

## Article

# Improved Cd Detection in Rice Grain Using LIBS with Husk-Based XGBoost Transfer Learning

Weiping Xie <sup>1</sup>, Jiang Xu <sup>1</sup>, Lin Huang <sup>2</sup>, Yuan Xu <sup>3</sup>, Qi Wan <sup>1</sup>, Yangfan Chen <sup>2</sup> and Mingyin Yao <sup>1,\*</sup>

<sup>1</sup> College of Engineering, Jiangxi Agricultural University, Nanchang 330045, China; xie.weiping@foxmail.com (W.X.); xujiangstart@163.com (J.X.)

<sup>2</sup> College of Bioscience and Bioengineering, Jiangxi Agricultural University, Nanchang 330045, China; cyfff000117@163.com (Y.C.)

<sup>3</sup> Ganzhou Agricultural Science Research Institute, Ganzhou 341000, China; yuanxu\_0726@163.com

\* Correspondence: mingyin800@126.com

**Abstract:** Cadmium (Cd) is a highly toxic metal that is difficult to completely eliminate from soil, despite advancements in modern agricultural and environmental technologies that have successfully reduced Cd levels. However, rice remains a key source of Cd exposure for humans. Even small amounts of Cd absorbed by rice can pose a potential health risk to the human body. Laser-induced breakdown spectroscopy (LIBS) has the advantages of simple sample preparation and fast analysis, which, combined with the transfer learning method, is expected to realize the real-time and rapid detection of low-level heavy metals in rice. In this work, 21 groups of naturally matured rice samples from potentially Cd-contaminated environments were collected. These samples were processed into rice husk, brown rice, and polished rice groups, and the reference Cd content was measured by ICP-MS. The XGBoost algorithm, known for its excellent performance in handling high-dimensional data and nonlinear relationships, was applied to construct both the XGBoost base model and the XGBoost-based transfer learning model to predict Cd content in brown rice and polished rice. By pre-training on rice husk source data, the XGBoost-based transfer learning model can learn from the abundant information available in rice husk to improve Cd quantification in rice grain. For brown rice, the XGBoost base model achieved  $R_C^2$  of 0.9852 and  $R_P^2$  of 0.8778, which were improved to 0.9885 and 0.9743, respectively, with the XGBoost-based transfer learning model. In the case of polished rice, the base model achieved  $R_C^2$  of 0.9838 and  $R_P^2$  of 0.8683, while the transfer learning model enhanced these to 0.9883 and 0.9699, respectively. The results indicate that the transfer learning method not only improves the detection capability for low Cd content in rice but also provides new insights for food safety detection.

**Keywords:** Laser-Induced Breakdown Spectroscopy (LIBS); transfer learning; rice husk; Cd quantitative analysis; XGBoost



**Citation:** Xie, W.; Xu, J.; Huang, L.; Xu, Y.; Wan, Q.; Chen, Y.; Yao, M. Improved Cd Detection in Rice Grain Using LIBS with Husk-Based XGBoost Transfer Learning. *Agriculture* **2024**, *14*, 2053. <https://doi.org/10.3390/agriculture14112053>

Academic Editor: Hyeon Tae Kim

Received: 14 October 2024

Revised: 11 November 2024

Accepted: 13 November 2024

Published: 14 November 2024



**Copyright:** © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Rice, the staple food for more than half of the world's population, is widely cultivated and consumed over the world [1]. However, rice is also the major cereal crop that absorbs the most heavy metals [2], making the heavy metal safety of rice of concern among many researchers. Recent advancements in food safety research and modern agricultural and environmental monitoring technologies have reduced the incidence of rice exceeding permitted heavy metal limits. Despite these improvements, heavy metal pollution in soil is difficult to completely eradicate [3], which means that rice may still be subject to different levels of contamination. Heavy metals are characterized by slow metabolism in the human body. Taking the heavy metal Cd as an example, its biological half-life is typically 16 to 30 years, and it tends to accumulate in the liver, kidneys, and bones [4]. Long-term consumption of rice with heavy metals, even at minimal exposure level, can still pose

potential health risks [5]. Therefore, detection of such rice with low levels of heavy metals becomes critical.

Among the various methods of heavy metals detection in rice, traditional techniques are often time-consuming, require chemical reagents, and have certain environmental impacts [6]. Spectroscopic techniques are widely employed in detection and analysis due to their non-polluting characteristics. There are numerous spectroscopic detection methods, including Raman spectroscopy, near-infrared spectroscopy (NIRS), and hyperspectral imaging (HSI), among others. However, each of these methods has its limitations when it comes to the detection of heavy metals. In the case of Raman, fluorescence background interference during detection interferes with the accurate detection of heavy metals [7]. NIR primarily relies on analyzing vibrational information to infer the composition of the sample, which is not particularly effective for the direct detection of heavy metal elements [8]. Similarly, the spectral characteristics of HSI do not distinctly highlight the content of heavy metal elements [9]. In contrast, laser-induced breakdown spectroscopy (LIBS) generates plasma by striking the surface of a sample with a high-energy laser pulse. The spectral characteristics are then analyzed to determine the elemental composition of the sample. It is applied in industrial [10], food [11], and biological fields [12] due to its rapid, environmentally friendly, and multi-element simultaneous detection capabilities [13], which are sensitive to heavy metals.

However, owing to the complex matrix of rice, and the relatively low content of heavy metals compared to other mineral elements, LIBS often suffers from poor accuracy when detecting heavy metals in rice. To solve this problem, Yang et al. [14] employed an ultrasonic-assisted extraction method for pretreatment of rice, which enhanced the rapid detection capability of LIBS for heavy metals in rice. Fu et al. [15], based on the interaction between mineral elements and heavy metals in rice, used the spectral data of mineral elements and heavy metals as model inputs, thereby improving the quantitative analysis capability for heavy metals in rice. Although these methods have improved the accuracy of LIBS for detecting heavy metals in rice, the sample pretreatment process increases experimental complexity. Moreover, while using the interaction between mineral elements and heavy metals, manually selecting relevant mineral elements is laborious and time-consuming.

In the field of machine learning, transfer learning is a method that enhances model performance in the target domain by leveraging knowledge learned from the source domain. It is characterized by its ability to solve data scarcity issues, improve model performance, and reduce training time and costs [16]. In spectral detection, transfer learning is effective for analyzing complex agricultural products. For instance, in tobacco analysis, Shen et al. [17] utilized transfer learning to analyze tobacco leaf spectral data collected from different NIRS machines, mitigating the need for re-detecting samples due to environmental and instrumental variations. Moreover, transfer learning is advantageous for handling variable agricultural products and limited data. Suarin et al. [18] applied transfer learning to honey produced in April, June, and August, finding the June-to-August model to be the most robust. Post transfer learning, the *RMSEP* decreased from 5.128% to 3.401%, significantly improving NIRS performance in honey quality analysis. Transfer learning also excels in enhancing analytical performance. Lin et al. [19] combined LIBS and transfer learning to identify the origins of rice, millet, and oats from five Chinese regions. The average classification accuracy was 88%, with the millet-to-oats model achieving the highest accuracy at 93.81%. The effectiveness of transfer learning in spectral analysis, as demonstrated in studies across various agricultural products and complex analytical tasks, highlights its potential for improving model performance in diverse applications.

In the context of rice analysis, this potential is promising. The data from the source and target domains in transfer learning should exhibit similar feature distributions, and the training tasks should be related. Rice husk, brown rice, and polished rice share similar biochemical properties and heavy metal accumulation patterns [20]. This similarity helps reduce data bias and improve model generalization, providing a foundation for transfer

learning. Moreover, the unique characteristics of rice husk further support its suitability as a source domain for transfer learning. Unlike brown rice and polished rice, which undergo additional milling processes, rice husks are obtained directly through hulling and retain their complete structure. Additionally, rice husk, being the largest by-product of rice processing, is readily available, inexpensive, and abundant [21]. These advantages make rice husk an ideal source domain for transfer learning in rice analysis.

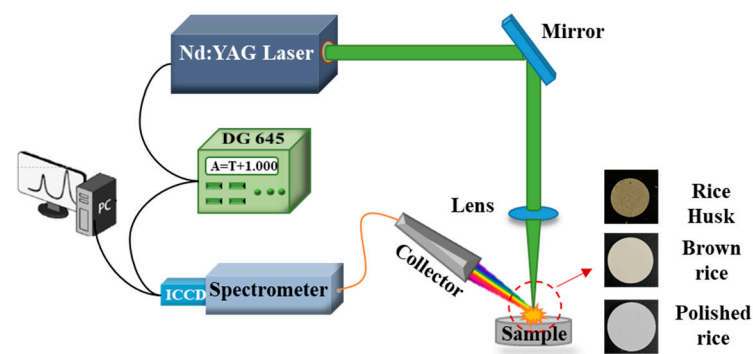
Building on these advantages, integrating eXtreme Gradient Boosting (XGBoost) with transfer learning offers a powerful enhancement to this process, compared to traditional machine learning algorithms such as Support Vector Machines (SVM), Logistic Regression (LR), and Random Forests (RF). XGBoost is an ensemble learning algorithm based on gradient boosting decision trees which offers rapid training and efficient computational performance, excelling in regression tasks [22]. It is widely applied in industry [23], medical diagnostics [24], and environmental assays [25], and has also shown progress in spectroscopy. For instance, to accurately predict arsenic (As) levels in soil by hyperspectral technology, Ye et al. [8] integrated Geographically Weighted Regression (GWR) with the XGBoost algorithm. The model achieved a prediction accuracy of 90%, establishing a great correlation between spectra and concentration. Similarly, Zeng et al. [26] integrated Raman spectroscopy with XGBoost and GBDT models to detect and analyze the novel coronavirus (SARS-CoV-2), where XGBoost demonstrated superior robustness. When combined with the SVM-REF algorithm, the model achieved an analysis accuracy of 93.55%. Furthermore, XGBoost has also been applied in transfer learning: Gao et al. [22] used Raman spectroscopy to detect lignin in cedar wood, comparing the quantitative analysis of XGBoost, LightGBM, and CatBoost models with and without transfer learning. The results showed improved performance for all models post transfer learning, where XGBoost showed more stability and accuracy. These studies demonstrate that XGBoost is robust in spectral analysis and advantageous in transfer learning applications. XGBoost's efficient computational capabilities and ability to manage large and complex datasets might enhance the performance of LIBS in identifying trace heavy metals in rice.

In this work, based on the rich data from rice husks, we constructed an XGBoost-based transfer learning model, then compared the performance of the XGBoost base model with the XGBoost-based transfer learning model. The coefficients of determination ( $R^2$ ) and root mean square error (RMSE) were used for quantitative analysis of Cd in brown rice and polished rice as the evaluation indexes. Thus, we explored the potential of transfer learning in LIBS analysis of low-content heavy metals in rice.

## 2. Materials and Methods

### 2.1. Experimental Setup

The LIBS experimental setup is shown in Figure 1. A Q-switched Nd:YAG laser (Beamtech, Vlite 200, Beijing, China) with an emission wavelength of 1064 nm, frequency of 2 Hz, and pulse duration of 8 ns serves as the laser source. The laser energy, measured by a laser energy meter (National Institute of Metrology E-1000, Beijing, China), is 170 mJ. The laser first passes through a 45° mirror and is then focused on the sample surface using a focusing lens. The plasma signal generated by the laser is collected by an optical fiber and transmitted to a spectrometer (Avantes, AvaSpec-2048FT-8R, Apeldoorn, The Netherlands) with a wavelength range of 200–1050 nm. The spectrometer converts the signal and transmits it to a computer. To prevent plasma instability caused by repeated laser ablation at the same sample location, the sample is placed on a two-dimensional rotating platform (Zolix, SC300-1A, Beijing, China). The timing between the spectrometer and the laser is controlled by a DG645 delay generator (Stanford Research Systems, Sunnyvale, CA, USA).



**Figure 1.** Schematic diagram of LIBS experimental setup.

## 2.2. Sample Preparation and Data Preprocessing

In order to take sufficient account of the geographic variability of Cd contamination, samples for this work were collected from 21 planting areas. The collected rice was dried, hulled, milled, and polished to produce samples of rice husk, brown rice, and polished rice. To minimize the impact of heterogeneous elemental distribution on detection results, physical preparation methods were applied to rice samples [27]. The prepared samples were crushed, passed through a 100-mesh sieve, and pressed into pellets with a diameter of 25 mm and a weight of 3 g. Three replicates were prepared for each sample, and 45 spectra were collected for each pellet. Following the Chinese national standard “GB/T 35876-2018” [28], ICP-MS was used to measure the concentration of heavy metal Cd in rice husks, brown rice, and polished rice from 21 planting areas. The reference contents are shown in Table 1. The results indicated that most brown rice and polished rice samples had low heavy metal content, although #21 rice sample and #19, #20 and #21 rice husk samples exceeded the maximum limit of 0.2 mg/kg stipulated in the Chinese national standard “GB 2762-2017” [29]. Based on the Target Hazard Quotient (THQ) standard, Cd has a THQ value of 0.1 mg/kg [30]. THQ is an indicator used to assess the health risks associated with long-term exposure to specific pollutants. When the THQ value exceeds 1, it indicates more health risks. For Cd, a THQ value of 0.1 mg/kg implies that long-term intake of cadmium at this concentration does not pose significant health risks. Therefore, defining rice samples with Cd content below 0.1 mg/kg as low Cd rice samples is based on this health risk assessment standard.

**Table 1.** Reference Cd content of rice by the ICP-MS method (mg/kg).

Sample	Rice Husk	Brown Rice	Polished Rice
#1	0.0118	0.0132	0.0127
#2	0.0122	0.0093	0.0096
#3	0.0206	0.0338	0.0298
#4	0.0238	0.0248	0.0249
#5	0.0268	0.0260	0.0288
#6	0.0287	0.0298	0.0293
#7	0.0326	0.0378	0.0383
#8	0.0496	0.0678	0.0630
#9	0.0614	0.0395	0.0412
#10	0.0670	0.0729	0.0730
#11	0.0776	0.0848	0.0823
#12	0.0804	0.0888	0.0890
#13	0.0822	0.0790	0.0788
#14	0.0910	0.0976	0.0938
#15	0.1180	0.1230	0.1240
#16	0.1230	0.0822	0.0692
#17	0.1240	0.1520	0.1430
#18	0.1890	0.1630	0.1560

Table 1. Cont.

Sample	Rice Husk	Brown Rice	Polished Rice
#19	0.2400	0.1660	0.1470
#20	0.2480	0.1860	0.1900
#21	0.2900	0.2000	0.2000

To enhance the quality of LIBS spectral data, the spectra are usually preprocessed. This work combines the Standard Normal Variate (SNV) and Moving Average (MA) preprocessing methods to mitigate drift and noise in the spectral data [31]. The source domain dataset was randomly divided into training, validation, and prediction sets in a 3:1:1 ratio, ensuring that most data is used for model training while retaining some for validation and testing to maintain the model's generalization. Meanwhile, the target domain dataset is divided into training, validation, and prediction sets in a 1:2:2 ratio, allocating more data for validation and testing. The XGBoost base model's dataset was randomly divided into training, validation, and prediction sets in a 3:1:1 ratio.

### 2.3. Transfer Learning Based on XGBoost

In the field of transfer learning, both deep learning algorithms and machine learning algorithms have been effectively utilized. Deep learning algorithms typically excel with large-scale datasets, but this study involves a relatively small dataset. Consequently, machine learning algorithms were chosen to build the transfer learning model. This helps to conserve computational resources and time. XGBoost, an ensemble learning algorithm based on gradient-boosting decision trees, is ideal for LIBS data analysis due to its efficiency and accuracy in handling high-dimensional data [32]. The construction process of XGBoost is shown in Figure 2. XGBoost evaluates the importance of spectral features by calculating the Gain, automatically selecting the most relevant spectral features that are most helpful for Cd quantitative analysis to construct the learning tree [33]. Unlike traditional decision tree algorithms that construct the entire tree at once, XGBoost builds each tree incrementally by correcting the errors of the previous tree. This step-by-step tree-building process allows the model to continuously adjust and optimize, enhancing its adaptation to data characteristics.

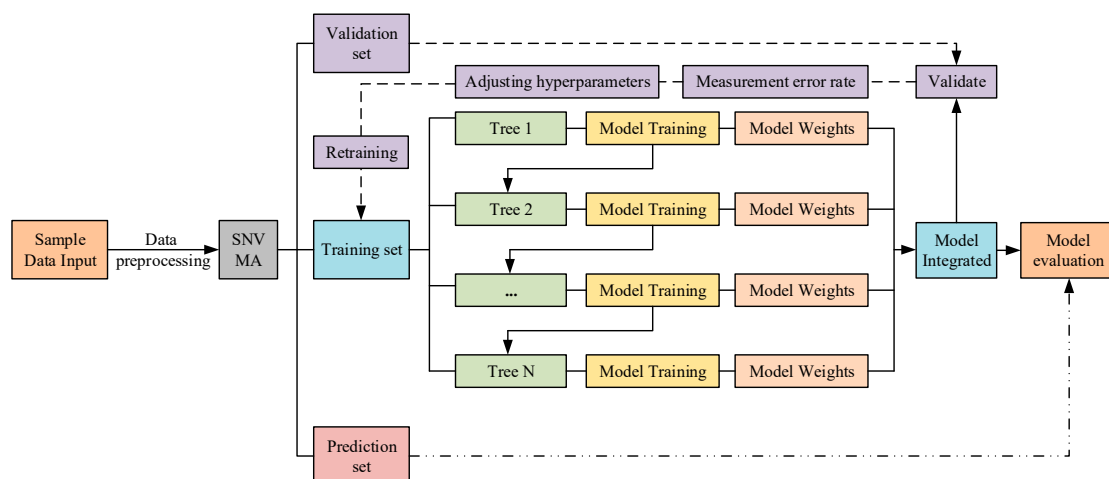
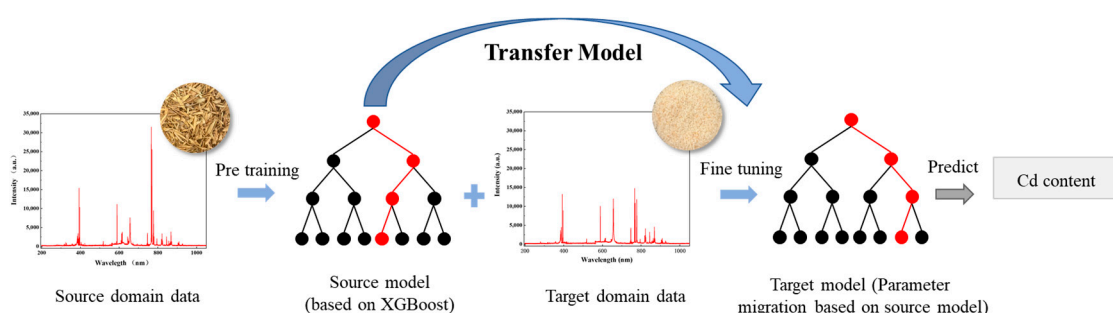


Figure 2. Schematic of the XGBoost structure used for modeling.

Although XGBoost is efficient and accurate, in practical applications, it often requires a large amount of data and computational resources and time to build models directly with XGBoost [34]. Transfer learning can improve the performance of new tasks by utilizing existing knowledge, and reducing the need for extensive data. Therefore, the XGBoost-based transfer learning method can be an ideal choice. This method realizes knowledge

transfer by sharing parameters or prior distributions between the source and target domains, to improve the prediction ability of target domains. The schematic diagram of XGBoost transfer learning is shown in Figure 3. Rice husk was selected as the source domain for the transfer learning model due to its rich spectral information, low cost, wide availability, and the shared growing environment with brown and polished rice. The pre-training phase involved learning the elemental features related to Cd contamination in rice husk and constructing a pre-training model. This model was then transferred to brown or polished rice, where it conducted further training based on the spectral data from rice grains. During this phase, parameter fine-tuning was performed to adapt to the specific characteristics of the samples in the target domain. The final model constructed is an XGBoost-based transfer learning model. Furthermore, the step-by-step optimization feature of XGBoost is better suited for transfer learning, allowing the model to fit the data distribution of brown rice and polished rice during fine-tuning, thereby enhancing the model's generalization ability.



**Figure 3.** Schematic of migration learning from pre-trained rice husk XGBoost model to target domain.

Additionally, the grid search method is used to automatically return the optimal parameter set after evaluating all parameter combinations. The hyperparameter settings for the XGBoost-based transfer learning model are detailed in Table 2. The hyperparameter `n_estimators` represents the number of trees to be trained; an appropriate number of trees ensures the model has enough complexity to capture data features. The `max_depth` stands for the maximum depth of each tree; in the case of more noise, a shallower depth can prevent the model from overfitting. The `gamma` controls the flexibility of splitting nodes, and lower `gamma` value allows the model to capture more feature details. The `learning_rate` balances the learning speed and stability of the model. Hyperparameter values are determined via grid search, eliminating biases from subjective selection and ensuring optimal model performance.

**Table 2.** Hyperparameters of XGBoost-based transfer learning models.

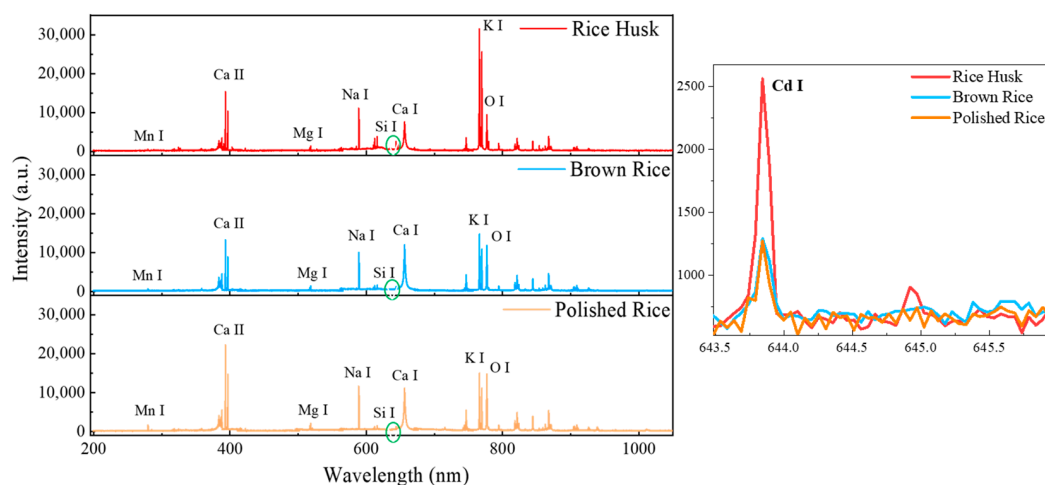
Hyperparameterization	Experimental Value	Range
<code>n_estimators</code>	200	100–300
<code>max_depth</code>	3	3–5
<code>gamma</code>	0.001	0–0.01
<code>learning_rate</code>	0.1	0–0.2

### 3. Conclusions and Analysis

#### 3.1. LIBS Spectral Analysis of Rice Husk, Brown Rice and Polished Rice

Taking the spectrogram of #19 sample as an example, the LIBS spectra of rice husk, brown rice, and polished rice from the same rice sample are compared in Figure 4. The analysis reveals that the main excited mineral elements are Mn I 279.827, Mn I 280.108, Mg I 517.268, Mg I 518.360, Na I 588.995, Na I 589.592, Si I 615.51, Ca II 393.366, Ca II 396.847, K I 766.490, and K I 769.90. The spectral intensities of elements vary significantly

among rice husk, brown rice, and polished rice. Mn is predominantly found in the bran layer and endosperm [35], while rice husk is mainly composed of cellulose and contains a large amount of Si [36]. The spectral intensity of Mn in brown rice and polished rice is significantly higher than in rice husk, while the Si spectral line intensity is highest in rice husk. The spectral trends correlate with the elemental composition of the rice.



**Figure 4.** Typical LIBS spectrum of #19 rice sample.

Due to the relatively low content of Cd in rice compared to other mineral elements, the intensity of the Cd spectral line in the spectrum is also weak. Referring to the analysis by Su et al. and Fu et al. [30,37], Cd I 643.847 was selected as the target for analysis. The complex matrix effect of agricultural products often results in poor accuracy for LIBS quantitative analysis of low Cd content in rice. XGBoost leverages its powerful learning capabilities and flexible feature processing to effectively uncover the complex relationships between Cd content and other variables.

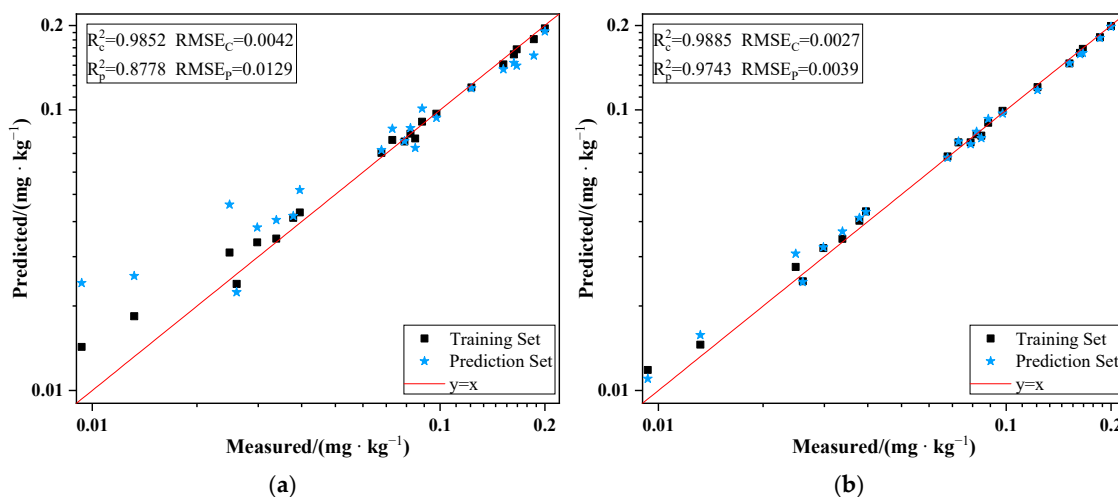
This work utilizes Gain for automatic feature extraction learning by XGBoost, and to build the XGBoost base model and XGBoost-based transfer learning model. Visualizing the XGBoost decision tree showed that the first tree constructed by XGBoost uses the Si and Na spectral feature as one of the splitting points for learning. What is more, research by Fu et al. [30] has demonstrated that Si, Na and other mineral elements respond to Cd stress. It can be observed that the features extracted by XGBoost have a certain intrinsic connection with Cd. The transfer learning model also compensates for the lack of spectral information in rice grains by learning the spectral characteristics of elements such as Si and Na in rice husk. While Figure 4 shows that the spectral intensities of most spectral lines were higher in rice husk than in brown and polished rice, the spectral intensities of Mn and Ca were lower compared to those in brown and polished rice. Despite this, these elements interact with Cd and were successfully feature-extracted and learned by the pre-training model. During the transfer learning process, an important parameter fine-tuning phase allows for further in-depth learning of Mn and Ca, ensuring that the final prediction model can adapt to the specific characteristics of brown and polished rice samples. This process of knowledge transfer and parameter fine-tuning significantly enhances the model's generalization ability.

### 3.2. XGBoost Base Model and XGBoost-Based Transfer Learning Model for Quantitative Analysis of Cd in Rice

The rice husk, brown rice, and polished rice share the same environmental conditions during growth, establishing an intrinsic connection among them. The rice husk, serving as the outer protective layer of the grain, contains various mineral elements and organic compounds. For example, Si is one of the components of rice husk, whereas the Si content

gradually decreases in brown rice and polished rice. This intrinsic connection between the rice husk and the inner grain provides a foundation for transfer learning.

Figure 5a shows the fitting curve of the XGBoost base model predicting Cd content in brown rice, while Figure 5b presents the prediction of the XGBoost-based transfer learning model. For the XGBoost base model of brown rice, the  $R_p^2$  is 0.8778, which improves to 0.9743 after transfer learning, and the  $RMSEP$  decreases from 0.0129 mg/kg to 0.0039 mg/kg, demonstrating that the transfer learning method significantly enhances the model's prediction performance. This improvement is attributed to the rice husk providing rich elemental characteristic information for the XGBoost-based transfer learning model.

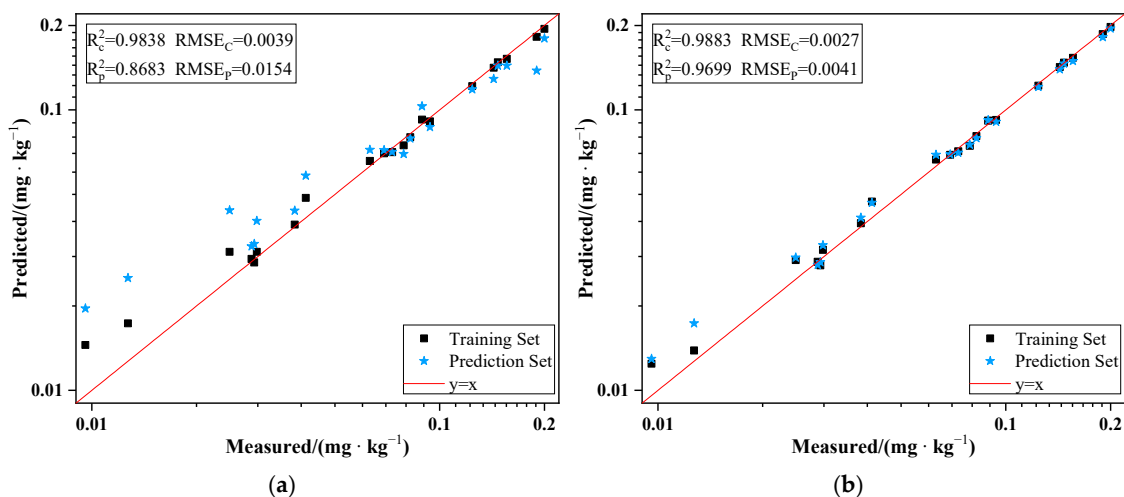


**Figure 5.** Analytical curve of Cd content in brown rice. (a) XGBoost base model; (b) XGBoost-based transfer learning model.

Due to the Cd content in brown rice being relatively low, the XGBoost base model exhibits a large error in predicting Cd levels. In Figure 5a, the prediction set data points are scattered, whereas in Figure 5b, the data points are more concentrated, particularly for samples with lower Cd content. Additionally, the XGBoost-based transfer learning model exhibits higher consistency and better generalization between the prediction and training sets. This demonstrates that pre-training on rice husk domain data can reduce overfitting risks on target domain data while enhancing prediction accuracy. Furthermore, rice husk is a major by-product of rice processing, which is abundant and inexpensive. Its wide availability and low cost make rice husk a feasible option for practical applications in transfer learning. Utilizing rice husk for spectral data analysis can significantly reduce experimental costs and make efficient use of agricultural waste.

To assess the performance of the XGBoost-based transfer learning model on different but related datasets, the rice husk model was transferred to polished rice to verify the effectiveness of transfer learning in various rice applications. Figure 6a shows the prediction results of the XGBoost base model using the spectrum data of polished rice. The direct application of XGBoost led to a scattered data distribution and poor prediction accuracy. Similar to the XGBoost base model constructed with brown rice data, lower Cd content led to more scattered prediction points. After applying transfer learning with rice husk data, the model's fit improved significantly: as shown in Figure 6b, the  $R_p^2$  increased from 0.8683 to 0.9699, and the  $RMSEP$  decreased from 0.0154 mg/kg to 0.0041 mg/kg. This demonstrates that the XGBoost-based transfer learning model performs well on both brown rice and polished rice datasets. By calculating the  $RMSEP$  for samples with lower Cd content (below 0.1 mg/kg), the  $RMSEP$  of polished rice was reduced from 0.0019 mg/kg to 0.0005 mg/kg after transfer learning, while that of brown rice was reduced from 0.0021 mg/kg to 0.0005 mg/kg. Transfer learning significantly improved the predictive accuracy of the model, especially for low Cd content in rice.





**Figure 6.** Analytical curve of Cd content in polished rice. (a) XGBoost base model; (b) XGBoost-based transfer learning model.

Compared to the XGBoost base model, the XGBoost-based transfer learning model pre-trained on rice husk data effectively learns more characteristic information of Cd and related elements. This approach enhances the model's overall performance, even when the spectral intensity of Cd is weaker than that of other mineral elements. Table 3 presents a comparison between the prediction results of the XGBoost base model and the XGBoost-based transfer learning model. The transfer learning model shows 10.99% improvement in  $R_p^2$  for brown rice and 11.70% for polished rice, along with a significant reduction in  $RMSEP$ . Thus, the transfer learning method improved the prediction of Cd content in brown rice and polished rice.

**Table 3.** Comparison of XGBoost predictions without and with migration learning.

Target Domains		$R_p^2$	$RMSEP$ (mg/kg)
Brown rice	XGBoost base model	0.8778	0.0129
	XGBoost-based transfer learning model	0.9743	0.0039
	Variations	10.99%	−69.77%
Polished rice	XGBoost base model	0.8683	0.0154
	XGBoost-based transfer learning model	0.9699	0.0041
	Variations	11.7%	−73.37%

#### 4. Conclusions

This work compared an XGBoost base model with an XGBoost-based transfer learning model using LIBS spectral data from rice husk, brown rice, and polished rice. The results indicate that the XGBoost-based transfer learning model significantly enhanced the prediction ability of Cd content in rice, particularly for low Cd content in rice, compared to the XGBoost base model directly using spectrum data from brown rice and polished rice. The inherent correlation between rice husk and rice grain allows the model to leverage the rich elemental information in the rice husk, improving performance in the rice grain target domain. The XGBoost-based transfer learning model achieved  $R_p^2$  of 0.9743 for brown rice, a 10.99% improvement over the XGBoost base model, and  $R_p^2$  of 0.9699 for polished rice, an 11.70% increase. This approach enhanced predictive accuracy and reduced overfitting risk.

Furthermore, the XGBoost-based transfer learning model offers a novel method for predicting heavy metal content in rice and provides new insights for rice quality and safety assessment.

**Author Contributions:** Conceptualization, J.X. and M.Y.; methodology, W.X. and L.H.; software, W.X. and J.X.; validation, Q.W.; formal analysis, Q.W.; investigation, Y.X.; resources, Y.X.; data curation, L.H.; writing—original draft preparation, W.X.; writing—review and editing, J.X. and M.Y.; visualization, Y.C.; supervision, M.Y.; project administration, M.Y.; funding acquisition, M.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially supported by the National Natural Science Foundation of China (NSFC) [Grant No. 32260632].

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Ali, W.; Mao, K.; Zhang, H.; Junaid, M.; Xu, N.; Rasool, A.; Feng, X.; Yang, Z. Comprehensive review of the basic chemical behaviours, sources, processes, and endpoints of trace element contamination in paddy soil-rice systems in rice-growing countries. *J. Hazard. Mater.* **2020**, *397*, 122720. [[CrossRef](#)] [[PubMed](#)]
2. Ma, J.F.; Shen, R.F.; Shao, J.F. Transport of cadmium from soil to grain in cereal crops: A review. *Pedosphere* **2021**, *31*, 3–10. [[CrossRef](#)]
3. Li, Z.; Liang, Y.; Hu, H.; Shaheen, S.M.; Zhong, H.; Tack, F.M.G.; Wu, M.; Li, Y.-F.; Gao, Y.; Rinklebe, J.; et al. Speciation, transportation, and pathways of cadmium in soil-rice systems: A review on the environmental implications and remediation approaches for food safety. *Environ. Int.* **2021**, *156*, 106749. [[CrossRef](#)]
4. Charkiewicz, A.E.; Omeljaniuk, W.J.; Nowak, K.; Garley, M.; Niklinski, J. Cadmium Toxicity and Health Effects—A Brief Summary. *Molecules* **2023**, *28*, 6620. [[CrossRef](#)] [[PubMed](#)]
5. Wei, R.; Chen, C.; Kou, M.; Liu, Z.; Wang, Z.; Cai, J.; Tan, W. Heavy metal concentrations in rice that meet safety standards can still pose a risk to human health. *Commun. Earth Environ.* **2023**, *4*, 84. [[CrossRef](#)]
6. Khan, Z.H.; Ullah, M.H.; Rahman, B.; Talukder, A.I.; Wahadoszamen, M.; Abedin, K.M.; Haider, A.F.M.Y.; Galić, N. Laser-Induced Breakdown Spectroscopy (LIBS) for Trace Element Detection: A Review. *J. Spectrosc.* **2022**, *2022*, 1–25. [[CrossRef](#)]
7. Wang, Y.; Fang, L.; Wang, Y.; Xiong, Z. Current Trends of Raman Spectroscopy in Clinic Settings: Opportunities and Challenges. *Adv Sci (Weinh)* **2024**, *11*, e2300668. [[CrossRef](#)]
8. Ye, M.; Zhu, L.; Li, X.; Ke, Y.; Huang, Y.; Chen, B.; Yu, H.; Li, H.; Feng, H. Estimation of the soil arsenic concentration using a geographically weighted XGBoost model based on hyperspectral data. *Sci. Total Environ.* **2023**, *858*, 159798. [[CrossRef](#)] [[PubMed](#)]
9. Khan, M.J.; Khan, H.S.; Yousaf, A.; Khurshid, K.; Abbas, A. Modern Trends in Hyperspectral Image Analysis: A Review. *IEEE Access* **2018**, *6*, 14118–14129. [[CrossRef](#)]
10. Pedarnig, J.D.; Trautner, S.; Grünberger, S.; Giannakaris, N.; Eschlböck-Fuchs, S.; Hofstadler, J. Review of Element Analysis of Industrial Materials by In-Line Laser—Induced Breakdown Spectroscopy (LIBS). *Appl. Sci.* **2021**, *11*, 9274. [[CrossRef](#)]
11. Velásquez-Ferrín, A.; Babos, D.V.; Marina-Montes, C.; Anzano, J. Rapidly growing trends in laser-induced breakdown spectroscopy for food analysis. *Appl. Spectrosc. Rev.* **2020**, *56*, 492–512. [[CrossRef](#)]
12. Gaudiuso, R.; Melikechi, N.; Abdel-Salam, Z.A.; Harith, M.A.; Palleschi, V.; Motto-Ros, V.; Busser, B. Laser-induced breakdown spectroscopy for human and animal health: A review. *Spectrochim. Acta Part. B At. Spectrosc.* **2019**, *152*, 123–148. [[CrossRef](#)]
13. Guo, L.-B.; Zhang, D.; Sun, L.-X.; Yao, S.-C.; Zhang, L.; Wang, Z.-Z.; Wang, Q.-Q.; Ding, H.-B.; Lu, Y.; Hou, Z.-Y.; et al. Development in the application of laser-induced breakdown spectroscopy in recent years: A review. *Front. Phys.* **2021**, *16*, 1–25. [[CrossRef](#)]
14. Yang, P.; Zhou, R.; Zhang, W.; Yi, R.; Tang, S.; Guo, L.; Hao, Z.; Li, X.; Lu, Y.; Zeng, X. High-sensitivity determination of cadmium and lead in rice using laser-induced breakdown spectroscopy. *Food Chem.* **2019**, *272*, 323–328. [[CrossRef](#)]
15. Fu, G.; Hu, W.; Xie, W.; Yao, X.; Xu, J.; Yang, P.; Yao, M. Quantitative analysis of Cd based on the stress effect of minerals in rice by laser-induced breakdown spectroscopy. *Anal. Methods* **2023**, *15*, 5867–5874. [[CrossRef](#)]
16. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. In Proceedings of the Institute of Electrical and Electronics Engineers, Bari, Italy, 7–10 September 2021; Volume 109, pp. 43–76. [[CrossRef](#)]
17. Shen, H.; Geng, Y.; Ni, H.; Wang, H.; Wu, J.; Hao, X.; Tie, J.; Luo, Y.; Xu, T.; Chen, Y.; et al. Across different instruments about tobacco quantitative analysis model of NIR spectroscopy based on transfer learning. *RSC Adv.* **2022**, *12*, 32641–32651. [[CrossRef](#)]
18. Suarin, N.A.S.; Chia, K.S.; Mohamad Fuzi, S.F.Z. Transfer learning in near infrared spectroscopy for stingless bee honey quality prediction across different months. *Knowl-Based Syst.* **2024**, *295*, 111817. [[CrossRef](#)]
19. Lin, P.; Wen, X.; Ma, S.; Liu, X.; Xiao, R.; Gu, Y.; Chen, G.; Han, Y.; Dong, D. Rapid identification of the geographical origins of crops using laser-induced breakdown spectroscopy combined with transfer learning. *Spectrochim. Acta Part B At. Spectrosc.* **2023**, *206*, 106729. [[CrossRef](#)]
20. Choi, S.H.; Choi, E.M.; Lee, Y.R.; Park, K.S. Study of the Transition Pattern of Heavy Metal Absorption in a Rice-Related Matrix. *Anal. Lett.* **2021**, *54*, 2171–2181. [[CrossRef](#)]

21. Singh, R.; Patel, M. Effective utilization of rice straw in value-added by-products: A systematic review of state of art and future perspectives. *Biomass Bioenergy* **2022**, *159*, 106411. [[CrossRef](#)]
22. Gao, W.; Jiang, Q.; Guan, Y.; Huang, H.; Liu, S.; Ling, S.; Zhou, L. Transfer learning improves predictions in lignin content of Chinese fir based on Raman spectra. *Int. J. Biol. Macromol.* **2024**, *269*, 132147. [[CrossRef](#)] [[PubMed](#)]
23. Nguyen, H.; Bui, X.-N.; Bui, H.-B.; Cuong, D.T. Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: A case study. *Acta Geophys.* **2019**, *67*, 477–490. [[CrossRef](#)]
24. Maleki, A.; Raahemi, M.; Nasiri, H. Breast cancer diagnosis from histopathology images using deep neural network and XGBoost. *Biomed. Signal Process. Control* **2023**, *86*, 105152. [[CrossRef](#)]
25. Tymoteusz, M.; Kozłowska, P.; Krzemińska, A.; Lewita, K.; Biedrzycka, J.; Geroch, K. Xgboost in Environmental Ecology: A Powerful Tool for Sustainable Insights. *Grail Sci.* **2023**, *34*, 163–170. [[CrossRef](#)]
26. Zeng, W.; Wang, Q.; Xia, Z.; Li, Z.; Qu, H. Application of XGBoost Algorithm in The Detection of SARS-CoV-2 Using Raman Spectroscopy. *J. Phys. Conf. Ser.* **2021**, *1775*, 012007. [[CrossRef](#)]
27. Yang, P.; Fu, G.; Wang, J.; Luo, Z.; Yao, M. A tutorial review on methods of agricultural product sample pretreatment and target analysis by laser-induced breakdown spectroscopy. *J. Anal. At. Spectrom.* **2022**, *37*, 1948–1960. [[CrossRef](#)]
28. GB/T 35876-2018; National Health and Family Planning Commission of the People's Republic of China; China Food and Drug Administration. Inspection of Grain and Oils-Determination of Sodium, Magnesium, Kalium, Calcium, Chromium, Manganese, Iron, Copper, Zinc, Arsenie, Selenium, Cadmium and Plumbum in Cereals and Derived Products-Inductively Coupled Plasamass Spectrometrie Method. National Standardization Administration of China, SAC: Beijing, China, 2018.
29. GB 2762-2017; National Health Commission of the People's Republic of China; National Food Safety Standard—Maximum Levels of Contaminants in Foods. National Standardization Administration of China, SAC: Beijing, China, 2017.
30. Fu, G.; Li, Z.; Xu, J.; Xie, W.; Yang, P.; Xu, Y.; Yao, M. Prediction of heavy metal Cd and stress on minerals in rice by analysis of LIBS spectra. *Appl. Opt.* **2022**, *61*, 2536–2541. [[CrossRef](#)] [[PubMed](#)]
31. Zhou, N.; Hu, T.; Wu, M.; Chen, Q.; Qi, C. Comparative analysis of machine learning algorithms for identifying cobalt contamination in soil using spectroscopy. *J. Environ. Chem. Eng.* **2024**, *12*, 113328. [[CrossRef](#)]
32. Alix, G.; Lymer, E.; Zhang, G.; Daly, M.; Gao, X. A comparative performance of machine learning algorithms on laser-induced breakdown spectroscopy data of minerals. *J. Chemom.* **2022**, *37*, e3400. [[CrossRef](#)]
33. Loss, F.P.; Pedro, H.; da Cunha, P.H.; Rocha, M.B.; Zanon, M.P.; de Lima, L.M.; Nascimento, I.T.; Rezende, I.; Canuto, T.R.P.; Rossoni, R.; et al. Skin cancer diagnosis using NIR spectroscopy data of skin lesions in vivo using machine learning algorithms. *arXiv* **2024**, arXiv:2401.01200.
34. Pan, S.; Zheng, Z.; Guo, Z.; Luo, H. An optimized XGBoost method for predicting reservoir porosity using petrophysical logs. *J. Pet. Sci. Eng.* **2022**, *208*, 109520. [[CrossRef](#)]
35. Ram, H.; Gandass, N.; Sharma, A.; Singh, A.; Sonah, H.; Deshmukh, R.; Pandey, A.K.; Sharma, T.R. Spatio-temporal distribution of micronutrients in rice grains and its regulation. *Crit. Rev. Biotechnol.* **2020**, *40*, 490–507. [[CrossRef](#)] [[PubMed](#)]
36. Zou, Y.; Yang, T. Rice Husk, Rice Husk Ash and Their Applications. In *Rice Bran Rice Bran Oil*; Elsevier: Alpharetta, GA, USA, 2019; pp. 207–246.
37. Su, L.; Shi, W.; Chen, X.; Meng, L.; Yuan, L.; Chen, X.; Huang, G. Simultaneously and quantitatively analyze the heavy metals in *Sargassum fusiforme* by laser-induced breakdown spectroscopy. *Food Chem.* **2021**, *338*, 127797. [[CrossRef](#)] [[PubMed](#)]

**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.