

Deep multi-block convolutional neural network for quantitative near-infrared spectral analysis

Hilmy Abidzar Tawakal^{1,2}, Yohanes Aris Purwanto³, Sony Hartono Wijaya¹, Shelve Nidya Neyman¹

¹School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

²Department of Informatics, Nurul Fikri College of Technology, Depok, Indonesia

³Division of Biosystem Engineering, Faculty of Engineering and Technology, IPB University, Bogor, Indonesia

Article Info

Article history:

Received Jul 16, 2025

Revised Jul 17, 2026

Accepted Aug 12, 2026

Keywords:

Deep learning

Multi-block convolutional neural network

Near-infrared spectra

Preprocessing

Rice quality

Spectral analysis

ABSTRACT

Near-infrared (NIR) spectroscopy is a powerful analytical technique, but its effectiveness is often hindered by instrumental noise and a dependency on expert-driven preprocessing and calibration. This study introduces a novel, intelligent deep learning architecture, a deep multi-block convolutional neural network (CNN) that addresses these challenges by performing end-to-end quantitative analysis directly on raw spectral data. By segmenting the NIR spectrum into multiple blocks and processing them in parallel, the model mimics multi-scale feature learning, automatically identifying and leveraging localized spectral regions relevant to target properties. This approach eliminates the need for manual preprocessing, a significant step toward creating autonomous analytical systems. This study evaluated the model on four rice quality datasets (moisture, protein, fatty acids, and total fat) and benchmarked it against traditional partial least squares regression (PLSR) and other deep learning models. The proposed multi-block CNN consistently achieved superior performance, demonstrating its ability to handle raw spectral data robustly. For instance, it achieved coefficient of determination (R^2) values of 0.996 ± 0.003 for total fat and 0.997 ± 0.001 for fatty acids. Statistical tests confirm that the model's performance on raw data is not significantly different from its performance with optimized preprocessing, validating its potential for real-world artificial intelligence applications in agriculture, food quality monitoring, and biosystems engineering where rapid, automated analysis is critical.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Yohanes Aris Purwanto

Division of Biosystem Engineering, Faculty of Agricultural Technology, IPB University
Bogor, Indonesia

Email: arispurwanto@apps.ipb.ac.id

1. INTRODUCTION

Near-infrared (NIR) spectroscopy offers rapid, non-destructive, and safe chemical analysis, making it invaluable in fields like agriculture and food science [1]–[3]. However, raw NIR spectra are frequently corrupted by noise, baseline drift, and light scattering effects, which can obscure the underlying chemical information [4]. Traditionally, this challenge is addressed through a multi-step pipeline that involves expert-driven preprocessing techniques followed by the construction of a calibration model, typically using methods such as partial least squares regression (PLSR). The success of this pipeline is highly dependent on human expertise, and an inappropriate choice of preprocessing can even degrade model performance more than

using no preprocessing at all [5].

Deep learning has emerged as a powerful alternative, capable of learning hierarchical features directly from data and thus automating the feature engineering process [6], a trend that has been extensively reviewed in recent literature [7]. This has led to a recent surge of interest in the field, alongside a critical discussion of its benefits and potential pitfalls [8]. Studies have shown that convolutional neural network (CNN) can effectively replace traditional preprocessing steps and achieve promising results in NIR spectral analysis [9]. Several researchers have successfully developed CNN architectures for NIR data that perform better on raw spectra than on preprocessed data [5], [10]–[13]. However, an important research gap remains: existing CNN-based spectral models often treat the spectrum as a monolithic whole and fail to capture localized spectral relevance effectively, even though chemical information in NIR spectra is not uniformly distributed. Specific organic compounds correlate strongly with narrow, distinct wavelength regions, while other regions may contain only noise.

To address this gap, this study proposes a novel deep multi-block CNN architecture that segments the NIR spectrum and learns localized features in parallel, building on recent advancements in parallel input models for spectral analysis [14]. Inspired by the inception layer concept and attention mechanisms [15], the architecture functions as an intelligent system that autonomously identifies and weighs the importance of different spectral regions. This study's primary objective is to develop and validate a preprocessing-free, end-to-end CNN architecture for robust NIR spectral regression. This study hypothesizes that by enabling the model to focus on localized spectral segments, it can enhance its ability to extract meaningful features, reduce the impact of noise, and build more accurate and generalizable calibration models directly from raw data. This work advances intelligent systems by demonstrating how a modular, block-based architecture can enable automated learning directly from raw sensor data, highlighting its potential for real-world artificial intelligence applications in quality control and process monitoring.

2. MATERIALS AND METHODS

2.1. Materials

This study used NIR spectra of rice to estimate essential properties: moisture, protein, fatty acids, and total fat, which are key indicators of rice quality [16]–[18]. Rice samples are placed in a glass container and sealed with dark paper to minimize environmental interference, such as ambient light, during measurement. Rice samples were stored in a container and measured monthly over a 5-month period. NIR spectra were collected using a spectrometer with a wavelength range of 1,350 to 2,500 nm, resulting in 257 spectral features per sample. For calibration, reference values for rice properties were obtained from certified laboratories.

2.2. Methods

2.2.1. Preprocessing methods

To establish a performance benchmark, this study compared models trained on raw data with those trained on data preprocessed with various conventional techniques, including scatter correction (standard normal variate (SNV), multiplicative scatter correction (MSC), extended multiplicative scatter correction (EMSC)), asymmetric least squares baseline correction (ALS), and smoothing (Savitzky-Golay). These methods address common issues in NIR spectra, such as baseline drift, noise, and scattering effects. They were selected based on prior research demonstrating their effectiveness [19]–[21].

2.2.2. Baseline calibration models

To evaluate the proposed method, this study benchmarked its performance against a diverse set of traditional and deep learning-based models. This study included PLSR as a representative of the standard linear chemometric methods widely used in spectral data analysis. To provide a rationale for using more complex architectures, two baseline models were implemented: a simple deep neural network (DNN) and a simple CNN. The simple DNN consisted of two hidden layers with 64 neurons each using rectified linear unit (ReLU) activation, designed to assess performance without convolutional feature extraction. The simple CNN featured a Conv1D layer with 64 filters, followed by MaxPooling1D, a second Conv1D layer with 128 filters, and a final dense output layer. Finally, the proposed multi-block architecture was compared with state-of-the-art models in the literature, specifically DeepSpectra [5] and spectral-visual geometry group (SpectraVGG) [13], which represent more advanced, deeper CNN architectures.

2.2.3. Proposed multi-block convolutional neural network architecture

The proposed multi-block CNN is an intelligent architecture designed for end-to-end analysis, integrating feature extraction and regression to operate directly on raw spectra. The design in this study is inspired by the effectiveness of parallel processing of spectral segments, a concept that has proven successful in recent chemometric models [14]. As illustrated in the conceptual diagram in Figure 1, the architecture's logic comprises three key stages: spectral segmentation, parallel feature extraction, and feature fusion for regression. This figure shows the flow from raw spectra → block segmentation → parallel CNN blocks → concatenation and dense layers → output.

- i) Block segmentation strategy: the input spectrum (1×257 vector) is first divided into B contiguous, non-overlapping blocks of equal width. This segmentation allows the model to learn localized features in parallel, since different spectral regions encode distinct chemical information. The number of blocks, B , is a key hyperparameter determined during optimization.
- ii) Parallel CNN feature extraction blocks: each spectral block is processed in parallel by an identical CNN sub-network designed to act as an automated feature extractor for its assigned region. Each sub-network consists of a 1D convolutional layer with a ReLU activation function, followed by a MaxPooling1D layer to reduce dimensionality. This is followed by a second 1D convolutional layer with ReLU activation and an AveragePooling1D layer, concluding with a flatten layer to produce the block's specific feature vector. The optimal number of filters and kernel sizes for these convolutional layers are determined through hyperparameter optimization.
- iii) Feature fusion and regression: the flattened feature vectors from all B blocks are concatenated into a single, comprehensive feature vector. This vector is then processed by a deep, dense network (comprising multiple fully connected layers with ReLU activation) that models the non-linear relationship between the spectral features and the target property. Batch normalization and dropout are applied to enhance generalization before the final single-neuron output layer.

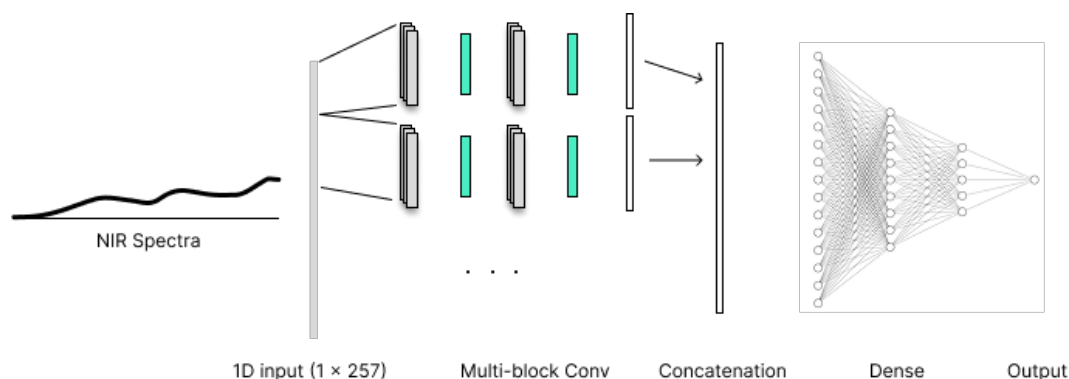


Figure 1. Conceptual diagram of the proposed multi-block CNN architecture

2.2.4. Intelligent hyperparameter optimization and training

To autonomously determine the optimal model configuration, this study employed an intelligent hyperparameter optimization strategy using the Optuna framework [22]. This automated approach is a practical implementation for tuning deep spectral models, as detailed in recent hands-on tutorials for the chemometrics community [23], [24]. The tree-structured Parzen estimator (TPE) algorithm was used to efficiently search the hyperparameter space over 100 trials, with the objective of maximizing the validation coefficient of determination (R^2) score.

The search space was defined to explore key architectural and training parameters, including the number of blocks ($B \in \{2, 4, 8\}$), the number of filters in convolutional layers ($layers \in [16, 64]$), kernel sizes ($kernel \in \{3, 5, 7\}$), the number of neurons in dense layers ($node \in [8, 32]$), the dropout rate ($do \in [0.1, 0.5]$), and the learning rate ($lr \in [10^{-4}, 10^{-2}]$). The optimal hyperparameters for the multi-block CNN were determined using Optuna, yielding $B = 8$ blocks. Each block consists of two convolutional layers: the first with 32 filters (kernel size 5) and the second with 16 filters (kernel size 5). The subsequent dense layers contain 16, 8, and 4 neurons, respectively. To prevent overfitting, a dropout rate of 0.2 was applied, and the model was trained with a learning rate of 0.001.

2.2.5. Experimental setup and evaluation

To ensure rigorous and unbiased performance estimation, we discarded the traditional train-test split in favor of a nested cross-validation (Nested CV) protocol. The Nested CV consisted of an outer loop (5 folds) dedicated strictly to final model evaluation and an inner loop (3 folds) dedicated to hyperparameter tuning. This guarantees that the hyperparameters are optimized without any data leakage from the test set. All reported performance metrics represent the mean $\pm$ standard deviation (SD) across the 5 outer test folds.

Model performance was evaluated using six key metrics: R^2 , mean absolute error (MAE), mean absolute percentage error (MAPE), cross-validation root mean squared error (RMSE_{cv}), prediction root mean squared error (RMSE_{p}), ratio of performance to deviation (RPD), and ratio of error to range (RER). The RMSE_{cv} is calculated during the inner tuning loop to rank hyperparameters, whereas RMSE_{p} measures the true generalization error on the held-out outer folds. In chemometrics, an $\text{RPD} > 3$ is generally considered acceptable for screening purposes, and an $\text{RPD} > 5$ is considered suitable for quality control. To ensure methodological transparency and facilitate reproducibility, all experimental code, hyperparameter search spaces, and model weights have been documented. Importantly, the test set in the outer loop of the nested cross-validation was strictly held out and completely isolated from the hyperparameter search process. Scripts for reproducing the results and generating synthetic data schemas (to protect proprietary datasets) are available upon request.

3. RESULTS AND DISCUSSION

3.1. Model performance on raw spectra

First, this study evaluated all models on raw spectral data to establish a baseline for their intrinsic learning capabilities without manual feature engineering. As shown in the heatmap in Figure 2, the proposed multi-block CNN achieved the highest R^2 scores across all four target properties. This figure demonstrates the superior performance of the multi-block CNN on raw spectral data. This result highlights its powerful, built-in ability to extract meaningful features directly from unprocessed data. In contrast, the simple DNN performed the worst, confirming that convolutional layers are essential for effective feature extraction from 1D spectral data. While the simple CNN and more complex architectures like DeepSpectra and SpectraVGG performed well, they were consistently outperformed by our multi-block approach, underscoring its gains.

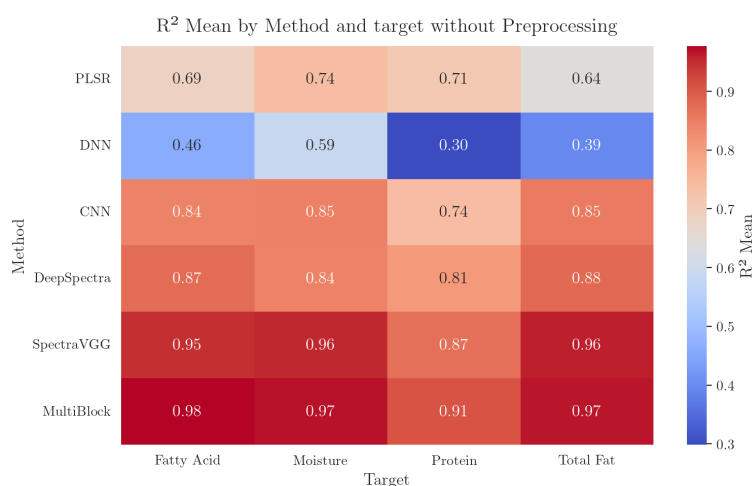


Figure 2. Heatmap of mean R^2 values for each model without any preprocessing

3.2. Impact of preprocessing

Next, this study evaluated each model using its optimal preprocessing method, identified through exhaustive testing. Table 1 summarizes both the selected best preprocessing method and the resulting mean $R^2 \pm \text{SD}$ for each target. The multi-block CNN, when combined with its optimal preprocessing, again outperformed all other models and showed low variability. More importantly, its performance on raw data (MultiBlock_RAW) remained superior to all other models, even when they used their optimal preprocessing.

Table 1. Summary of optimal preprocessing and corresponding mean $R^2 \pm$ SD for each model and dataset

Method	Total fat	Fatty acid	Moisture	Protein
PLSR	0.655 $\pm$ 0.046 [SG]	0.940 $\pm$ 0.059 [SG]	0.754 $\pm$ 0.040 [SG]	0.715 $\pm$ 0.055 [SG]
DNN	0.966 $\pm$ 0.010 [Cmb]	0.963 $\pm$ 0.008 [Cmb]	0.940 $\pm$ 0.043 [Cmb]	0.854 $\pm$ 0.024 [Cmb]
CNN	0.964 $\pm$ 0.008 [MMS]	0.959 $\pm$ 0.011 [MMS]	0.940 $\pm$ 0.038 [MMS]	0.862 $\pm$ 0.031 [MMS]
DeepSpectra	0.948 $\pm$ 0.012 [AS]	0.954 $\pm$ 0.010 [Cmb]	0.905 $\pm$ 0.030 [SG]	0.857 $\pm$ 0.033 [SG]
SpectraVGG	0.983 $\pm$ 0.007 [MMS]	0.978 $\pm$ 0.009 [MMS]	0.966 $\pm$ 0.025 [MMS]	0.896 $\pm$ 0.023 [SS]
MultiBlock-RAW	0.995 $\pm$ 0.004 [Raw]	0.997 $\pm$ 0.001 [Raw]	0.991 $\pm$ 0.014 [Raw]	0.939 $\pm$ 0.019 [Raw]
MultiBlock	0.996 $\pm$ 0.003 [AS]	0.997 $\pm$ 0.001 [Cmb]	0.991 $\pm$ 0.014 [Raw]	0.941 $\pm$ 0.017 [AS]

SG: Savitzky-Golay, Cmb: combination, MMS: min-max scaling, AS: auto scaling, SS: standard scaling

3.3. Deep dive into multi-block convolutional neural network performance

3.3.1. Nested cross-validation performance

Table 2 presents the rigorous nested cross-validation results for the optimal multi-block CNN across all four target properties. The model demonstrated exceptional predictive accuracy. The reported values represent the mean and SD across the five outer folds of the nested cross-validation procedure. The 95% confidence intervals (CI) were computed using the Student's t-distribution with four degrees of freedom ($t_{0.975,4} = 2.776$), according to $CI = t \cdot \frac{SD}{\sqrt{n}}$, where $n = 5$ denotes the number of outer folds. The relatively narrow CI, particularly for total fat and fatty acid prediction, indicate stable generalization performance across folds and confirm the robustness of the proposed multi-block CNN under strict test-set isolation.

From an application perspective, the model achieved RPD scores ranging from 4.18 (protein) to 22.02 (fatty acid). According to standard industry thresholds, an RPD > 5 indicates that the model is highly suitable for strict quality control, while an RPD > 3 is excellent for screening. Therefore, the proposed multi-block CNN is fully capable of replacing manual chemical assays for rigorous postharvest rice evaluation.

Table 2. Nested cross-validation performance of the proposed multi-block CNN (mean $\pm$ SD; 95% CI shown in brackets)

Target	R^2 (Mean $\pm$ SD)	RMSEcv	RMSEp	RPD	RER
Total fat	0.9954 $\pm$ 0.0038	0.0858 $\pm$ 0.0139	0.0191 $\pm$ 0.0075	17.35 $\pm$ 5.23	45.94 $\pm$ 13.89
95% CI	[± 0.0047]	[± 0.0172]	[± 0.0093]	[± 6.50]	[± 17.24]
Fatty acid	0.9974 $\pm$ 0.0013	0.0133 $\pm$ 0.0020	0.00216 $\pm$ 0.00063	22.02 $\pm$ 7.12	65.70 $\pm$ 21.33
95% CI	[± 0.0017]	[± 0.0025]	[± 0.00078]	[± 8.84]	[± 26.48]
Moisture	0.9912 $\pm$ 0.0139	0.2845 $\pm$ 0.0371	0.0775 $\pm$ 0.0589	17.07 $\pm$ 6.57	46.47 $\pm$ 17.94
95% CI	[± 0.0173]	[± 0.0460]	[± 0.0732]	[± 8.15]	[± 22.27]
Protein	0.9385 $\pm$ 0.0188	0.1305 $\pm$ 0.0147	0.0754 $\pm$ 0.0078	4.18 $\pm$ 0.65	15.41 $\pm$ 1.46
95% CI	[± 0.0234]	[± 0.0182]	[± 0.0097]	[± 0.80]	[± 1.81]

3.3.2. Significance test

This study evaluated the paired cross-validation folds globally ($N = 20$) using a two-sided Wilcoxon signed-rank test. To prevent type I errors across the six evaluation metrics, we applied the Benjamini-Hochberg procedure to control the false discovery rate ($\alpha = 0.05$). The statistical analysis revealed no significant difference in performance between the raw and preprocessed pipelines. As detailed in Table 3, the corrected p -values for primary metrics such as R^2 ($q = 0.5429$) and RPD ($q = 0.1322$) confirm that the addition of preprocessing does not yield a statistically significant improvement. In this table, corrected p -values were adjusted using the Benjamini-Hochberg procedure to control the false discovery rate.

Table 3. Paired Wilcoxon signed-rank test results: preprocessed vs. raw data ($N = 20$)

Metric	Wilcoxon stat	Raw p -value	Corrected p -value	Significant ($\alpha = 0.05$)
R^2	82.0	0.4091	0.5429	False
RMSEp	80.0	0.3683	0.5429	False
MAE	96.0	0.7562	0.7562	False
MAPE	84.0	0.4524	0.5429	False
RPD	51.0	0.0441	0.1322	False
RER	51.0	0.0441	0.1322	False

3.3.3. Data efficiency

To assess model generalization as a function of data availability, this study conducted a data-efficiency analysis by varying the training set size (25%, 50%, 75%, and 100%). As detailed in Table 4, performance

scales consistently with available data. With 25% of the data, the model undergeneralized (R^2 values ranging from 0.48 for moisture to 0.73 for total fat). Performance improved sharply as the sample size increased, and with 75% of the data used for training, all targets achieved $R^2 \geq 0.90$. Using 100% of the data produced peak performance ($R^2 > 0.93$ for all targets, with fat and fatty acid above 0.99), while gains from 75% to 100% became smaller, indicating an approaching plateau.

Table 4. Summary of R^2 scores across varying training data fractions

Target	25%	50%	75%	100%
Total fat	0.73 ± 0.08	0.87 ± 0.04	0.96 ± 0.01	0.995 ± 0.004
Fatty acid	0.61 ± 0.13	0.79 ± 0.12	0.94 ± 0.04	0.997 ± 0.001
Moisture	0.48 ± 0.22	0.85 ± 0.09	0.92 ± 0.04	0.991 ± 0.014
Protein	0.49 ± 0.17	0.74 ± 0.16	0.90 ± 0.01	0.939 ± 0.019

3.3.4. Ablation study

To systematically verify the architectural contributions of the proposed model, an ablation study was conducted. This study compared the optimal multi-block CNN against a monolithic CNN variant (single block) with a comparable parameter count, which effectively removes the spectral segmentation mechanism. As shown in Table 5, the monolithic CNN (1 block) failed to model the data effectively under the nested CV protocol. This drastic performance gap confirms that segmenting the spectrum enables the network to learn localized, distinct feature maps that are otherwise lost when global pooling is applied to the entire raw spectrum.

Table 5. Impact of block segmentation on model performance (total fat dataset)

Architecture	R^2 (Mean $\pm$ SD)	$RMSE_p$	RPD	RER
Multi-block (proposed)	0.995 ± 0.004	0.019 ± 0.007	17.35 ± 5.23	45.94 ± 13.89
Monolithic (single block)	0.267 ± 0.068	0.250 ± 0.012	1.18 ± 0.05	3.21 ± 0.15

3.3.5. Permutation test for model robustness

To verify that the proposed deep multi-block CNN learns genuine wavelength–chemical relationships rather than spurious correlations or dataset artifacts, a target permutation test (y-randomization) was performed. In this experiment, the target variables (fat, fatty acid, moisture, and protein) were randomly permuted with respect to the NIR spectral inputs, thereby completely destroying the structural relationship between the predictor matrix (X) and the response vector (y). Using the previously optimized hyperparameter configurations, the model was re-evaluated under identical 5-fold cross-validation conditions. Under proper model behavior, removal of the X – y dependency should lead to a collapse in predictive performance. The results are summarized in Table 6.

Table 6. Model performance following y-randomization across 5-fold cross-validation (mean $\pm$ SD)

Target	R^2	$RMSE_p$	RPD	RER
Total fat	0.0087 ± 0.0182	0.2983 ± 0.0167	1.0124 ± 0.0093	2.6841 ± 0.1513
Fatty acid	-0.0089 ± 0.0330	0.0436 ± 0.0021	1.0038 ± 0.0160	3.0099 ± 0.1439
Moisture	0.0035 ± 0.0637	1.0006 ± 0.0360	1.0108 ± 0.0306	2.7647 ± 0.0771
Protein	-0.0020 ± 0.0353	0.3114 ± 0.0336	1.0073 ± 0.0176	3.7476 ± 0.3962

As expected, eliminating the true X – y relationship reduced predictive performance to near-random levels across all targets. The R^2 approached zero or became negative, indicating no explanatory power. Likewise, the RPD converged to approximately 1.0, which in chemometric interpretation corresponds to performance equivalent to predicting the mean value for all samples. These results confirm that the high predictive accuracy observed on the structured dataset arises from meaningful spectral–chemical correlations rather than overfitting, memorization, or background artifacts. The permutation analysis therefore provides strong statistical evidence of genuine model generalization.

3.3.6. Stress testing via artificial noise injection

To further quantify the model's resilience to instrumental degradation, a systematic stress test was conducted by injecting artificial Gaussian noise into the test set spectra. Gaussian noise, defined as $\mathcal{N}(0, \sigma^2)$, was added to the standardized spectral inputs, where σ represents the SD of the noise level relative to the original spectral variance. The noise intensity was incrementally scaled from $\sigma = 0.01$ (low noise) to $\sigma = 0.10$

(severe noise). The pre-trained multi-block CNN was then evaluated on these corrupted datasets across all 5-folds without retraining the model. The resulting average R^2 for each property is presented in Table 7.

Table 7. Model performance degradation (R^2) under increasing levels of artificial Gaussian noise

Noise level (σ)	Total fat (R^2)	Fatty acid (R^2)	Moisture (R^2)	Protein (R^2)
0.00 (baseline)	0.9924	0.9951	0.9888	0.9368
0.01 (low)	0.9891	0.9934	0.9882	0.9318
0.05 (moderate)	0.8351	0.8576	0.8887	0.8209
0.10 (severe)	0.4614	0.3385	0.6592	0.2465

As shown in Table 7, the multi-block CNN demonstrates exceptional robustness at low noise levels ($\sigma = 0.01$). The predictive performance remains virtually identical to the baseline, with R^2 degrading by less than 1% across all target properties. Even under moderate noise corruption ($\sigma = 0.05$), the architecture exhibits graceful degradation, maintaining strong predictive capabilities with R^2 values remaining above 0.82 for all properties. A substantial collapse in predictive performance is only observed under severe artificial corruption ($\sigma = 0.10$), where the injected noise variance begins to heavily mask the underlying spectral variance. However, such extreme degradation generally exceeds the typical operational variance observed in modern, calibrated Fourier transform near-infrared (FT-NIR) spectrometers. Ultimately, these results empirically validate the model's operational reliability in real-world scanning environments, confirming its ability to maintain high accuracy without the strict need for traditional scatter-correction preprocessing algorithms.

3.4. Discussion

The experimental results demonstrate that the proposed multi-block CNN model significantly outperforms both traditional and other deep learning-based approaches for quantitative NIR analysis. Its ability to operate directly on raw spectra presents a major advantage. This is demonstrated by its exceptional performance on the raw fatty acid dataset ($R^2 = 0.997 \pm 0.001$).

A key architectural advantage is the balance between model capacity and generalization. While deeper architectures like DeepSpectra and SpectraVGG have shown strong performance, they can risk overfitting on smaller datasets [5], [13]. The multi-block CNN combines depth (in its dense layers) with width (via spectral segmentation), which proved to be a more robust strategy for this task.

Despite its success, the model has several limitations. The lowest performance was observed for protein prediction ($R^2 = 0.939 \pm 0.019$), a well-known challenge in traditional NIR analysis due to subtle, often noisy spectral signals. To investigate how our model handled this challenge, we generated saliency maps to visualize its internal feature-ranking process [5]. The results, shown in Figure 3, provide a clear visual explanation for the performance disparity. The model in the figure identifies multiple, distributed, and subtle spectral regions for the more challenging protein prediction. In contrast, it focuses on a single, highly concentrated, and dominant region for the high-performing fatty acid prediction. Bright yellow indicates high importance.

The saliency map for the protein shown in Figure 3(a) reveals that the model identifies multiple diffuse regions of importance across the spectrum. This confirms that the spectral signature for protein is complex and lacks a single dominant feature. In stark contrast, Figure 3(b) presents the saliency map for fatty acid, one of the best-performing models. The model focuses its attention on a single, highly concentrated, and dominant region at the start of the spectrum. This visual evidence strongly suggests that the model's superior performance on fatty acids is due to its ability to identify a localized, information-rich spectral signature, a task that is significantly more challenging for the subtle, distributed features of proteins.

In terms of generalizability, the modular, block-based architecture is not inherently limited to rice spectra. Its ability to learn localized features from 1D data makes it a strong candidate for other quantitative spectral analyses, building on the growing body of work applying CNNs to various spectral challenges [25]. The core principle of segmenting a signal to isolate localized information is broadly applicable. In addition to absolute performance, the SDs across runs were low, confirming the model's reproducibility and stability. The low root mean squared error (RMSE) and MAPE values for fatty acid and moisture predictions, in particular, emphasize the model's suitability for precise, real-time applications such as food quality monitoring or on-site agricultural screening.

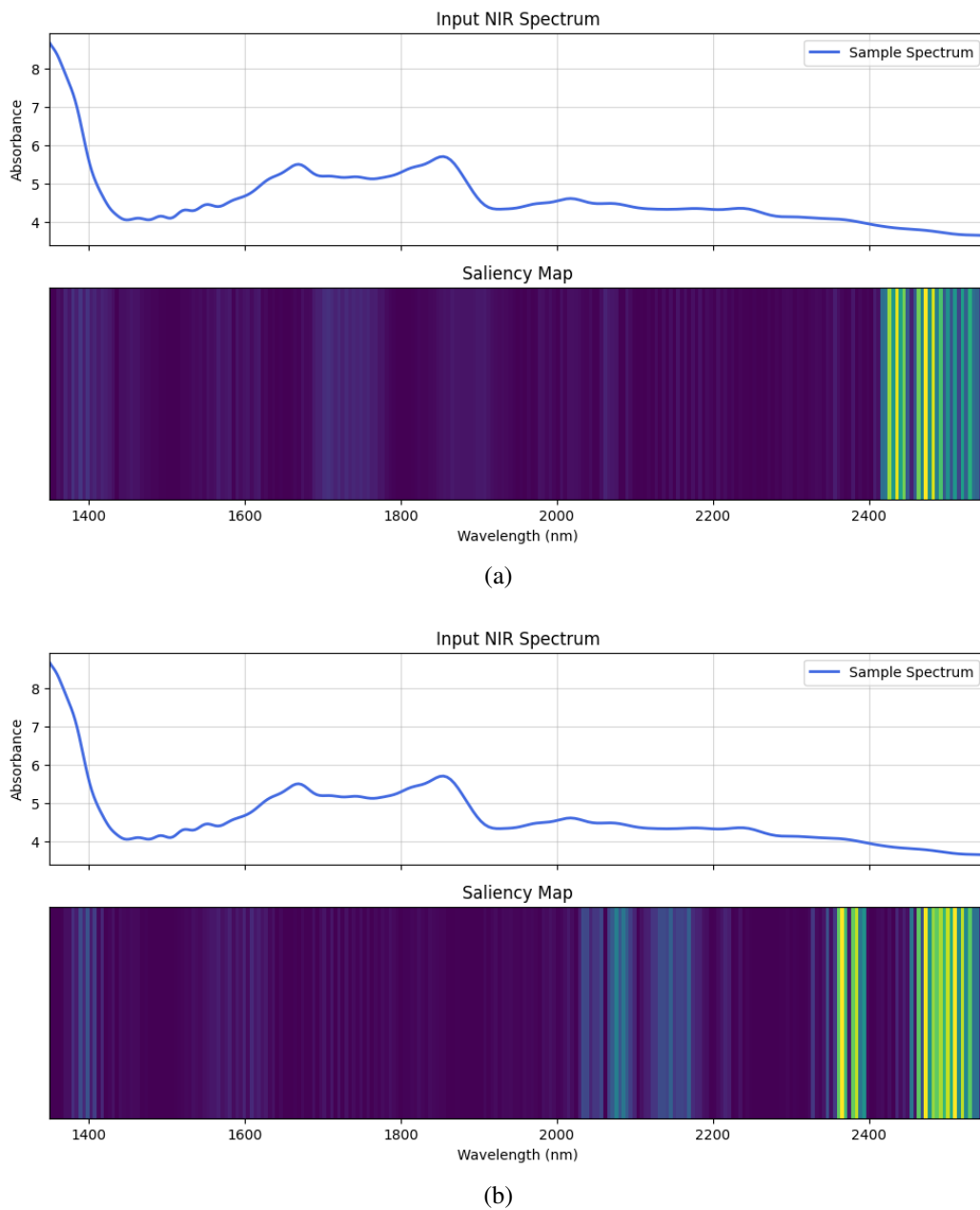


Figure 3. Feature importance maps for (a) protein prediction and (b) fatty acid prediction

3.5. Future work

Future work should focus on improving performance for challenging targets like proteins. This could involve exploring hybrid architectures that integrate the multi-block CNN with transformer or attention mechanisms to better model long-range spectral dependencies and emphasize subtle features. Additionally, investigating transfer learning to adapt the model across different crops or instruments and deploying the model on portable NIR devices for edge artificial intelligence applications represent promising directions for making this technology more accessible for real-time quality control.

4. CONCLUSION

This study introduced a novel deep multi-block CNN architecture for quantitative NIR spectral analysis, integrating feature extraction and calibration into a single, intelligent system. The proposed model

demonstrated superior performance over both traditional chemometric methods and other deep learning architectures, achieving state-of-the-art results in predicting fatty acid, moisture, and total fat in rice. The key contribution of this work is an end-to-end model that can operate directly on raw spectra without conventional preprocessing, thereby simplifying the analysis pipeline and reducing dependence on domain expertise. This study statistically validated that the model's performance on raw data is equivalent to its performance with optimized preprocessing. The architecture's novelty lies in its use of spectral segmentation and parallel processing, which enables it to learn localized features effectively and autonomously, a significant improvement over monolithic, end-to-end models.

FUNDING INFORMATION

This research is funded by the Ministry of Research, Technology, and Education through the doctoral dissertation research program (PDD), with the title "Preprocessing method of NIR spectrum based on CNN for estimation of rice quality" and the number 0536/E5/PG.02.00/2023.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Hilmy Abidzar Tawakal	✓	✓	✓	✓	✓	✓		✓	✓		✓			
Yohanes Aris Purwanto	✓				✓		✓			✓		✓	✓	✓
Sony Hartono Wijaya		✓		✓	✓					✓				
Shelvie Nidya Neyman		✓		✓	✓					✓				

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal Analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project Administration

Fu : Funding Acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Furthermore, the authors confirm that this study was conducted independently and that no external interests, affiliations, or personal relationships influenced the research design, data analysis, interpretation of the results, or preparation of the manuscript.

INFORMED CONSENT

This section is not applicable. This study did not involve human participants, human data, or any personally identifiable information. All data used were either publicly available, fully anonymized, or derived from non-human sources, and therefore no informed consent was required from individuals.

ETHICAL APPROVAL

This section is not applicable. This research did not involve human subjects, human biological materials, or experimental procedures on animals. The work was conducted solely on computational models, publicly available datasets, or non-sensitive data that did not require intervention with living organisms. Therefore, ethical approval from an institutional review board or animal ethics committee was not necessary for this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [YAP], upon reasonable request. The data, which were collected from proprietary rice samples measured with portable NIR spectroscopy, are not publicly available due to restrictions related to ongoing research and institutional data-sharing policies. Synthetic data schemas replicating the structure of the original datasets, along with the experimental code used to produce results in this paper, are available upon request to facilitate reproducibility.

REFERENCES

- [1] K. B. Walsh, V. A. McGlone, and D. H. Han, "The uses of near infra-red spectroscopy in postharvest decision support: a review," *Postharvest Biology and Technology*, vol. 163, May 2020, doi: 10.1016/j.postharvbio.2020.111139.
- [2] Y. Liu, H. Pu, and D.-W. Sun, "Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices," *Trends in Food Science & Technology*, vol. 113, pp. 193–204, Jul. 2021, doi: 10.1016/j.tifs.2021.04.042.
- [3] M. Nadimi and J. Paliwal, "Recent applications of near-infrared spectroscopy in food quality analysis," *Foods*, vol. 13, no. 16, Aug. 2024, doi: 10.3390/foods13162633.
- [4] J. Westermayr and P. Marquetand, "Machine learning spectroscopy to advance computation and analysis," *Chemical Science*, vol. 16, no. 46, pp. 21660–21676, 2025, doi: 10.1039/D5SC05628D.
- [5] X. Zhang, T. Lin, J. Xu, X. Luo, and Y. Ying, "DeepSpectra: an end-to-end deep learning approach for quantitative spectral analysis," *Analytica Chimica Acta*, vol. 1058, pp. 48–57, Jun. 2019, doi: 10.1016/j.aca.2019.01.002.
- [6] Y. Xiao, L. Zhou, Y. Zhao, H. Qi, Y. Pu, and C. Zhang, "Deep learning-based regression of food quality attributes using near-infrared spectroscopy and hyperspectral imaging: a review," *Food Chemistry*, vol. 493, No. 4, Nov. 2025, doi: 10.1016/j.foodchem.2025.145932.
- [7] W. Zhang, L. C. Kasun, Q. J. Wang, Y. Zheng, and Z. Lin, "A review of machine learning for near-infrared spectroscopy," *Sensors*, vol. 22, no. 24, Dec. 2022, doi: 10.3390/s22249764.
- [8] P. Mishra *et al.*, "Deep learning for near-infrared spectral data modelling: hypes and benefits," *TrAC Trends in Analytical Chemistry*, vol. 157, Dec. 2022, doi: 10.1016/j.trac.2022.116804.
- [9] C. Zhang, L. Zhou, Y. Zhao, S. Zhu, F. Liu, and Y. He, "Noise reduction in the spectral domain of hyperspectral images using denoising autoencoder methods," *Chemometrics and Intelligent Laboratory Systems*, vol. 203, Aug. 2020, doi: 10.1016/j.chemolab.2020.104063.
- [10] P. Nie, J. Zhang, X. Feng, C. Yu, and Y. He, "Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning," *Sensors and Actuators B: Chemical*, vol. 296, Oct. 2019, doi: 10.1016/j.snb.2019.126630.
- [11] C. Zhang, W. Wu, L. Zhou, H. Cheng, X. Ye, and Y. He, "Developing deep learning based regression approaches for determination of chemical compositions in dry black goji berries (*Lycium ruthenicum* Murr.) using near-infrared hyperspectral imaging," *Food Chemistry*, vol. 319, Jul. 2020, doi: 10.1016/j.foodchem.2020.126536.
- [12] X. Zhang, J. Yang, T. Lin, and Y. Ying, "Food and agro-product quality evaluation based on spectroscopy and deep learning: a review," *Trends in Food Science & Technology*, vol. 112, pp. 431–441, Jun. 2021, doi: 10.1016/j.tifs.2021.04.008.
- [13] U. Blazhko, V. Shapaval, V. Kovalev, and A. Kohler, "Comparison of augmentation and pre-processing for deep learning and chemometric classification of infrared spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 215, Aug. 2021, doi: 10.1016/j.chemolab.2021.104367.
- [14] P. Mishra and D. Passos, "Deep multiblock predictive modelling using parallel input convolutional neural networks," *Analytica Chimica Acta*, vol. 1163, Jun. 2021, doi: 10.1016/j.aca.2021.338520.
- [15] L. Shuai, Z. Li, Z. Chen, D. Luo, and J. Mu, "A research review on deep learning combined with hyperspectral imaging in multiscale agricultural sensing," *Computers and Electronics in Agriculture*, vol. 217, Feb. 2024, doi: 10.1016/j.compag.2023.108577.
- [16] L. Lin, Y. He, Z. Xiao, K. Zhao, T. Dong, and P. Nie, "Rapid-detection sensor for rice grain moisture based on NIR spectroscopy," *Applied Sciences*, vol. 9, no. 8, Apr. 2019, doi: 10.3390/app9081654.
- [17] X. Miao *et al.*, "Classification of rice based on storage time by using near infrared spectroscopy and chemometric methods," *Microchemical Journal*, vol. 171, Dec. 2021, doi: 10.1016/j.microc.2021.106841.
- [18] P. S. Sampaio, B. Carbas, A. Soares, I. Sousa, and C. Brites, "Spectral markers and machine learning: revolutionizing rice evaluation with near infrared spectroscopy," *Food Chemistry*, vol. 492, Nov. 2025, doi: 10.1016/j.foodchem.2025.145569.
- [19] P. Mishra, J. M. Roger, D. N. Rutledge, and E. Woltering, "SPORT pre-processing can improve near-infrared quality prediction models for fresh fruits and agro-materials," *Postharvest Biology and Technology*, vol. 168, Oct. 2020, doi: 10.1016/j.postharvbio.2020.111271.
- [20] J.-M. Roger, A. Biancolillo, and F. Marini, "Sequential preprocessing through ORThogonalization (SPORT) and its application to near infrared spectroscopy," *Chemometrics and Intelligent Laboratory Systems*, vol. 199, Apr. 2020, doi: 10.1016/j.chemolab.2020.103975.
- [21] L.-H. Xie *et al.*, "Simultaneous determination of apparent amylose, amylose and amylopectin content and classification of waxy rice using near-infrared spectroscopy (NIRS)," *Food Chemistry*, vol. 388, Sep. 2022, doi: 10.1016/j.foodchem.2022.132944.
- [22] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: a next-generation hyperparameter optimization framework," in *25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Jul. 2019, pp. 2623–2631, doi: 10.1145/3292500.3330701.
- [23] D. Passos and P. Mishra, "A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks," *Chemometrics and Intelligent Laboratory Systems*, vol. 223, Apr. 2022, doi: 10.1016/j.chemolab.2022.104520.
- [24] F. Ding, M. A. R. Delgado, R. Zha, and J. F. G.-Martín, "Optuna-optimized boosting models for predicting quality traits in multiple juice types using NIRS: interpretability analysis via SHAP," *Food Control*, vol. 182, Apr. 2026, doi: 10.1016/j.foodcont.2025.111878.

- [25] D. Passos and P. Mishra, "An automated deep learning pipeline based on advanced optimisations for leveraging spectral classification modelling," *Chemometrics and Intelligent Laboratory Systems*, vol. 215, Aug. 2021, doi: 10.1016/j.chemolab.2021.104354.

BIOGRAPHIES OF AUTHORS



Hilmy Abidzar Tawakal    is a Ph.D. candidate at the School of Data Science, Mathematics, and Informatics, IPB University. He also serves as a lecturer at the Department of Informatics, Nurul Fikri College of Technology. His research focuses on the application of deep learning for chemometrics and spectral analysis. He can be contacted at email: hilmi.tawakal@apps.ipb.ac.id.



Yohanes Aris Purwanto    is a professor in the Division of Biosystem Engineering, Faculty of Engineering and Technology at IPB University. He holds a Ph.D. in Agricultural Engineering from the University of Tokyo. His expertise lies in postharvest engineering, agricultural machinery, and the application of non-destructive methods like NIR spectroscopy for quality evaluation of agricultural products. He is the corresponding author of this paper. He can be contacted at email: arispurwanto@apps.ipb.ac.id.



Sony Hartono Wijaya    is a lecturer in the School of Data Science, Mathematics, and Informatics, IPB University. He earned his Ph.D. from Nara Institute of Science and Technology (NAIST), Japan. His research interests include bioinformatics, machine learning, data mining, information retrieval, and software engineering. He can be contacted at email: sony@apps.ipb.ac.id.



Shelve Nidya Neyman    is a lecturer in the School of Data Science, Mathematics, and Informatics, IPB University. She holds a doctorate from the Institute of Technology Bandung (ITB). Her research interests include information security, cryptography, steganography, and watermarking. She can be contacted at email: shelve@apps.ipb.ac.id.