1D-CNN	R^2 = 0.999 RMSE = 0.001	Pre-harvest monitoring (heading to optimal harvest)

Note. Abbreviations: CARS, Competitive Adaptive Reweighted Sampling; PCA, Principal Component Analysis; ELM, Extreme Learning Machine; SPA, Successive Projections Algorithm.

5. Conclusions

In this study, a moisture content prediction model for paddy rice from the heading stage to harvest maturity was developed by integrating near-infrared (NIR) spectroscopy with machine learning and deep learning techniques. To enhance model performance, seven spectral preprocessing methods were applied, and optimal prediction models were established using PLSR, SVR, DNN, and 1D-CNN architectures based on VGGNet and EfficientNet.

Among the developed models, the 1D-CNN model based on the EfficientNet architecture with Savitzky–Golay 1st order derivative preprocessing achieved the highest prediction accuracy (significantly different at $p < 0.001$), with an R_p^2 of 0.999 and an $RMSE_P$ of 0.001. All models showed improved predictive performance when derivative preprocessing was applied, indicating the effectiveness of Savitzky–Golay derivatives in enhancing moisture-related feature detection within the NIR spectral region.

In contrast to previous studies, this research utilized whole paddy rice samples including the husk to acquire NIR reflectance spectra in a non-destructive manner, achieving high prediction accuracy using deep learning approaches. These results demonstrate the applicability of the proposed method in the field of non-destructive quality assessment of grains. The developed technology has broad potential for practical implementation, including real-time determination of harvest timing in the field, immediate post-harvest quality analysis of research samples, moisture content evaluation during storage at Rice Processing Complexes (RPCs), and integration into combine harvesters for real-time quality assessment.

Future work will involve the development of generalized and practical moisture prediction models through the construction of large-scale datasets and comparative analysis of spectral features across various rice cultivars.

Author Contributions: H.-E.Y.: Conceptualization, Data curation, Formal analysis, Investigation, Software, Validation, Writing—original draft. H.-G.L.: Conceptualization, Methodology, Software. J.-E.L.: Data curation, Investigation. J.-Y.S.: Data curation, Investigation. W.-G.S.: Project administra-

tion, Resources. B.-K.C.: Conceptualization, Validation. C.M.: Project administration, Supervision, Conceptualization, Funding acquisition, Methodology, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Rural Development Administration as "Cooperative Research Program for Agriculture Science and Technology Development [Project Nos. RS2022-RD010389] and [Project Nos. RS2025-RD02215069]".

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available because the research is still in progress and part of an ongoing project.

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 data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

NIR	Near-infrared
1D-CNNs	One-dimensional convolutional neural networks
MC	Moisture content
ML	Machine learning
SNV	Standard normal variate
MSC	Multiplicative scatter correction
RBF	Radial Basis Function
DL	Deep learning
PLSR	Partial least squares regression
SVR	Support vector regression
ANN	Artificial neural networks
DNN	Deep neural networks
SNR	Signal-to-noise ratio
MLP	Multilayer perceptron
ELU	Exponential Linear Unit
RMSE	Root Mean Squared Error
$RMSE_C$	Root mean square error of calibration
$RMSE_V$	Root mean square error of validation
$RMSE_P$	Root mean square error of prediction

Appendix A Details of Model Architectures and Hyperparameters

Appendix A.1 Machine Learning Hyperparameters

The PLSR and SVR models in this study were developed using The Unscrambler X software (Version 10.4, Camo Analytics, Oslo, Norway). The specific hyperparameters and configuration settings employed for the model development are summarized in (Table A1).

Table A1. Details of Machine Learning Model Configurations (PLSR and SVR).

Model	Parameter	Setting/Value
PLSR	Algorithm	Kernel PLS
	Polynomial Order (Savitzky–Golay)	2
	Max. Latent Variables	20
SVR	Kernel Type	Radial Basis Function (RBF)
	C (Cost)	1.0
	γ (Gamma)	1/[number of features]
	ε (Epsilon)	0.1

Appendix A.2 Optimized Spectral Preprocessing Parameters

The specific Savitzky–Golay parameters for each model are summarized in Table A2. The window size was optimized for each algorithm to maximize predictive performance, based on the instrument's spectral resolution (3.8 nm). A second-order polynomial was consistently used for the fitting process.

Table A2. Optimized Savitzky–Golay preprocessing parameters for each model.

Model	Derivative Order	Gap Size (nm)	Window Size (Points)	Polynomial Order (*n*)
PLSR	1 2	22.8 7.6	13 5	2 2
SVR	1 2	3.8 7.6	3 5	2 2
DNN	1 2	3.8 7.6	3 5	2 2
1D-CNN	1 2	22.8 7.6	13 5	2 2

Appendix A.3 Deep Learning Architectures and Layer Configurations

The detailed network architectures of the three deep learning models employed in this study—DNN (RiceANN), Modified VGG-CNN (RiceCNN), and EfficientNet-1D—are summarized in Table A3.

Table A3. Network architecture of the DNN and CNN (Modified VGG Net, Efficient Net-1D).

Palatino Linotype		Layer (Type)	Configuration Details (Kernel, Stride, Channels/Nodes)	Activation
DNN		Input Layer Hidden 1–2 Hidden 3 Hidden 4 Hidden 5 Output	Number of spectral features (Input) Linear (200), BatchNorm Linear (100), BatchNorm Linear (50), BatchNorm Linear (25), BatchNorm Linear (1)	ELU
1D-CNN	VGG Net Based	Conv Block 1	Conv1d (16, k3, s1), Conv1d (16, k3, s1), BatchNorm, MaxPool (s2)	ELU
		Conv Block2	Conv1d (32, k3, s1), Conv1d (32, k3, s1), BatchNorm, MaxPool (s2)	
		Conv Block 3	Conv1d (64, k3, s1), Conv1d (64, k3, s1), BatchNorm, MaxPool (s2)	
		Conv Block 4	Conv1d (128, k3, s1), Conv1d (128, k3, s1), BatchNorm, MaxPool (s2)	
		Conv Block 5	Conv1d (256, k3, s1), Conv1d (256, k3, s1), BatchNorm, MaxPool (s2)	
		Conv Block 6	Conv1d (512, k3, s1), Conv1d (512, k3, s1), BatchNorm, MaxPool (s2)	
		Fully-Conn.	Linear (2048) → 256 → 128 → 64 → 1	
	Efficient Net Based	Stem	Conv1d (32, k3, s1, p1), BatchNorm	SiLU
		MBConv Blocks	7 Stages of MBConv1D (Expansion 1 or 6, SE-ratio 0.25)	
		Head	Conv1d (1280, k1), BatchNorm, GlobalAvgPool	
		Final FC	Linear (1280) → 128 → 1	

Note. k, s, and p denote the kernel size, stride, and padding of the convolutional layers, respectively.

Appendix A.4 Detailed Parameters of the EfficientNet-1D MBConv Blocks

The specific parameters of the MBConv blocks for the EfficientNet-based 1D-CNN model are summarized in Table A4.

Table A4. Detailed architectural parameters of the MBConv blocks in the EfficientNet-1D.

Stage	Block Type	Expansion Factor	Output Channels	Stride	Kernel Size	SE-Ratio
1	MBConv1	1	16	1	3	0.25
2	MBConv2	6	24	2	3	0.25
3	MBConv3	6	40	2	3	0.25
4	MBConv4	6	80	2	3	0.25
5	MBConv5	6	112	1	3	0.25
6	MBConv6	6	192	2	3	0.25
7	MBConv7	6	320	1	3	0.25

Appendix B Stratified Error Analysis for the 1D-CNN Model

Reliability Assessment of the 1D-CNN Model by Moisture Level

To further evaluate the practical reliability of the best-performing model (1D-CNN based on EfficientNet), a stratified error analysis was conducted. The test dataset was categorized into three moisture content ranges: Low (<20%), Medium (20–25%), and High (>25%). This detailed analysis demonstrates the model's robustness and its ability to provide stable predictions across all growth and harvest stages, ensuring precise harvest timing.

Table A5. Stratified error analysis of the EfficientNet-based 1D-CNN model across different moisture content levels.

Model	Range	Level (%)	Sample Size (*n*)	RMSEP	Bias
1D-CNN (EfficientNet)	Low	<20	392	0.00119	0.0007
	Medium	20–25	2830	0.00118	0.0006
	High	>25	6000	0.00148	0.0005

References

1. Muthayya, S.; Sugimoto, J.D.; Montgomery, S.; Maberly, G.F. An overview of global rice production, supply, trade, and consumption. *Ann. N. Y. Acad. Sci.* **2014**, *1324*, 7–14. [CrossRef]
2. Hashim, N.; Ali, M.M.; Mahadi, M.R.; Abdullah, A.F.; Wayayok, A.; Kassim, M.S.M.; Jamaluddin, A. Smart Farming for Sustainable Rice Production: An Insight into Application, Challenge, and Future Prospect. *Rice Sci.* **2023**, *31*, 47–61. [CrossRef]
3. Siebenmorgen, T.J.; Bautista, R.C.; Counce, P.A. Optimal harvest moisture contents for maximizing milling quality of long- and medium-grain rice cultivars. *Appl. Eng. Agric.* **2007**, *23*, 517–527. [CrossRef]
4. Sarkar, T.K.; Ryu, C.-S.; Kang, J.-G.; Kang, Y.-S.; Jun, S.-R.; Jang, S.-H.; Park, J.-W.; Song, H.-Y. Artificial neural network-based model for predicting moisture content in rice using UAV remote sensing data. *Korean J. Remote Sens.* **2018**, *34*, 611–624. [CrossRef]
5. Islam, M.; Shimizu, N.; Kimura, T. Energy requirement in parboiling and its relationship to some important quality indicators. *J. Food Eng.* **2004**, *63*, 433–439. [CrossRef]
6. Cnossen, A.G.; Siebenmorgen, T.J. The glass transition temperature concept in rice drying and tempering: Effect on milling quality. *Trans. ASAE* **2000**, *43*, 1661–1667. [CrossRef]
7. Lu, R.; Siebenmorgen, T.J.; Dilday, R.H.; Costello, T.A. Modeling long-grain rice milling quality and yield during the harvest season. *Trans. ASAE* **1992**, *35*, 1905–1913. [CrossRef]
8. Heman, A.; Hsieh, C.-L. Measurement of moisture content for rough rice by visible and near-infrared (NIR) spectroscopy. *Eng. Agric. Environ. Food* **2016**, *9*, 280–290. [CrossRef]
9. Liu, J.; Qiu, S.; Wei, Z. Real-time measurement of moisture content of paddy rice based on microstrip microwave sensor assisted by machine learning strategies. *Chemosensors* **2022**, *23*, 376. [CrossRef]

10. Sun, W.; Wan, L.; Che, G.; Xu, P.; Wang, H.; Qu, T. Design and experiment of capacitive rice online moisture detection device. *Sensors* **2023**, *23*, 5753. [CrossRef]
11. Yang, Y.; Cai, H.; Wu, J.; Guo, Z.; Zhou, T.; Zhang, M.; Hou, N.; Huang, W.; Jiang, X.; Yin, J.; et al. Measurement system based on microwave pseudo waveguide and LSTM neural network for moisture content of rice grains. *J. Food Process. Eng.* **2025**, *48*, e70188. [CrossRef]
12. Moran, M.; Inoue, Y.; Barnes, E. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote. Sens. Environ.* **1997**, *61*, 319–346. [CrossRef]
13. Czaja, T.P.; Engelsen, S.B. Why nothing beats NIRS technology: The green analytical choice for the future sustainable food production. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2024**, *325*, 125028. [CrossRef]
14. Du, Z.; Tian, W.; Tilley, M.; Wang, D.; Zhang, G.; Li, Y. Quantitative assessment of wheat quality using near-infrared spectroscopy: A comprehensive review. *Compr. Rev. Food Sci. Food Saf.* **2022**, *21*, 2956–3009. [CrossRef]
15. Vincent, B.; Dardenne, P. Application of NIR in Agriculture. In *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 331–345. [CrossRef]
16. Williams, P.; Norris, K. *Near-Infrared Technology in the Agricultural and Food Industries*; American Association of Cereal Chemists: St. Paul, MN, USA, 2020.
17. Cheng, T.; Song, R.; Li, D.; Zhou, K.; Zheng, H.; Yao, X.; Tian, Y.; Cao, W.; Zhu, Y. Spectroscopic estimation of biomass in canopy components of paddy rice using dry matter and chlorophyll indices. *Remote Sens.* **2017**, *9*, 319. [CrossRef]
18. Zheng, H.; Cheng, T.; Zhou, M.; Li, D.; Yao, X.; Tian, Y.; Cao, W.; Zhu, Y. Improved estimation of rice aboveground biomass combining textural and spectral analysis of UAV imagery. *Precis. Agric.* **2019**, *20*, 611–629. [CrossRef]
19. He, J.; Zhang, N.; Su, X.; Lu, J.; Yao, X.; Cheng, T.; Zhu, Y.; Cao, W.; Tian, Y. Estimating leaf area index with a new vegetation index considering the influence of rice panicles. *Remote Sens.* **2019**, *11*, 1809. [CrossRef]
20. Prabhakar, M.; Gopinath, K.A.; Ravi Kumar, N.; Thirupathi, M.; Sai Sravan, U.; Srasvan Kumar, G.; Samba Siva, G.; Chandana, P.; Singh, V.K. Mapping leaf area index at various rice growth stages in southern India using airborne hyperspectral remote sensing. *Remote Sens.* **2024**, *16*, 954. [CrossRef]
21. Tan, S.; Pei, J.; Zou, Y.; Fang, H.; Wang, T.; Huang, J. Improving rice yield prediction with multi-modal UAV data: Hyperspectral, thermal, and LiDAR integration. *Geo-Spat. Inf. Sci.* **2025**. [CrossRef]
22. Zhang, X.; Yang, J.; Lin, T.; Ying, Y. Food and agro-product quality evaluation based on spectroscopy and deep learning: A review. *Trends Food Sci. Technol.* **2021**, *112*, 431–441. [CrossRef]
23. Lin, L.; Lu, F.; Chang, Y. Development of a Near-Infrared Imaging System for Determination of Rice Moisture. *Cereal Chem.* **2006**, *83*, 498–504. [CrossRef]
24. Weng, S.; Tang, L.; Wang, J.; Zhu, R.; Wang, C.; Sha, W.; Zheng, L.; Huang, L.; Liang, D.; Hu, Y.; et al. Detection of amylase activity and moisture content in rice by reflectance spectroscopy combined with spectral data transformation. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2023**, *290*, 122311. [CrossRef]
25. Mishra, P.; Passos, D.; Marini, F.; Xu, J.; Amigo, J.M.; Gowen, A.A.; Jansen, J.J.; Biancolillo, A.; Roger, J.M.; Rutledge, D.N.; et al. Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TrAC Trends Anal. Chem.* **2022**, *157*, 116804. [CrossRef]
26. Chen, H.; Chen, A.; Xu, L.; Xie, H.; Qiao, H.; Lin, Q.; Cai, K. A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources. *Agric. Water Manag.* **2020**, *240*, 106303. [CrossRef]
27. Xiao, Q.; Wu, N.; Tang, W.; Zhang, C.; Feng, L.; Zhou, L.; Shen, J.; Zhang, Z.; Gao, P.; He, Y. Visible and near-infrared spectroscopy and deep learning application for the qualitative and quantitative investigation of nitrogen status in cotton leaves. *Front. Plant Sci.* **2022**, *13*, 1080745. [CrossRef]
28. Bagchi, T.B.; Sharma, S.; Chattopadhyay, K. Development of NIRS models to predict protein and amylose content of brown rice and proximate compositions of rice bran. *Food Chem.* **2016**, *191*, 21–27. [CrossRef]
29. Fazeli Burestan, N.; Afkari Sayyah, A.H.; Taghinezhad, E. Prediction of some quality properties of rice and its flour by near-infrared spectroscopy (NIRS) analysis. *Food Sci. Nutr.* **2021**, *9*, 1099–1105. [CrossRef]
30. Yang, H.-E.; Kim, N.-W.; Lee, H.-G.; Kim, M.-J.; Sang, W.-G.; Yang, C.; Mo, C. Prediction of protein content in paddy rice (*Oryza sativa* L.) combining near-infrared spectroscopy and deep-learning algorithm. *Front. Plant Sci.* **2024**, *15*, 1398762. [CrossRef] [PubMed]
31. Lee, I.-S.; Ko, D.-Y.; Choi, C.-H. Sindongjin Rice Has Changed the Image of Jeonbuk Rice: A Review. In *Proceedings of the Korean Society of Crop Science Conference*; The Korean Society of Crop Science: Suwon-si, Republic of Korea, 2023; p. 43.
32. Kim, H.-G.; Park, S.-Y.; Son, W.-C.; Lee, J.-I. Analysis of management efficiency of Sindongjin rice farms using DEA. *J. Korea Acad.-Ind. Coop. Soc.* **2020**, *21*, 61–69. [CrossRef]
33. *JSAM S4102*; Standard Method for Determination of Moisture Content of Rough Rice. Japan Society of Agricultural Machinery: Tokyo, Japan, 1984.

34. Carneiro, L.d.O.; Coradi, P.C.; Rodrigues, D.M.; Lima, R.E.; Teodoro, L.P.R.; de Moraes, R.S.; Teodoro, P.E.; Nunes, M.T.; Leal, M.M.; Lopes, L.R.; et al. Characterizing and predicting the quality of milled rice grains using machine learning models. *Agriengineering* **2023**, *5*, 1196–1215. [CrossRef]
35. Mevik, B.-H.; Wehrens, R. TheplsPackage: Principal component and partial least squares regression in R. *J. Stat. Softw.* **2007**, *18*, 1–23. [CrossRef]
36. Abdi, H. Partial least square regression (PLS regression). *Encycl. Res. Methods Soc. Sci.* **2003**, *6*, 792–795.
37. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [CrossRef]
38. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
39. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1995.
40. Basak, D.; Pal, S.; Patranabis, D.C. Support vector regression. *Neural Inf. Process.-Lett. Rev.* **2007**, *11*, 203–224.
41. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef]
42. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
43. Acquarelli, J.; van Laarhoven, T.; Gerretzen, J.; Tran, T.N.; Buydens, L.M.; Marchiori, E. Convolutional neural networks for vibrational spectroscopic data analysis. *Anal. Chim. Acta* **2017**, *954*, 22–31. [CrossRef]
44. Kiranyaz, S.; Avci, O.; Abdeljaber, O.; Ince, T.; Gabbouj, M.; Inman, D.J. 1D convolutional neural networks and applications: A survey. *Mech. Syst. Signal Process.* **2021**, *151*, 107398. [CrossRef]
45. Nguyen, T.-H.; Nguyen, T.-N.; Ngo, B.-V. A VGG-19 Model with transfer learning and image segmentation for classification of tomato leaf disease. *Agriengineering* **2022**, *4*, 871–887. [CrossRef]
46. Singh, R.; Rana, R.; Singh, S.K. Performance evaluation of VGG models in detection of wheat rust. *Asian J. Comput. Sci. Technol.* **2018**, *7*, 76–81. [CrossRef]
47. Turaev, S.; Almisreb, A.A.; Saleh, M.A. Application of Transfer Learning for Fruits and Vegetable Quality Assessment. In *2020 14th International Conference on Innovations in Information Technology (IIT)*; IEEE: Piscataway, NJ, USA, 2020; pp. 7–12.
48. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]
49. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*; PMLR: New York, NY, USA, 2019; pp. 6105–6114.
50. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* **1937**, *32*, 675–701. [CrossRef]
51. Conover, W.J. *Practical Nonparametric Statistics*; John Wiley & Sons: Hoboken, NJ, USA, 1999.
52. Jeon, D.H.; Lee, W.W.; Han, S.W.; Cho, Y.C.; Kim, H.D. *Study on Quality Improvement of Gyeonggi Rice: Determination of Optimal Harvest Timing by Transplanting Period for High-Quality Rice Production*; Research Report; Gyeonggi-do Agricultural Research & Extension Services: Suwon, Republic of Korea, 2005; pp. 71–79.
53. Rachmawati; Rohaeti, E.; Rafi, M. Combination of near infrared spectroscopy and chemometrics for authentication of taro flour from wheat and sago flour. *J. Phys. Conf. Ser.* **2017**, *835*, 012011. [CrossRef]
54. Guo, W.; Zhao, F.; Dong, J. Nondestructive measurement of soluble solids content of kiwifruits using near-infrared hyperspectral imaging. *Food Anal. Methods* **2015**, *9*, 38–47. [CrossRef]
55. Kamruzzaman, M.; ElMasry, G.; Sun, D.-W.; Allen, P. Application of NIR hyperspectral imaging for discrimination of lamb muscles. *J. Food Eng.* **2011**, *104*, 332–340. [CrossRef]
56. Morón, A.; García, A.; Sawchik, J.; Cozzolino, D. Preliminary study on the use of near-infrared reflectance spectroscopy to assess nitrogen content of undried wheat plants. *J. Sci. Food Agric.* **2006**, *87*, 147–152. [CrossRef]
57. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]
58. Kim, M.-J.; Lee, J.-E.; Back, I.; Lim, K.J.; Mo, C. Estimation of total nitrogen content in topsoil based on machine and deep learning using hyperspectral imaging. *Agriculture* **2023**, *13*, 1975. [CrossRef]
59. Das, B.; Sahoo, R.N.; Pargal, S.; Krishna, G.; Verma, R.; Viswanathan, C.; Sehgal, V.K.; Gupta, V.K. Evaluation of different water absorption bands, indices and multivariate models for water-deficit stress monitoring in rice using visible-near infrared spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *247*, 119104. [CrossRef] [PubMed]
60. Song, Y.; Cao, S.; Chu, X.; Zhou, Y.; Xu, Y.; Sun, T.; Zhou, G.; Liu, X. Non-destructive detection of moisture and fatty acid content in rice using hyperspectral imaging and chemometrics. *J. Food Compos. Anal.* **2023**, *121*, 105397. [CrossRef]
61. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]
62. Workman, J., Jr.; Weyer, L. *Practical Guide to Interpretive Near-Infrared Spectroscopy*; CRC Press: Boca Raton, FL, USA, 2007; Chapter 6.2; pp. 63–65. [CrossRef]

63. 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. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2023**, *304*, 123402. [CrossRef] [PubMed]
64. Ruffin, C.; King, R.L. The analysis of hyperspectral data using savitzky-golay filtering-theoretical basis. In *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293)*; IEEE: Piscataway, NJ, USA, 1999; pp. 756–758. [CrossRef]
65. Vestergaard, R.-J.; Vasava, H.B.; Aspinall, D.; Chen, S.; Gillespie, A.; Adamchuk, V.; Biswas, A. Evaluation of optimized preprocessing and modeling algorithms for prediction of soil properties using vis-nir spectroscopy. *Sensors* **2021**, *21*, 6745. [CrossRef] [PubMed]
66. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE on Compute Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2015; pp. 770–778.
67. Yan, J.; Tian, H.; Wang, S.; Wang, Z.; Xu, H. Paddy moisture on-line detection based on ensemble preprocessing and modeling for combine harvester. *Comput. Electron. Agric.* **2022**, *198*, 107050. [CrossRef]
68. Lin, L.; He, Y.; Xiao, Z.; Zhao, K.; Dong, T.; Nie, P. Rapid-detection sensor for rice grain moisture based on NIR spectroscopy. *Appl. Sci.* **2019**, *9*, 1654. [CrossRef]
69. Malvandi, A.; Feng, H.; Kamruzzaman, M. Application of NIR spectroscopy and multivariate analysis for Non-destructive evaluation of apple moisture content during ultrasonic drying. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *269*, 120733. [CrossRef] [PubMed]
70. Toscano, G.; Leoni, E.; Gasperini, T.; Picchi, G. Performance of a portable NIR spectrometer for the determination of moisture content of industrial wood chips fuel. *Fuel* **2022**, *320*, 123948. [CrossRef]
71. Xue, H.; Xu, X.; Yang, Y.; Hu, D.; Niu, G. Rapid and Non-Destructive Prediction of Moisture Content in Maize Seeds Using Hyperspectral Imaging. *Sensors* **2024**, *24*, 1855. [CrossRef]
72. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
73. Navarro, E.; Costa, N.; Pereira, A. A systematic review of IoT solutions for smart farming. *Sensors* **2020**, *20*, 4231. [CrossRef]
74. Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of remote sensing in precision agriculture: A review. *Remote Sens.* **2020**, *12*, 3136. [CrossRef]
75. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [CrossRef]
76. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine learning in agriculture: A comprehensive updated review. *Sensors* **2021**, *21*, 3758. [CrossRef]
77. Huang, Z.; Sanaeifar, A.; Tian, Y.; Liu, L.; Zhang, D.; Wang, H.; Ye, D.; Li, X. Improved generalization of spectral models associated with Vis-NIR spectroscopy for determining the moisture content of different tea leaves. *J. Food Eng.* **2021**, *293*, 110374. [CrossRef]
78. Makky, M.; Santosa Putri, R.E.; Nakano, K. Determination of moisture content in rice using non-destructive short-wave near infrared spectroscopy. *AIP Conf. Proc.* **2019**, *2155*, 020014. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.